

# **Structural Equation Models for LSC Teacher Questionnaire and Classroom Observation Data**

by

Eric R. Banilower  
Amy A. Germuth

**June 2004**

Prepared For: The National Science Foundation  
4201 Wilson Boulevard  
Arlington, VA 22230

Prepared By: Horizon Research, Inc.  
326 Cloister Court  
Chapel Hill, NC 27514

## INTRODUCTION

In 1995 the National Science Foundation initiated the Local Systemic Change (LSC) through Teacher Enhancement program to improve instruction in science, mathematics, and technology. As part of this program, each teacher in a LSC-targeted project was expected to participate in 130 hours of professional development over the course of the LSC project with the emphasis on preparing teachers to implement exemplary science and mathematics instructional materials and lessons in their classes.<sup>1</sup> This study, utilizing cross-sectional teacher questionnaire and classroom observation data collected between 1999 and 2003,<sup>2</sup> uses structural equation modeling to examine the relationships among teachers' participation in LSC professional development; their use of LSC-designated instructional materials; and their attitudes, perceptions, and classroom practice.<sup>3</sup> Additionally the models examine how well teachers' perceptions of principal support predict teachers' classroom practice.

## METHODOLOGY

In this report two structural equation models are examined using MPLUS v. 3.0. Structural equation modeling (SEM) is a confirmatory form of multiple regression which takes into account the modeling of interactions, nonlinearities, correlated independents, measurement error, and correlated error terms without leading to an increased Type I error rate. Advantages of SEM compared to multiple regression include the desirability of simultaneously testing all of the relationships in a model rather than individual coefficient one at a time. SEM also allows for models with multiple dependent variables, mediating variables, and the modeling of error terms.

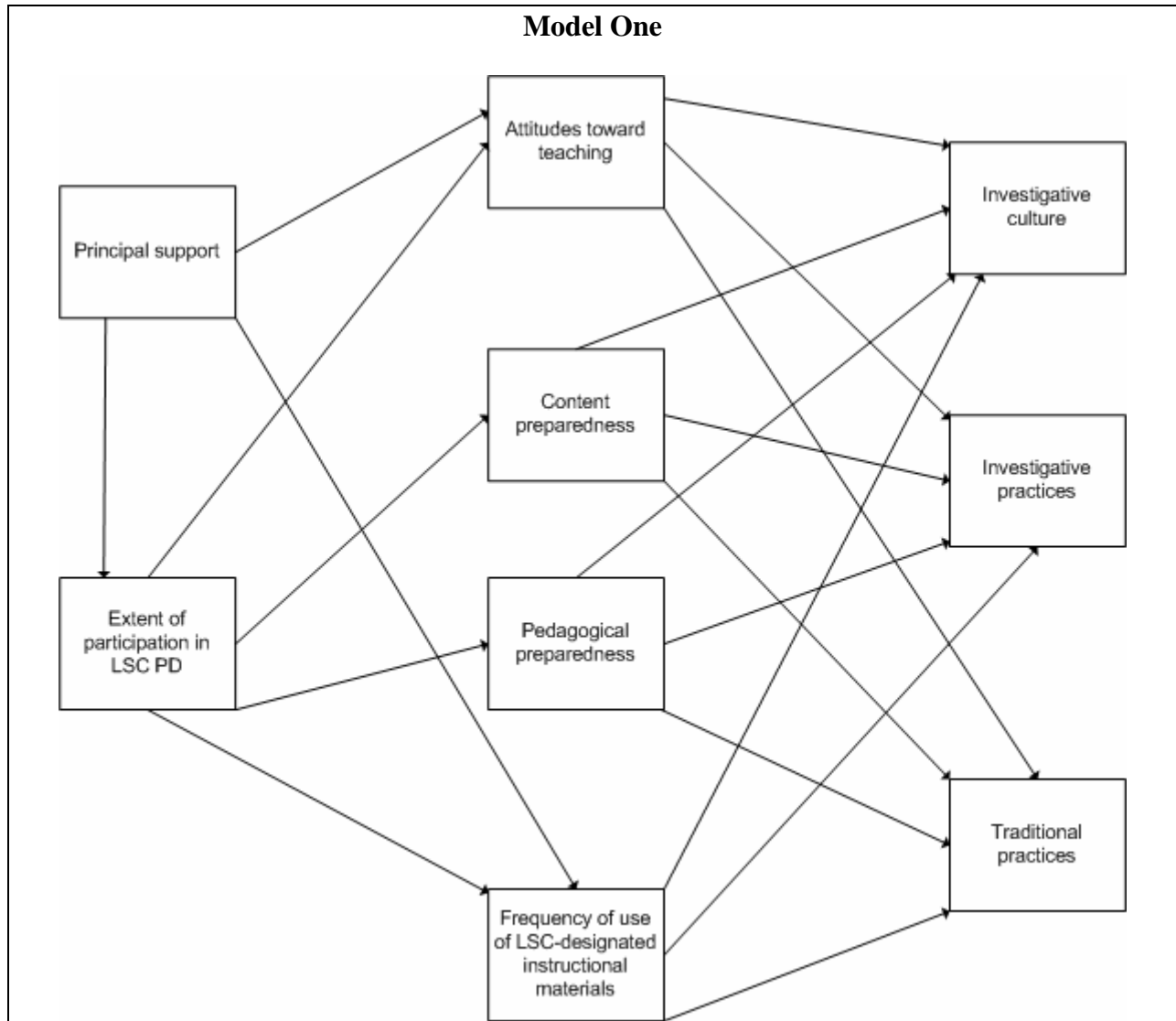
These analyses examine two models. Model One examines the relationship among several key variables from the LSC teacher questionnaire, including teacher participation in LSC professional development and a number of composite variables measuring their perceptions of preparedness and principal support, and the frequency of use of various classroom activities (see Figure 1). This model utilizes data from K–8 science, K–8 mathematics, 6–12 science, and 6–12 mathematics projects.

---

<sup>1</sup> Prior to 1999, the requirement for K–8 projects was 100 hours.

<sup>2</sup> Although the first cohort of LSC projects and the core evaluation began in 1995, 1999 was the first year the full set of the indicators used in these analyses were included on the instruments.

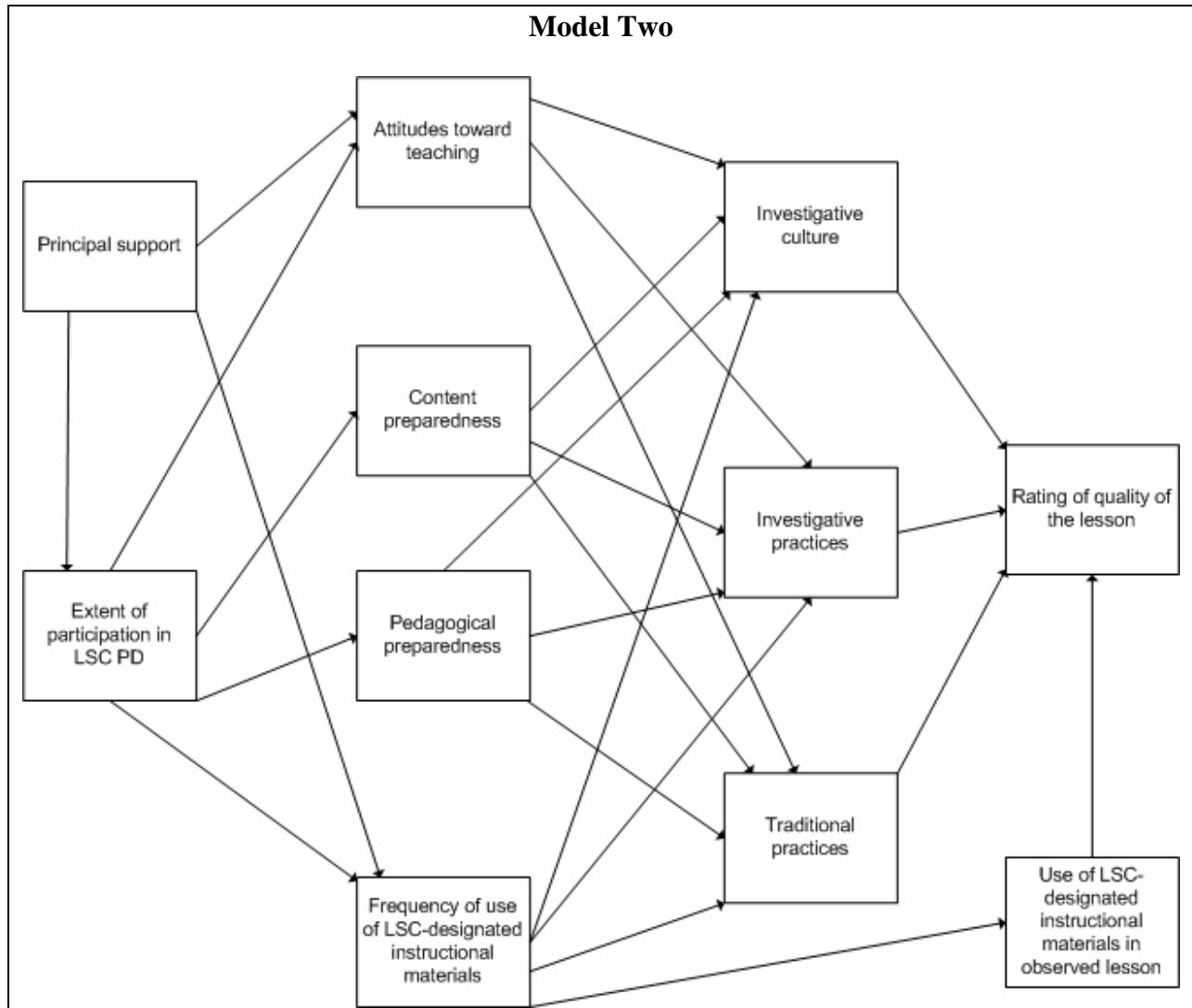
<sup>3</sup> There are three composite variables measuring classroom practice (investigative practices, investigative culture, and traditional practices). These three composites are independent; a teacher can score highly on all three composites.



