Scoring Evolution and Creationism

The Academic Effects of Creationist Curricula in Public Schools

Scott Anderson
Undergraduate Honors Thesis
Sanford School of Public Policy
Duke University
Durham, NC

December, 2011

Contents

Contents 2
Abstract 3
Introduction 3
Historical Background 4
Theoretical Framework 6
Methods 9
Results and Analysis 19
Limitations 30
Conclusions 32
References 36
Acknowledgements: 37

Abstract

From 2002 to 2006, Cobb County, Georgia placed controversial stickers on high school biology textbooks which were critical of evolution. This study uses the differences-in-differences method to analyze what effects the sticker policy had on biology student test performance. Data collected from 2003 to 2010 suggest that the removal of the stickers in 2006 resulted in both a higher failure rate and a lower rate of pass plus scores in Cobb County schools. However, these results suggest that any effect the stickers had on student performance was minimal. This study concludes that students learn better when the material being learned relates to events outside of the classroom and are seen by students as controversial.

Introduction

Does placing a sticker on biology textbooks claiming that students should think critically about evolution and casts doubt the factuality of evolution affect understanding of evolution and performance in the science classroom? This study of suburban Atlanta suggests that the sticker has a small, positive effect on student performance when their knowledge of evolution is tested. Perhaps this is because students take more interest in those subjects which are seen to be controversial.

This paper will begin with a discussion of the history of evolution and creationism, specifically on how the debate has taken shape relative to what is taught in the classroom. Cobb County, Georgia represents a more recent chronicle in this history, with its decision to place stickers critical of evolution on its biology textbooks in 2002 and decision to remove those stickers in 2006 and will be the focus of this study.

This paper will then focus on previous studies relating to the debate surrounding evolution and creationism in curricula. While many studies look into student beliefs on this debate, no studies directly examine the relationship between the curriculum taught in the science classroom and the performance of those students. Previous studies do however suggest that those student exposure to a curriculum more favorable to evolution will perform better when tested on evolution. Next, this paper will discuss the method used to study the relationship between curriculum and student performance. This section primarily concerns itself with the use of the differences-in-differences statistical method and its application to the performance of students observed in this study.

A discussion of the results and analysis of those results will follow the explanation of the method. The focus of this section is the interpretation of the models used. Following the discussion of results is a series of limitation, which will discuss the aspects of the study which may cause doubt to the conclusions or provide an opportunity to make future studies stronger. The final section of this paper is a conclusions section describing what one can learn from the observed relationship between the sticker policy and student performance along with the policy implications for teachers and policy-makers who deal with this debate.

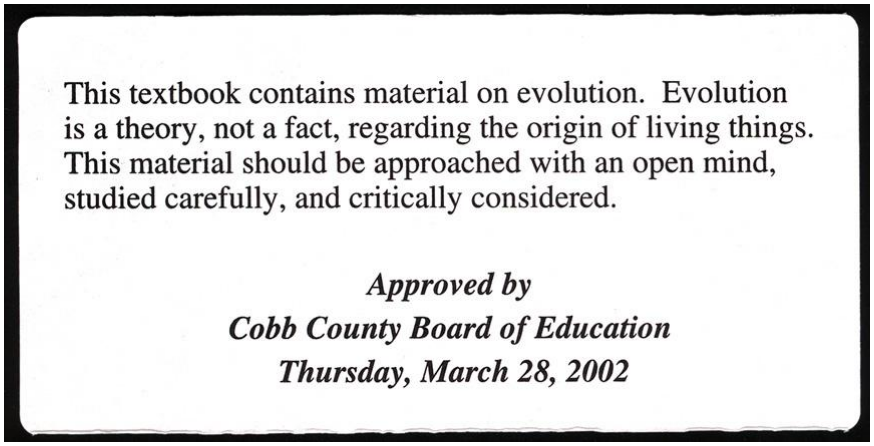
Historical Background

Since Charles Darwin introduced evolution as a theory in *On the Origin of Species* (Darwin 1859), traditional viewpoints on the formation of human life have come into question. As evolution established itself as scientific theory, questions developed as to how to present this theory to students in light of its apparent rejection of creationism. While many states were hesitant to allow evolution in the classroom, by the 1980s, evolution was consistently taught in biology classrooms. However, over the past 20 years, the debate raged on, as more than 15 states changed their curriculum regarding the subject over that time. Additionally, many areas, including Beebe, Arkansas, Rio Rancho, New Mexico, and Grantsburg, Wisconsin, deviated from the state standards regarding the teaching of evolution. Additionally, two states with somewhat creationist curricula have had districts within them deviate from state policy, teaching curricula emphasizing evolution in ways not allowed for in the state standards.