*Figure 1*

Model Two builds on Model One by adding data from classroom observations, specifically, the evaluator rating of the quality of the observed lesson and whether the lesson was based on the LSC-designated instructional materials (see Figure 2). Because the observation sample is a subgroup of the questionnaire sample, Model Two is based upon a smaller sample of cases than Model One and was calculated for K–8 science, K–8 mathematics, and 6–12 mathematics projects.<sup>4</sup> (The exact statistical models are shown in the Appendix.) Since the relationships among the variables in the models might vary by subject and grade-level, each model is estimated separately for the four subject/grade-range groups targeted by the LSC program.

<sup>4</sup> There were too few observed teachers (N = 75) to run this model for 6–12 science projects.



*Figure 2*

In conducting these analyses, a number of methodological issues had to be considered including the complex sampling design under which the data were collected, the impact of missing data, and the effect of non-normally distributed variables. The following describes how these issues were handled.

### **Independence of Data**

Due to limitations in modeling software available at the time these analyses were conducted, it was not possible to take advantage of the longitudinal data collected through the core evaluation. Thus, the structural models are based upon cross-sectional data. In order to maintain the independence of the different treatment groups, teachers with multiple data points were identified and for each teacher, a single data point was randomly selected. Consequently, the results of these analyses should not be interpreted as changes in teachers' attitudes, preparedness,

or practices, but rather as correlations between extent of participation in professional development and these outcomes.

## **Design Weights**

Although the number of targeted teachers in LSC projects varies greatly, with 41 percent targeting fewer than 500 teachers and 34 percent targeting more than 1,000 teachers, the core evaluation requires each project to administer questionnaires to a random sample of 300 teachers (or to the population if the project targets 350 or fewer teachers). Similarly, the size of the classroom observation sample is the same for each project regardless of the number of teachers targeted by that project. Prior to the 1999–2000 data collection year, each project was required to observe 10 lessons each year. Beginning with the 1999–2000 data collection year, projects were required to observe 16 lessons during their Baseline Year, Year Two, and Final Year.

Because of the unequal sampling probabilities, design weights are needed to estimate program-wide results. Without the use of design weights, all projects would have an equal impact on the results, regardless of the number of teachers targeted. Design weights are calculated by taking the inverse of the sampling probability. In addition, the design weights were then normalized so that the weighted number of cases was the same as the actual sample size. This procedure reduces the underestimation of standard errors, and thus the likelihood of Type I errors (incorrectly rejecting the null hypothesis, or in other words, saying a finding is significant when it really is not) that would otherwise occur.

## **Missing Data**

As with any large data collection, the LSC data sets contain missing data as a result of teacher non-response to various items on the questionnaire. There are a number of procedures for dealing with missing data. These procedures are (in order of least to most complex): listwise deletion, single imputation using mean substitution, single imputation using an estimation maximization algorithm, and multiple imputation.

Although multiple imputation may be the most appropriate technique for the LSC data, it is extremely computationally demanding, especially given the number of cases and variables in the data set. Initial trials indicate that implementing the multiple imputation procedure would take several months utilizing a top-of-the-line desktop computer. Thus, it was decided to try a single imputation using estimation maximization to examine the impact of missing data on the results. The results given for both the imputed and unimputed data sets (i.e., listwise deletion) were essentially equivalent—the regression coefficients for the imputed model were within  $\pm$  two standard errors of the coefficients for the unimputed model. Thus, the existence of missing data does not appear to affect the results of these models. For simplicity purposes, results from the unimputed data are shown in this report.

## Normality Procedures

Because the statistical procedures planned for the analyses assume multivariate normality, the continuous variables utilized in these models were assessed for deviations from normality. All were non-normally distributed, tending to display negative skewness and leptokurtosis. Various transformations were used on the variables in an attempt to normalize their distributions. The models were computed using the transformed variables and the results were compared to those generated when using the untransformed variables. The results were very similar across all subject/grade-ranges within each model, suggesting the effects of non-normality were minimal. To facilitate the interpretation of the results, the results of the model using untransformed variables are presented in this report.

## THE SAMPLE

### Model One

Model One utilizes over 20,000 questionnaires from the 88 LSC projects; the number of teachers and projects in each targeted subject/grade-range is shown in Table 1. The data used in these analyses reflect the number of projects funded in each targeted subject/grade-range.

**Table 1**  
**Teachers and Projects Included in Model One, by Subject/Grade-Range**

	<b>Number of Projects</b>	<b>Number of Teachers</b>	<b>Percent of Teachers</b>
K-8 Science	42	9,674	48
K-8 Mathematics	29	6,995	34
6-12 Mathematics	19	2,733	14
6-12 Science	7	695	4
<b>Total<sup>†</sup></b>	<b>85</b>	<b>20,097</b>	<b>100</b>

<sup>†</sup> The sum of projects is greater than the total as some projects target more than one subject/grade-range.

Table 2 shows the descriptive statistics for the continuous variables included in Model One by subject/grade-range prior to their transformation to standardized scores (z-scores). The distribution of responses, by subject/grade-range, for the only categorical variable in Model One, frequency of use of the LSC-designated instructional materials, is shown in Table 3. Roughly 8 out of 10 K-8 science and mathematics teachers use the LSC-designated instructional materials in their instruction; 7 out of 10 6-12 mathematics teachers and 6 out of 10 6-12 science teachers use the LSC-designated instructional materials in their teaching.

**Table 2**  
**Descriptive Statistics for Continuous Variables Included in Model One**

	Mean (standard deviation)			
	K-8 Science	K-8 Mathematics	6-12 Mathematics	6-12 Science
Number of hours of LSC professional development	34.43 (45.27)	42.42 (51.55)	46.37 (57.00)	25.98 (44.14)
Teacher perception of principal support composite	74.62 (14.30)	77.55 (14.29)	73.94 (13.60)	73.31 (14.79)
Attitudes toward reform-oriented teaching composite	86.24 (9.82)	87.60 (10.03)	81.62 (11.39)	85.15 (10.11)
Pedagogical preparedness composite	75.54 (13.85)	78.20 (13.58)	74.80 (13.10)	74.76 (13.38)
Content preparedness composite	59.11 (17.04)	72.29 (17.54)	73.70 (14.86)	78.97 (15.53)
Traditional teaching practices composite	51.34 (18.21)	73.37 (16.00)	76.15 (11.20)	70.27 (12.42)
Investigative culture composite	77.46 (14.87)	82.05 (13.75)	80.98 (12.47)	77.25 (11.87)
Investigative practices composite	54.00 (14.14)	52.99 (14.85)	47.27 (14.05)	52.77 (12.22)

**Table 3**  
**Descriptive Statistics for Categorical Variables Included in Model One:**  
**Frequency of Use of LSC-Designated Instructional Materials**

	Percent of Teachers			
	K-8 Science	K-8 Mathematics	6-12 Mathematics	6-12 Science
Never	20	23	29	39
Rarely	13	9	14	18
Sometimes	19	19	16	19
Often	21	23	13	16
All or Almost All the Time	27	27	28	8

## Model Two

As can be seen in Table 4, fewer teachers were utilized in Model Two than in Model One, due to the smaller quantity of observational data. The number of cases from 6-12 science projects was too small for including in this model.

**Table 4**  
**Teachers and Projects Included in Model Two, by Subject/Grade-Range**

	Number of Projects	Number of Teachers	Percent of Teachers
K-8 Science	41	450	43
K-8 Mathematics	29	368	35
6-12 Mathematics	19	231	22
<b>Total<sup>†</sup></b>	<b>79</b>	<b>1,049</b>	<b>100</b>

<sup>†</sup> The sum of projects is greater than the total as some projects target more than one subject/grade-range.

Table 5 shows the descriptive statistics for the continuous variables included in Model Two, by subject/grade-range, again before z-transformations were applied to continuous variables. With the exception of hours of professional development, the data are very similar to those utilized in Model One. The higher levels of professional development for teachers included in Model Two (i.e., those teachers that were observed) may be an indication of selection bias in the sample. Since both the questionnaire and observation samples were randomly drawn, this difference may be due to teachers with lower treatment levels declining to be observed at a higher rate than teachers with greater treatment levels. Thus, the results of Model Two may not be representative of all teachers in the LSC districts, but rather of those teachers who participated the most in LSC professional development.

**Table 5**  
**Descriptive Statistics for Continuous Variables Included in Model Two**

	Mean (standard deviation)		
	K-8 Science	K-8 Mathematics	6-12 Mathematics
Number of hours of LSC professional development	46.07 (52.24)	52.77 (56.68)	74.87 (69.83)
Teacher perception of principal support composite	75.07 (13.47)	77.82 (13.48)	73.73 (14.37)
Attitudes toward reform-oriented teaching composite	86.67 (9.38)	87.25 (10.22)	81.91 (12.31)
Pedagogical preparedness composite	76.48 (12.98)	77.88 (13.71)	76.46 (12.23)
Content preparedness composite	59.10 (17.01)	71.76 (17.40)	75.94 (13.03)
Traditional teaching practices composite	50.50 (17.87)	72.72 (15.65)	75.35 (10.18)
Investigative culture composite	78.99 (13.18)	82.36 (13.17)	82.71 (12.05)
Investigative practices composite	54.48 (13.52)	51.48 (13.73)	48.03 (14.31)

Table 6 shows the frequencies for each level of the categorical variables included in Model Two by subject/grade-range. The frequency distribution of use of the LSC-designated instructional materials is roughly the same as in Model One. About two-thirds of K-8 science lessons and about half of mathematics lessons, regardless of grade-range, are based upon the LSC-designated instructional materials. Although the Classroom Observation Protocol asks evaluators to rate lessons on a seven-point scale, ratings on that scale are not well distributed. Thus, for these analyses, evaluator ratings of lesson quality were collapsed to a 3-point scale of low, medium and high quality. Ratings on this three-point scale are fairly evenly distributed among the three levels.

**Table 6**  
**Descriptive Statistics for Categorical Variables Included in Model Two**

	Percent of Teachers		
	K-8 Science	K-8 Mathematics	6-12 Mathematics
<b>Frequency of Use of LSC-Designated Instructional Materials</b>			
Never	14	20	19
Rarely	12	8	17
Sometimes	17	15	15
Often	26	27	15
All or Almost All the Time	30	30	35
<b>Observed Lesson Based upon LSC-Designated Instructional Materials</b>			
No	37	52	49
Yes	63	48	51
<b>Quality Rating of Observed Lesson</b>			
Low	34	31	34
Medium	32	39	42
High	34	30	25

## RESULTS

### Model One

Model One examines the relationships among a number of key indicators from the teacher questionnaire. Results are presented for each subject/grade-range targeted by the LSC program.

#### *K-8 Science*

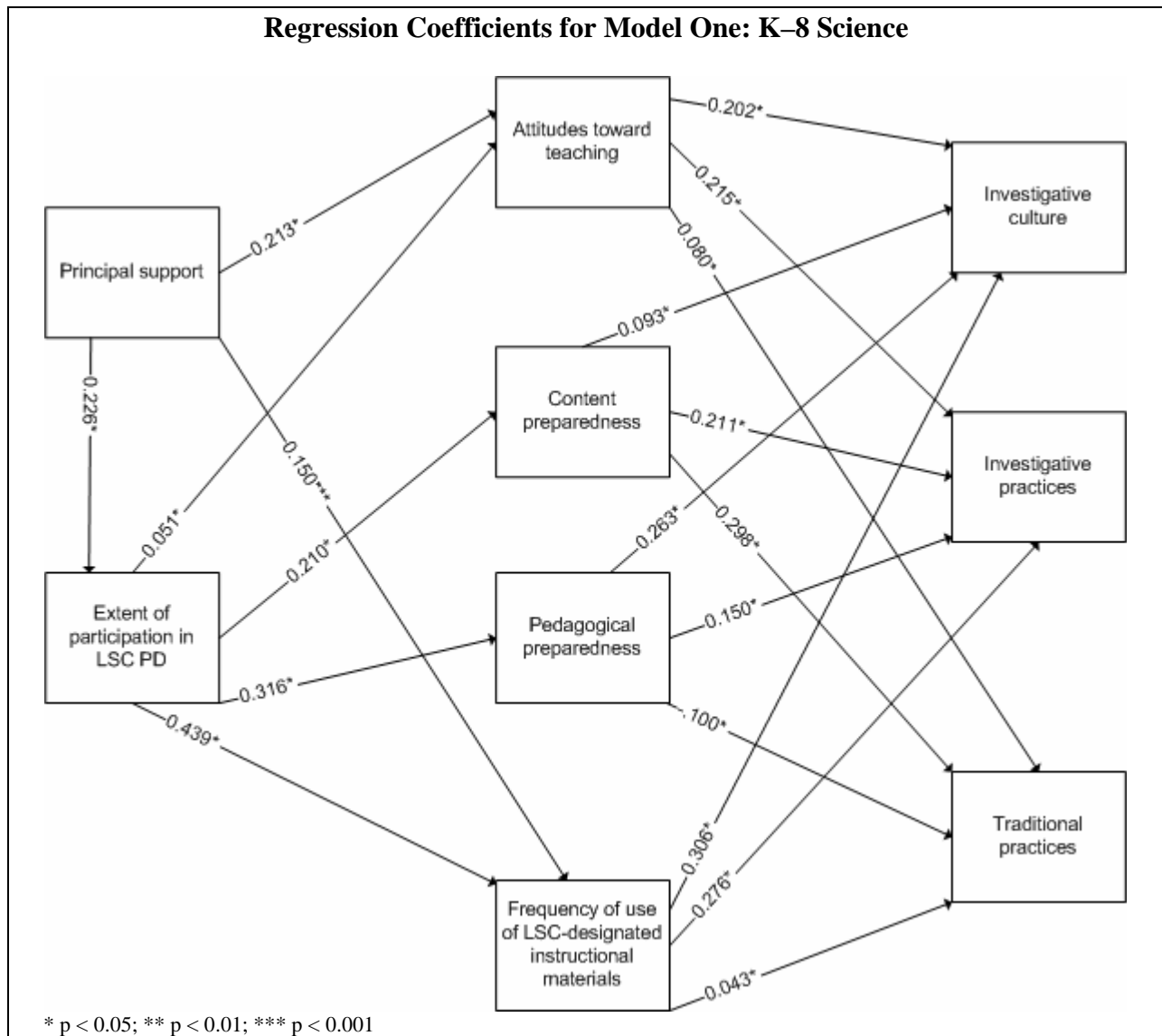
Figure 3 shows the standardized regression coefficients for K-8 science teachers. All regression paths in the model are statistically significant and some are relatively substantial. Participation in LSC professional development is a moderate predictor of teacher content preparedness, pedagogical preparedness, and use of LSC materials (effect sizes<sup>5</sup> of 0.21, 0.32, and 0.44 standard deviations, respectively). The teachers' perceptions of principal support composite is moderately and positively related to teachers' involvement in professional development (effect size of 0.23 standard deviations) and is similarly predictive of teacher's attitudes toward reform-based teaching (effect size of 0.21 standard deviations).

Composite scores of teachers' perceptions of their pedagogical preparedness are more strongly indicative of scores on the investigative culture composite (an effect size of 0.26 standard deviations) than scores on the investigative practice composite (an effect size of 0.15 standard deviations). The composite measuring teachers' attitudes toward reform-oriented teaching positively predicts scores on both the investigative culture and investigative practices composites, with effect sizes of 0.20 and 0.22 standard deviations, respectively. Interestingly, higher scores on the science content composite suggest a higher frequency of use of traditional teaching practices (an effect size of 0.30 standard deviations), perhaps reflecting a tendency for

<sup>5</sup> Following standard conventions, effect sizes of 0.2 are considered small effects, 0.5 medium effects, and 0.8 large effects (Jacob Cohen, *Statistical Power Analysis for the Behavior Sciences*, Hillsdale, NJ: Lawrence Erlbaum Associates, 1988).

elementary teachers who are better prepared in science content to spend more time teaching science.

Teachers' reports of the frequency at which they use the LSC-designated instructional materials is also a positive predictor of their reports of classroom practice. More frequent use of the LSC materials is associated with higher composite scores on the investigative practices and investigative culture composites (effect sizes of 0.31 and 0.28, respectively).



**Figure 3**

In addition to looking at the direct relationships between pairs of variables, SEM allows one to combine direct and indirect “effects” (i.e., those mediated by other variables) to get the “total”

effect of one variable on another.<sup>6</sup> Table 7 shows the total effects of a number of variables, including teachers’ participation in LSC professional development, teachers’ perception of principal support, and teachers’ attitudes toward reform-oriented teaching on the three teaching practice composites. The extent of teacher participation in LSC professional development, as mediated by their attitudes, perceptions of preparedness, and use of LSC materials, is a moderate predictor of teachers’ use of investigative practices and investigative culture (effect sizes of 0.22 and 0.25 standard deviations, respectively). In contrast, teachers’ perception of principal support is a relatively weak predictor of teachers’ scores on these composites (effect sizes of 0.15, and 0.14 standard deviations, respectively).

**Table 7**  
**Standardized Total Effects, by Outcome Variable: Model One K–8 Science**

	Total Effect		
	Investigative Culture	Investigative Practices	Traditional Practices
Frequency of use of LSC-designated instructional materials	0.31*	0.28*	0.04*
Pedagogical preparedness composite	0.26*	0.15*	-0.10*
Number of hours of LSC professional development	0.25*	0.22*	0.05*
Attitudes toward reform-oriented teaching composite	0.20*	0.22*	0.08*
Teacher perceptions of principal support composite	0.15*	0.14*	0.04*
Content preparedness composite	0.09*	0.21*	0.30*

\* p < 0.05

The largest predictors of the investigative culture and investigative practices composites are the extent of teachers’ participation in LSC professional development, their frequency of use of the LSC-designated instructional materials, and scores on the attitudes toward reform-oriented teaching and pedagogical preparedness composites. The content preparedness composite is a moderate predictor of the investigative practices composite and is the strongest predictor of scores on the traditional practices composite.

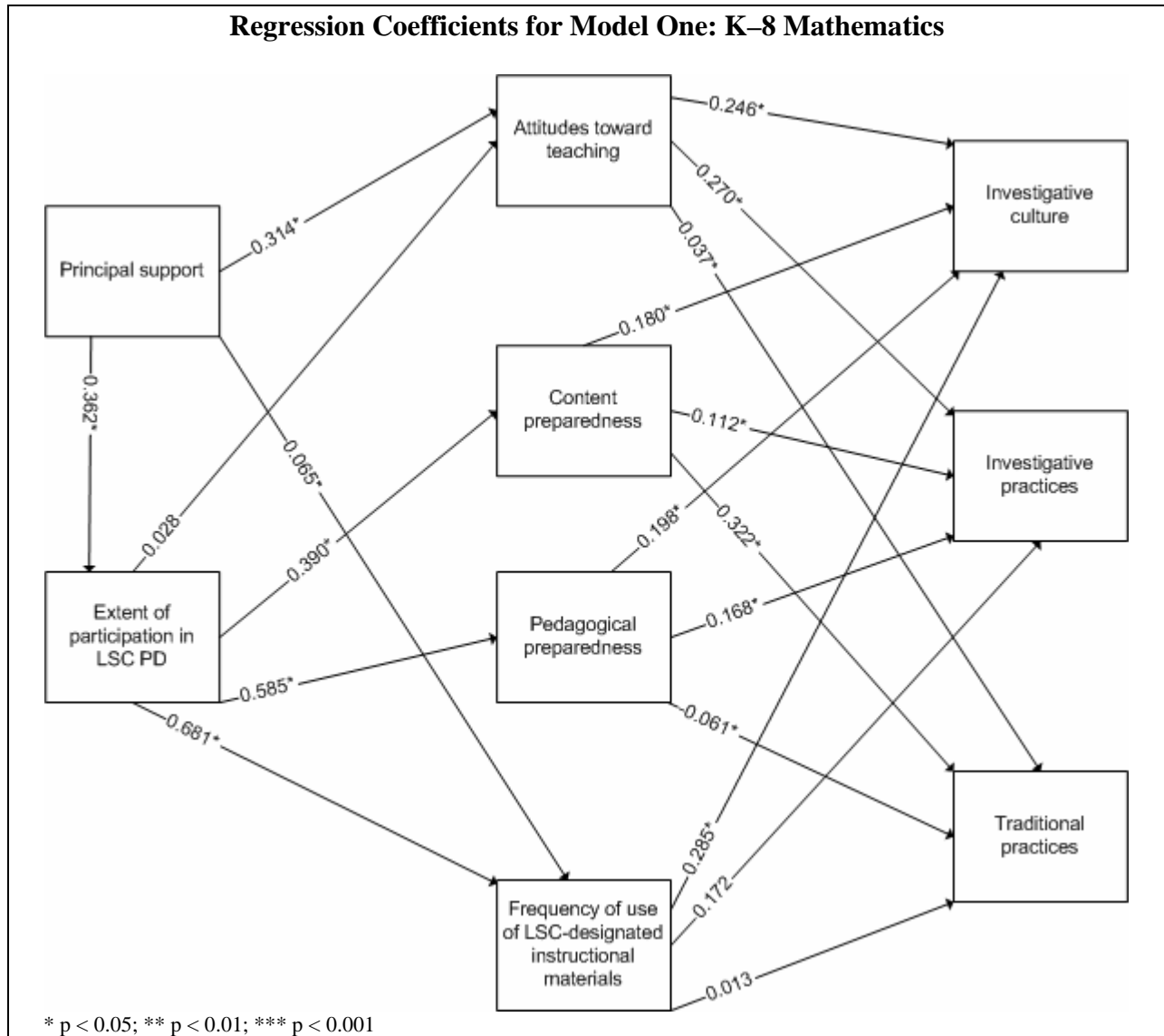
### ***K–8 Mathematics***

Figure 4 shows the standardized regression coefficients for K–8 mathematics teachers. Findings from this analysis suggest that participation in LSC professional development is positively and strongly related to K–8 mathematics teachers’ frequency of use of LSC materials and pedagogical preparedness (effect sizes of 0.68 and 0.59 standard deviations, respectively), and moderately predictive of teachers’ content preparedness (an effect size of 0.39 standard deviations). The teachers’ perceptions of principal support composite is moderately correlated with the extent of teachers’ involvement in professional development and attitudes toward reform-oriented teaching (effect size of 0.36 and 0.31 standard deviations, respectively).

The composite measuring teachers’ perceptions of their pedagogical preparedness is weakly related to their scores on the investigative culture and investigative practice composites (effect sizes of 0.20 and 0.17 standard deviations, respectively). The attitudes toward reform-oriented teaching composite moderately predicts scores for the investigative culture and investigative

<sup>6</sup> It should be noted that the term “effect,” a convention used in SEM, does not imply causality.

practice composites with effect sizes of 0.25 and 0.27 standard deviations, respectively. Interestingly, scores on the mathematics content composite more strongly predicts teachers' use of traditional teaching practices (an effect size of 0.32 standard deviations) than their use of investigative practices (an effect size of 0.11 standard deviations) or creation of an investigative culture (an effect size of 0.18 standard deviations). The use of the LSC-designated instructional materials is a moderate predictor of scores on the investigative culture composite (an effect size of 0.29 standard deviations).



**Figure 4**

Total effects, shown in Table 8, indicate that the extent of teacher participation in LSC professional development is positively related to teachers' scores on the investigative practices and investigative culture composites (effect sizes of 0.27 and 0.39 standard deviations,

respectively). The teacher perceptions of principal support composite is also associated positively with scores on the investigative culture and investigative practice composites (effect sizes of 0.24 and 0.19 standard deviations, respectively).

**Table 8**  
**Standardized Total Effects: Model One K–8 Mathematics**

	Total Effect		
	Investigative Culture	Investigative Practices	Traditional Practices
Number of hours of LSC professional development	0.38*	0.27*	0.10*
Frequency of use of LSC-designated instructional materials	0.29*	0.17*	0.01
Attitudes toward reform-oriented teaching composite	0.25*	0.27*	0.04*
Teacher perceptions of principal support composite	0.24*	0.19*	0.05*
Pedagogical preparedness composite	0.20*	0.17*	-0.06*
Content preparedness composite	0.18*	0.11*	0.32*

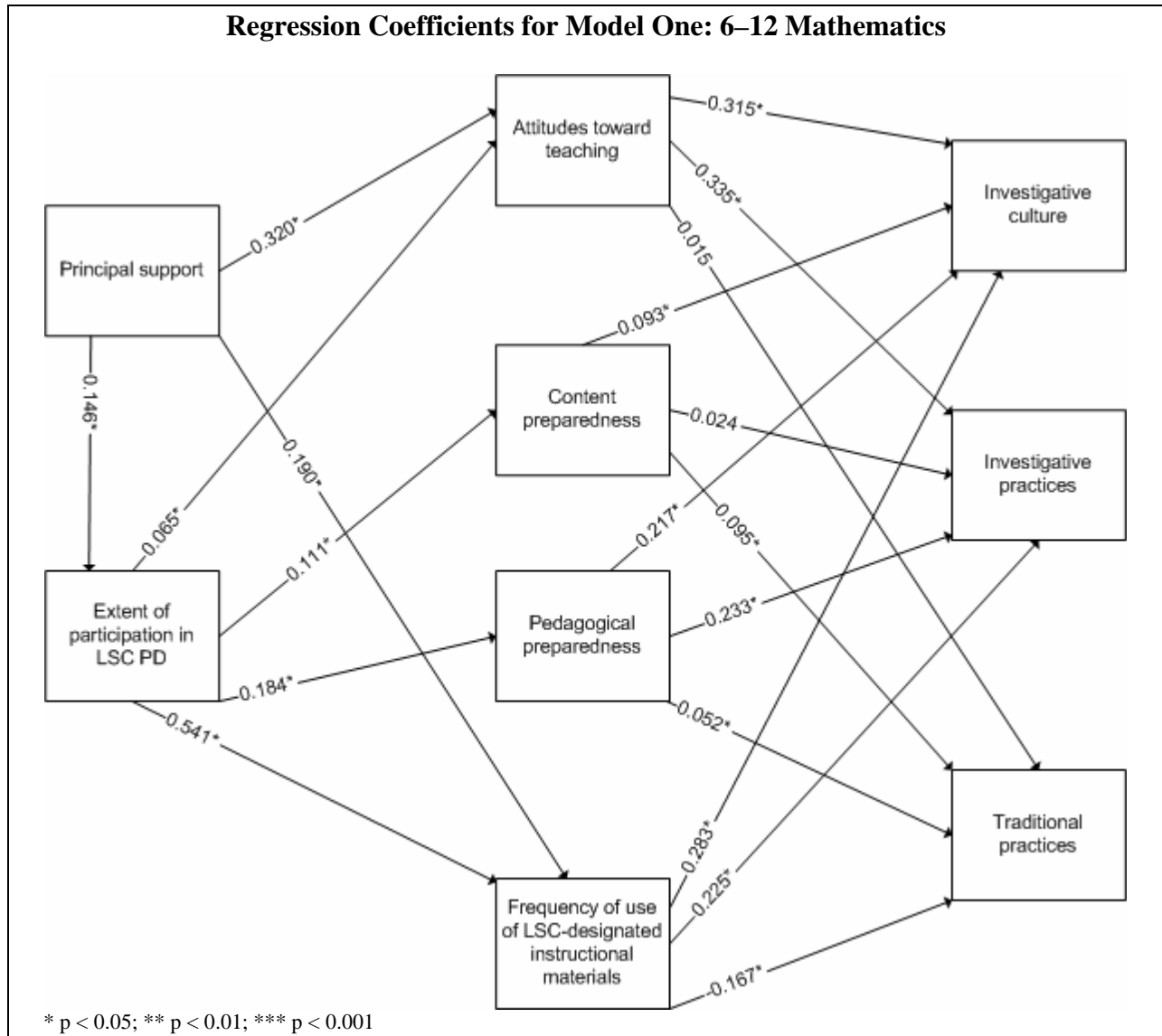
\*  $p < 0.05$

The largest predictors of the investigative culture and investigative practices composites are the extent of teachers’ participation in LSC professional development, their frequency of use of the LSC-designated instructional materials, and scores on the attitudes toward reform-oriented teaching composite. The strongest predictor of scores on the traditional practices composite is the composite measuring teachers’ perceptions of their content preparedness.

### ***6–12 Mathematics***

Participation in LSC professional development is strongly and positively related to 6–12 mathematics teachers’ use of LSC materials (effect size of 0.54 standard deviations), but is only weakly predictive of teachers’ attitudes toward reform-oriented teaching, content preparedness, and pedagogical preparedness (effect sizes of 0.07, 0.11, and 0.18 standard deviations, respectively). The teachers’ perceptions of principal support composite is a moderate predictor of teachers’ attitudes toward reform-based teaching (effect size of 0.32 standard deviations) and a weak predictor of teachers’ involvement in professional development and use of the LSC-designated instructional materials (effect sizes of 0.15 and 0.19 standard deviations, respectively).

The teachers’ perceptions of their pedagogical preparedness composite is positively related to their scores on the investigative culture and investigative practice composites (effect size of 0.22 and 0.23 standard deviations, respectively). The attitudes toward reform-oriented teaching composite is moderately correlated with scores on the investigative culture and investigative practices composites with effect size of 0.32 and 0.34 standard deviations, respectively. The frequency of use of the LSC-designated instructional materials is a moderate predictor of scores on the investigative practices and investigative culture composites (effect sizes of 0.23 and 0.28, respectively).



**Figure 5**

Total effects calculated as part of this analysis indicate that teacher participation in LSC professional development is a positive predictor of the investigative practices and investigative culture composites (effect sizes of 0.19 and 0.22 standard deviations, respectively). The teachers' perception of principal support composite is predictive of the investigative practices and investigative culture composites (effect sizes of 0.18 and 0.19 standard deviations, respectively).

**Table 9**  
**Standardized Total Effects: Model One 6–12 Mathematics**

	Total Effect		
	Investigative Culture	Investigative Practices	Traditional Practices
Attitudes toward reform-oriented teaching composite	0.32*	0.34*	0.02
Frequency of use of LSC-designated instructional materials	0.28*	0.23*	-0.17*
Pedagogical preparedness composite	0.22*	0.23*	0.05*
Number of hours of LSC professional development	0.22*	0.19*	-0.07*
Teacher perceptions of principal support composite	0.19*	0.18*	-0.04*
Content preparedness composite	0.09*	0.02	0.10*

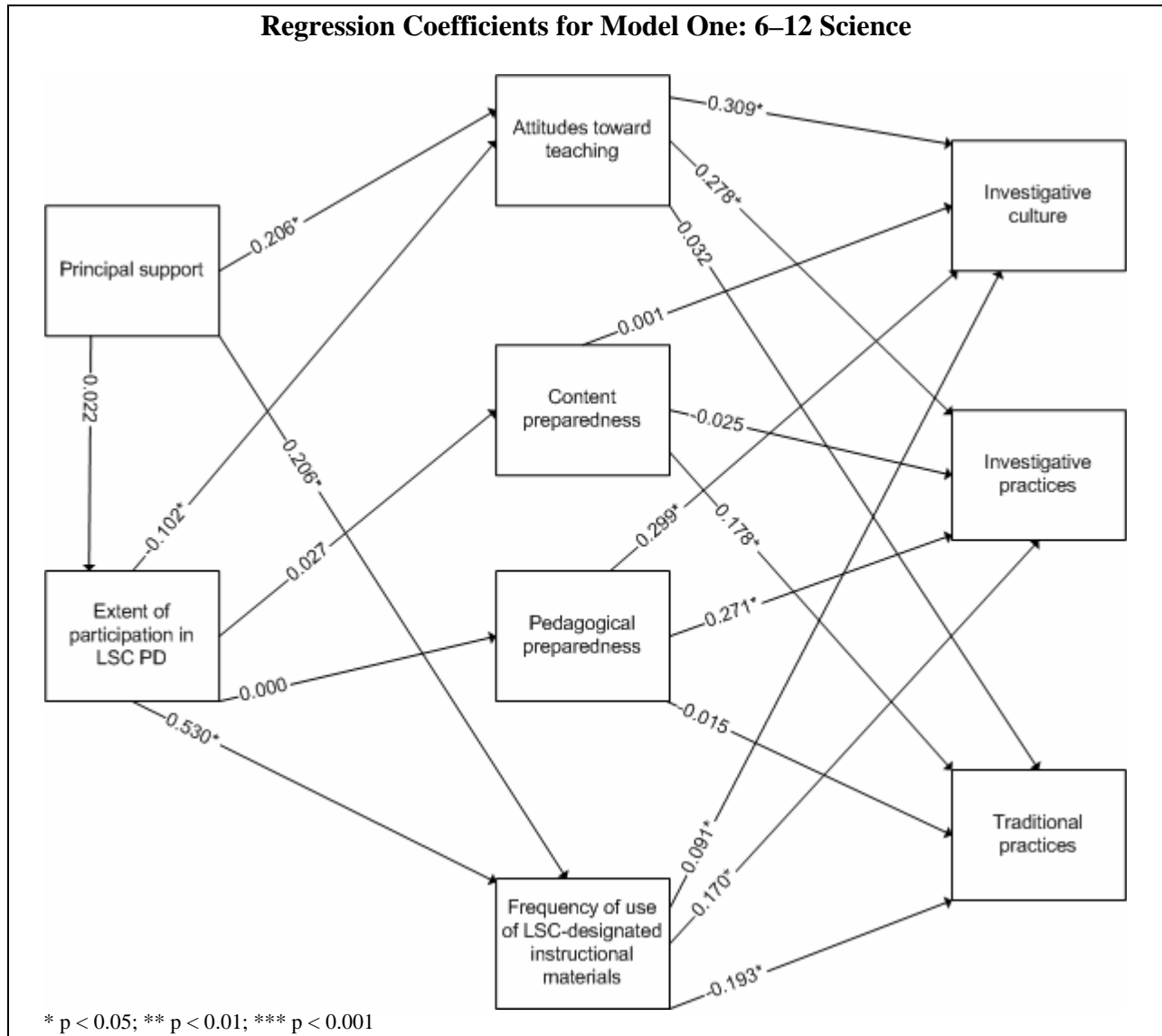
\*  $p < 0.05$

The largest predictors of the investigative culture and investigative practices composites are teachers' scores on the attitudes toward reform-oriented teaching composite, frequency of use of the LSC-designated instructional materials, extent of participation in LSC professional development, and scores on the pedagogical preparedness composite. The frequency of use of the LSC materials is the strongest predictor of teachers' scores on the traditional practices composite, with greater use of the materials associated with lower scores on this composite. The strongest positive predictor of traditional practices is the composite measuring teachers' perceptions of their content preparedness, although the relationship is quite weak.

### ***6–12 Science***

Figure 6 shows the standardized regression coefficients for the analysis of questionnaires completed by 6–12 science teachers. Regression coefficients in this model indicate that participation in LSC professional development is a strong positive indicator of teachers' use of LSC materials (effect size of 0.53 standard deviations). The extent of teachers' participation in LSC professional development is not related to their pedagogical preparedness or content preparedness. The composite measuring teachers' perceptions of principal support is a moderate predictor of teachers' attitudes toward reform-oriented teaching and their use of the LSC-designated instructional materials (effect sizes of 0.21 standard deviations).

Teachers' perceptions of their pedagogical preparedness are moderately predictive of their scores on the investigative culture composite and investigative practice composite (effect sizes of 0.30 and 0.27 standard deviations, respectively). The attitudes toward reform-oriented teaching composite is moderately correlated with scores on the investigative culture and investigative practices composites, with effect sizes of 0.31 and 0.28 standard deviations, respectively. The composite measuring teachers' perceptions of their content preparedness is a significant predictor of their use of traditional teaching practices (an effect size of 0.18 standard deviations). The frequency of use of the LSC-designated instructional materials is a significant but weak predictor of all three teaching practices composites.



**Figure 6**

The total effects for the 6–12 science analyses are shown in Table 10. Participation in LSC professional development is not significantly related to teacher use of investigative practices and investigative culture, though it is negatively, albeit weakly, related to their use of traditional practices (an effect size of -0.10 standard deviations). The teacher perceptions of principal support composite is a weak positive predictor of scores on the investigative culture composite, and a weak negative predictor of scores on the traditional practices composite (effect sizes of 0.08 and -0.04 standard deviations, respectively).

**Table 10**  
**Standardized Total Effects: Model One 6–12 Science**

	Total Effect		
	Investigative Culture	Investigative Practices	Traditional Practices
Attitudes toward reform-oriented teaching composite	0.31*	0.28*	0.03
Pedagogical preparedness composite	0.30*	0.27*	-0.02
Frequency of use of LSC-designated instructional materials	0.09*	0.17*	-0.19*
Teacher perceptions of principal support composite	0.08*	0.09	-0.04*
Number of hours of LSC professional development	0.02	0.06	-0.10*
Content preparedness composite	0.00	-0.03	0.18*

\*  $p < 0.05$

Interestingly, the strongest predictors of the investigative culture and investigative practices composites are teachers’ attitudes towards reform-oriented teaching and their perceptions of pedagogical preparedness. The strongest predictors of scores on the traditional practices composite are the use of the LSC-designated instructional materials (negatively correlated with traditional practices) and perceptions of content preparedness (positively correlated with traditional practices).

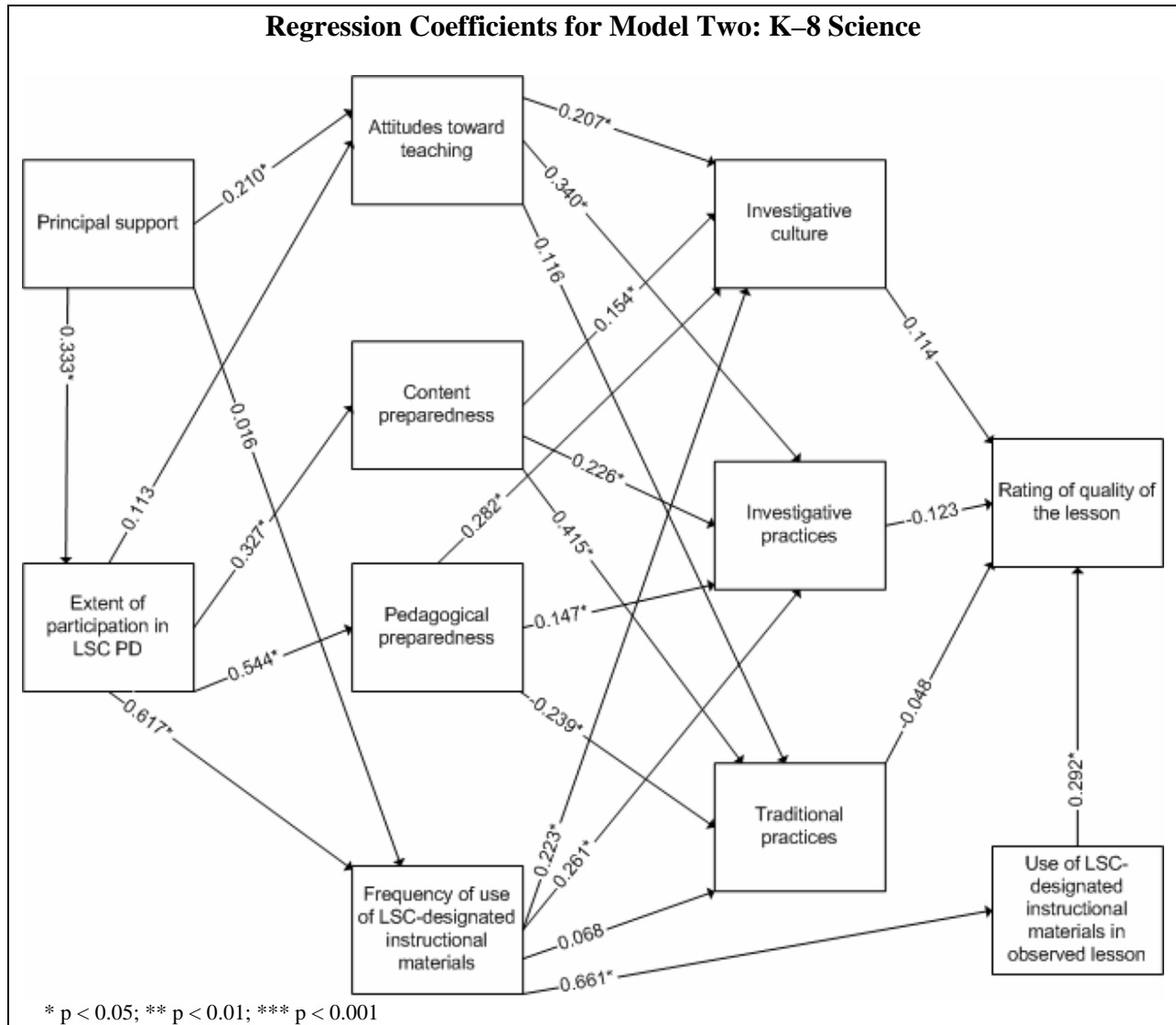
## Model Two

Model Two builds on Model One by adding data collected through classroom observations: whether the lesson was based on the LSC-designated instructional materials and the evaluator rating of lesson quality. Because of the smaller number of classroom observations required by the core evaluation, the analyses for Model Two are based upon a much smaller sample of teachers than in Model One. Again, results are reported for each targeted subject/grade-range.

### *K–8 Science*

Figure 7 shows the standardized regression coefficients for K–8 science teachers. Most of the regression coefficients for the paths that were included in Model One are similar in magnitude in Model Two. As mentioned previously, although the subset used in Model Two is similar to the full data set used in Model One on most characteristics, teachers in the subset tended to have higher levels of participation in LSC professional development than those in the full data set. Thus, the differences in the regression coefficients may be due to a selection bias.

None of the teaching practice composites is a significant predictor of lesson quality. In other words, the lessons of teachers who indicated they used various classroom practices more frequently were no more likely to be rated highly than the lessons of teachers who used those practices less frequently (or vice-versa). In contrast, the use of the LSC-designated instructional materials was a significant predictor of lesson quality with an effect size of 0.29. Teachers’ responses to the questionnaire item asking about their frequency of use of the LSC materials is a strong predictor of whether the teacher used those materials in the observed lesson (an effect size of 0.66), an indicator of the validity of this teacher questionnaire item.



**Figure 7**

The total effects shown in Table 11 indicate that the extent of teacher participation in LSC professional development is a weak, but positive, predictor of lesson ratings (an effect size of 0.11). Questionnaire responses about the frequency of use of LSC-designated instructional materials are somewhat predictive of lesson quality, with an effect size of 0.18. The strongest predictor of lesson quality is whether the lesson was based upon the LSC materials (an effect size of 0.29).

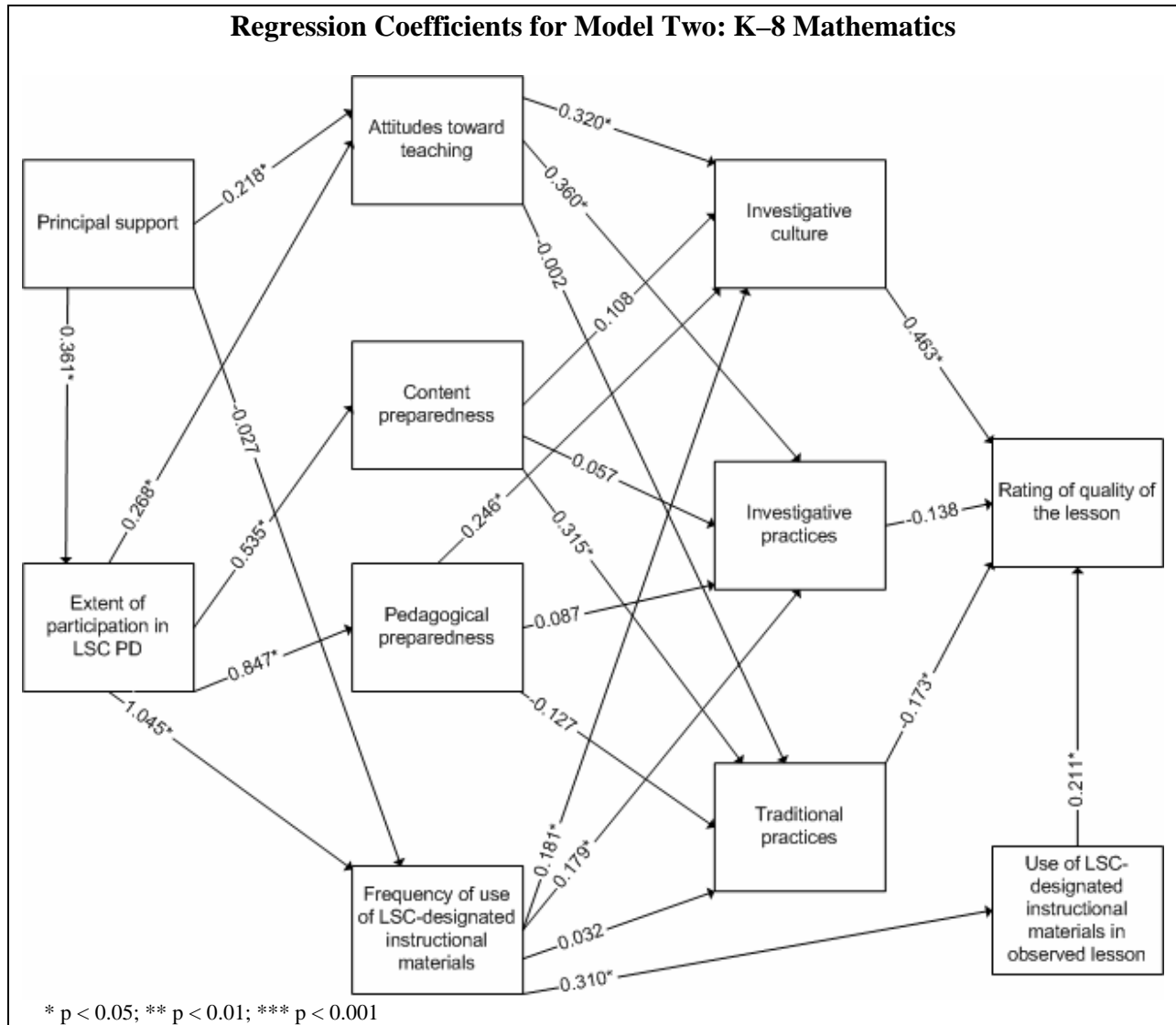
**Table 11**  
**Standardized Total Effects: Model Two K–8 Science**

	<b>Capsule Rating</b>
Use of LSC-designated instructional materials in observed lesson	0.29*
Frequency of use of LSC-designated instructional materials	0.18*
Number of hours of LSC professional development	0.11*
Investigative culture composite	0.11
Teacher perceptions of principal support composite	0.04
Pedagogical preparedness composite	0.03
Attitudes toward reform-oriented teaching composite	-0.02
Content preparedness composite	-0.03
Traditional teaching practices composite	-0.05
Investigative practices composite	-0.12

\*  $p < 0.05$

***K–8 Mathematics***

The results for K–8 mathematics are somewhat different than those for K–8 science. As can be seen in the Figure 8, the investigative culture composite is a significant and positive predictor of lesson quality (an effect size of 0.46). In addition, the traditional teaching practices composite is negatively related to lesson quality (an effect size of -0.17). As was the case in K–8 science, K–8 mathematics lessons that were based on the LSC-designated instructional materials tended to be rated higher than those that were not based on those materials.



**Figure 8**

Although the investigative culture composite has the strongest relationship, a number of other variables are significant predictors of lesson quality, including the extent of teacher participation in LSC professional development and whether the lesson was based on the LSC-designated instructional materials (effect sizes of 0.25 and 0.21, respectively). The frequency of use of the LSC materials in general, along with the pedagogical preparedness, perceptions of principal support, and attitudes toward reform-oriented teacher composites are weak predictors of lesson quality with effect sizes around 0.10.

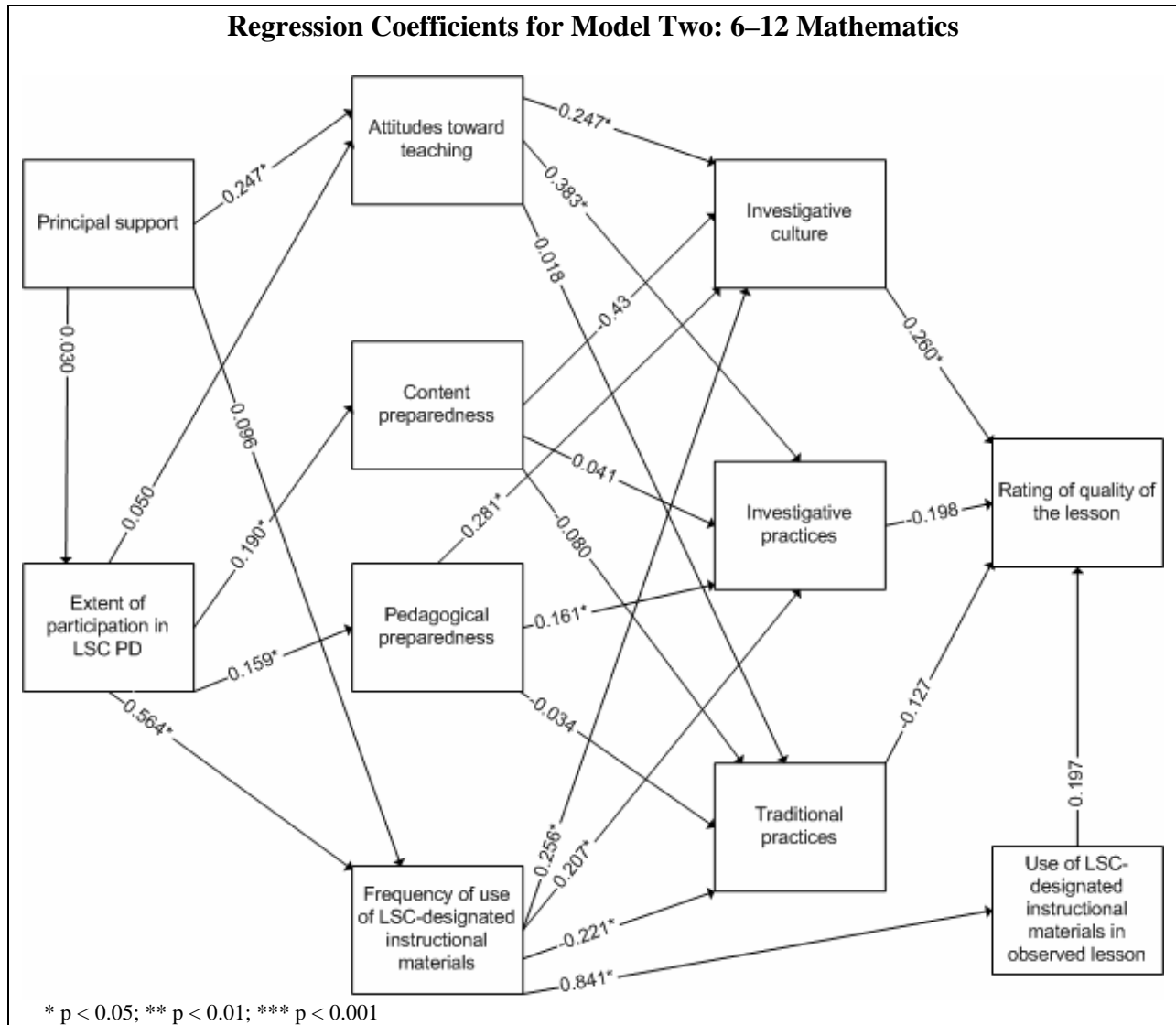
**Table 12**  
**Standardized Total Effects: Model Two K–8 Mathematics**

	<b>Capsule Rating</b>
Investigative culture composite	0.46*
Number of hours of LSC professional development	0.25*
Use of LSC-designated instructional materials in observed lesson	0.21*
Frequency of use of LSC-designated instructional materials	0.12*
Pedagogical preparedness composite	0.12*
Teacher perceptions of principal support composite	0.11*
Attitudes toward reform-oriented teaching composite	0.10*
Content preparedness composite	-0.01
Investigative practices composite	-0.14
Traditional teaching practices composite	-0.17*

\*  $p < 0.05$

***6–12 Mathematics***

As was the case for K–8 mathematics, the investigative culture composite is a significant predictor of quality in 6–12 mathematics lessons with an effect size of 0.26. None of the other direct effects are significant, though this result may be due to the relatively small sample size available for this analysis ( $N = 231$ ).



**Figure 9**

In addition to the investigative culture composite, the frequency of use of LSC-designated instructional materials and the extent of teacher participation in LSC professional development are significant predictors of lesson quality (effect sizes of 0.22 and 0.13, respectively). None of the other variables was a significant predictor of lesson quality.

**Table 13**  
**Standardized Total Effects: Model Two 6–12 Mathematics**

	<b>Capsule Rating</b>
Investigative culture composite	0.26*
Frequency of use of LSC-designated instructional materials	0.22*
Use of LSC-designated instructional materials in observed lesson	0.20
Number of hours of LSC professional development	0.13*
Pedagogical preparedness composite	0.05
Teacher perceptions of principal support composite	0.02
Attitudes toward reform-oriented teaching composite	-0.01
Content preparedness composite	-0.01
Traditional teaching practices composite	-0.13
Investigative practices composite	-0.20

\*  $p < 0.05$

## CONCLUSIONS

In this study, several structural models based on the underlying logic of the LSC program were tested. Although this study is not longitudinal, the number of statistically significant pathways in the models supports the underlying logic of the program. Important relationships demonstrated through the study include that greater amounts of LSC professional development are related to greater content and pedagogical preparedness and use of LSC-designated instructional materials, which in turn are related to greater use of standards-based teaching practices. Also, the use of LSC-designated instructional materials has a moderate, positive relationship with ratings of the quality of observed lessons. Table 14 provides a summary of the findings from these analyses.

**Table 14**  
**Summary of Results**

	Total Effect			
	Investigative Culture	Investigative Practices	Traditional Practices	Capsule Rating
<b>Number of hours of LSC professional development</b>				
K-8 Science	0.25*	0.22*	0.05*	0.11*
K-8 Mathematics	0.38*	0.27*	0.10*	0.25*
6-12 Mathematics	0.22*	0.19*	-0.07*	0.13*
6-12 Science	0.02	0.06	-0.10*	—
<b>Frequency of use of LSC-designated instructional materials</b>				
K-8 Science	0.31*	0.28*	0.04*	0.18*
K-8 Mathematics	0.29*	0.17*	0.01	0.12*
6-12 Mathematics	0.28*	0.23*	-0.17*	0.22*
6-12 Science	0.09*	0.17*	-0.19*	—
<b>Teacher perceptions of principal support composite</b>				
K-8 Science	0.15*	0.14*	0.04*	0.04
K-8 Mathematics	0.24*	0.19*	0.05*	0.11*
6-12 Mathematics	0.19*	0.18*	-0.04*	0.02
6-12 Science	0.08*	0.09	-0.04*	—
<b>Attitudes toward reform-oriented teaching composite</b>				
K-8 Science	0.20*	0.22*	0.08*	-0.02
K-8 Mathematics	0.25*	0.27*	0.04*	0.10*
6-12 Mathematics	0.32*	0.34*	0.02	-0.01
6-12 Science	0.31*	0.28*	0.03	—
<b>Pedagogical preparedness composite</b>				
K-8 Science	0.26*	0.15*	-0.10*	0.03
K-8 Mathematics	0.20*	0.17*	-0.06*	0.12*
6-12 Mathematics	0.22*	0.23*	0.05*	0.05
6-12 Science	0.30*	0.27*	-0.02	—
<b>Content preparedness composite</b>				
K-8 Science	0.09*	0.21*	0.30*	-0.03
K-8 Mathematics	0.18*	0.11*	0.32*	-0.01
6-12 Mathematics	0.09*	0.02	0.10*	-0.01
6-12 Science	0.00	-0.03	0.18*	—
<b>Use of LSC-designated instructional materials in observed lesson</b>				
K-8 Science	—	—	—	0.29*
K-8 Mathematics	—	—	—	0.21*
6-12 Mathematics	—	—	—	0.20
6-12 Science	—	—	—	—

\* p < 0.05

In K–8 science, K–8 mathematics, and 6–12 mathematics, there is a consistent relationship between the extent of participation in LSC professional development and the use of standards-based teaching practices as measured by the investigative practices and investigative culture composites. Hours of professional development also predicts, albeit weakly, evaluator ratings of the quality of observed lessons.

Teachers' reports of the frequency with which they use the LSC-designated instructional materials are also associated with higher scores on the investigative practices and culture composites and with higher evaluator ratings of the observed lesson quality. In K–8 science and mathematics projects, observed lesson quality was also related to whether that specific lesson was based on the LSC materials. Further, teacher responses to the questionnaire item regarding frequency of use of the materials was strongly correlated with observation data regarding materials use for all three subject/grade-range groups for which these data were available, an indication of the validity of the questionnaire item. In 6–12 science and mathematics science projects, greater frequency of use of the LSC materials is associated with lower scores on the traditional practices composite.

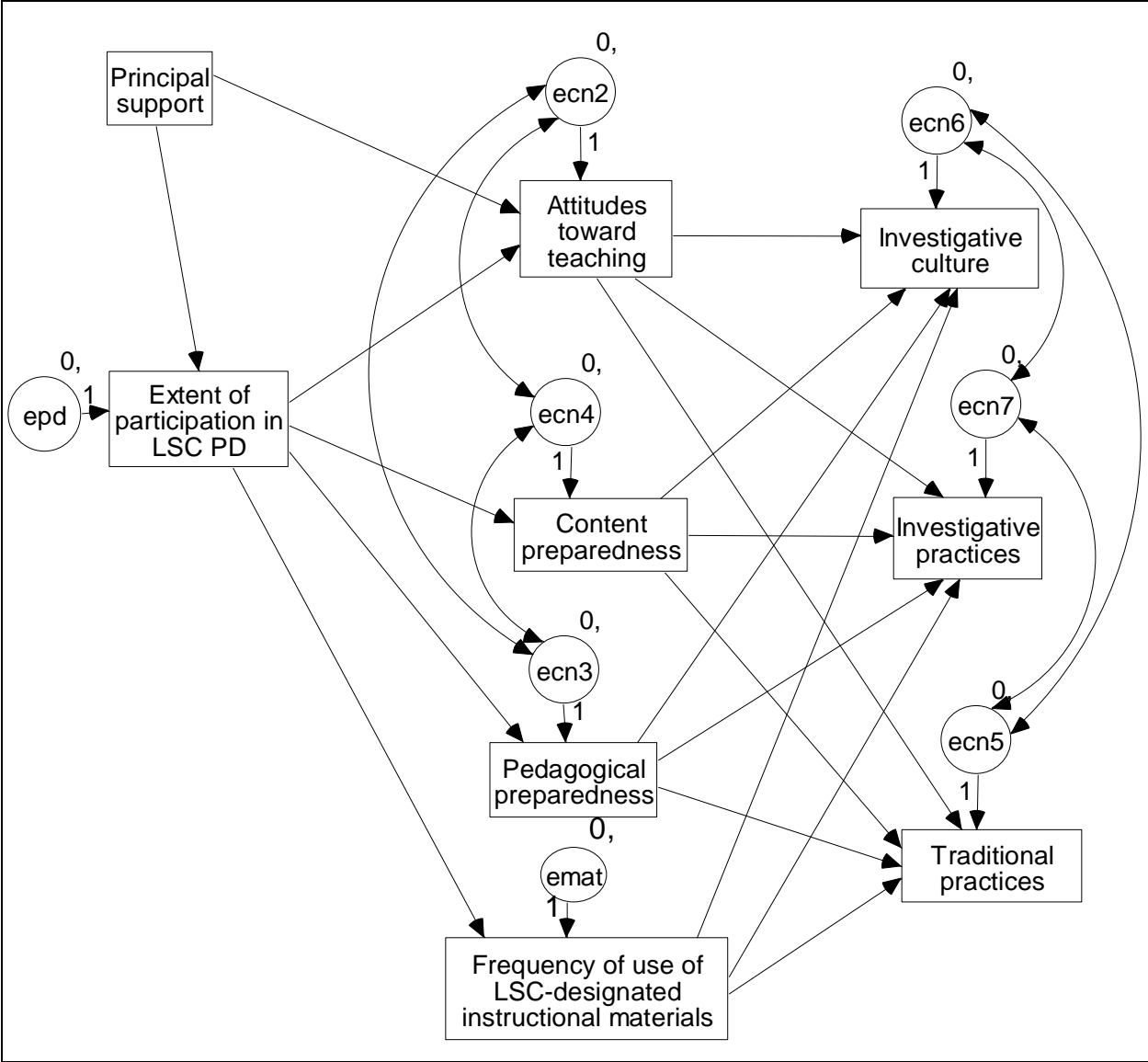
The composite measuring teachers' perceptions of principal support tends to be predictive of teachers' use of investigative classroom strategies. However, the perceptions of principal support composite is significantly related to the quality ratings of observed lessons only in K–8 mathematics projects.

Both the attitudes toward reform-oriented teaching and pedagogical preparedness composites are associated with higher scores on the investigative culture and investigative practices composites. The content preparedness composite is a consistent predictor of higher scores on the traditional practices composite. None of the attitudes or preparedness composites, nor the teaching practice composites, is a significant predictor of observed lesson quality.

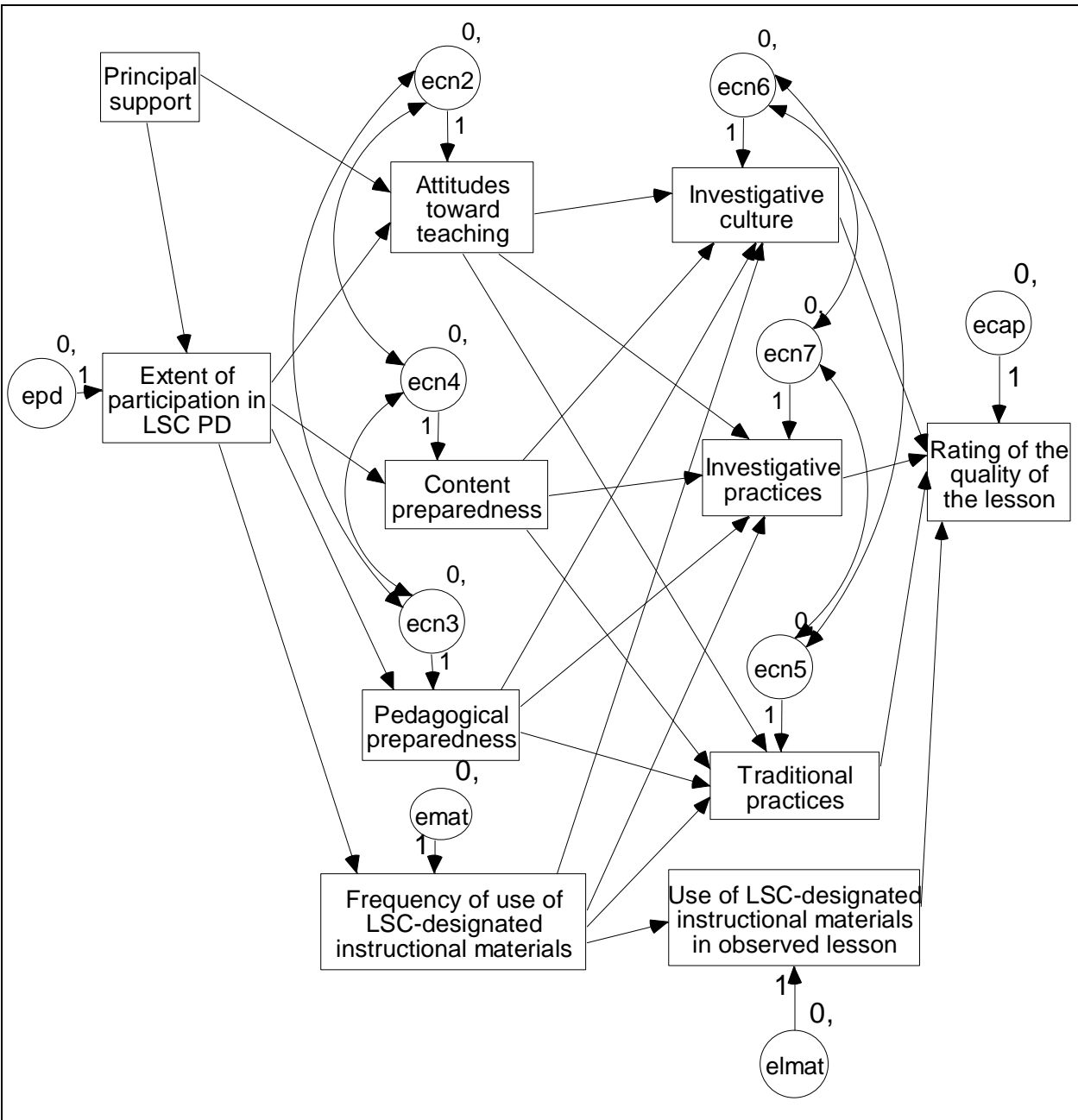
It is important to note that aside from the evaluator ratings of lesson quality, all of the data used in these analyses are self-reported by teachers. In addition, the classroom practice composites measure only the frequency of use of various practices, not their quality. Thus, results should be interpreted with some caution. It is also important to note that even though the LSC was intended to target all teachers in a jurisdiction, in practice teacher participation in the professional development tends to be voluntary, so there is a danger of selection bias in the sample (i.e., teachers who decide to participate may be the better teachers). Further, teachers who agree to participate in the classroom observations tend to have higher levels of professional development than teachers completing the questionnaire, indicating a potential selection bias affecting observational data. Finally, although path models such as structural equation modeling give the appearance of causality, they do not in fact test causality. These models illustrate correlations, not cause and effect.

## **APPENDIX**

### **Exact Statistical Models**



*Model One*



*Model Two*