Previously, researchers rated states on how favorable their policies are towards the scientific consensus on evolution (Lerner 2000; Mead and Mates 2009). However, this study will look beyond the intricacies of policy and instead analyze how a policy relates to the performance of students in the classroom. While it is interesting to see what policies are taught by a variety of states, the debate between creationism and evolution is largely a theoretical one, affecting students' beliefs rather than their understanding of scientific concepts vital to their success beyond academia. Therefore, the goal of this study is to determine whether or not the policies that are taught to a student affect their knowledge of biology in a significant way, which could have more far reaching implications than simply understanding what the curricula taught are.

In 2002, 2,300 parents of Cobb County students petitioned against the teaching of evolution. As a result, the Cobb County Board of Education required the placement of stickers on textbooks describing evolution as "a theory, not a fact" and calling for students to approach evolution in a critical and careful manner. These stickers remained on biology textbooks until December, 2006, when the Cobb County Board of Education settled out of court with those who challenged the stickers on constitutional grounds. The sticker placed on the textbook can be seen in Figure 1.

Figure 1



This textbook contains material on evolution. Evolution is a theory, not a fact, regarding the origin of living things. This material should be approached with an open mind, studied carefully, and critically considered.

Approved by
Cobb County Board of Education
Thursday, March 28, 2002

While this sticker does not make any patently false claims about evolution, it relies on a general ambiguity between the scientific and colloquial senses of the word "theory". Scientists use the word "theory" to indicate that something is accepted as true even though it cannot be proven. However, the coloquial sense of the word "theory" suggests that evolution is untested speculation. The sticker and the creationists behind its inception are therefore using the term theory to draw doubt on what is commonly accepted by scientists (Isaak 1995).

Theoretical Framework

The central question of this study is "How do the policies taken by individual school districts regarding the teaching of creationism or evolution affect student understanding of science?"

There are two major hypotheses that could be true about the relationship between the sticker policy used to question evolution and the performance of students. The first hypothesis would be that having stickers on the biology textbooks would call into question the validity of evolution. This would lead students to question what they were being taught in class, not take that material seriously, and perform more poorly when examined at the end of the course. The other possible hypothesis would be that the stickers on the textbooks intrigue teenage students who are naturally drawn to controversy. The controversy stirred up by the stickers would inspire an interest by the students in evolution. Thus the students would pay more attention to that material in class, engage with the arguments for and against evolution, and perform better when their tested at the end of the course.

Studies have already shown an association between students' positive attitudes toward evolution and academic success in classes covering subjects relating to evolution (Ingram and Nelson 2006). Ingram and Nelson examined the relationships between the attitudes of a student towards evolution both prior to biology courses and following the biology course with the final grade that the student received in the course. This study found that students' attitudes prior to biology courses have little effect on their understanding of evolution when compared to their attitudes towards evolution at the completion of the course. However, the observed effect of positive attitudes toward evolution on academic success is small. Ingram and Nelson also found that curricula designed to answer students' questions about the subject matter are more effective in changing attitudes towards evolution and improving student achievement than curricula which teach evolution as an uncontested fact. Studies like this may indicate that students who are taught evolution in a way that addresses the concerns that creationists have towards evolution will have success in other measures correlated with academic success, including success in science based standardized tests measured in the research.

Whereas the Ingram and Nelson study related curricula, beliefs, and grades, this research will focus on classroom performance as measured by standardized tests rather than a grade given by the same teacher who taught the class. For the purpose of this research, grades are less useful than standardized test scores. A standardized test score measures how well a student learned the concepts which a higher body, such as a state, deems part of the biology curriculum. However, a grade merely reflects how well a student performed according to a certain teacher. For example, a student who has little understanding of evolution could perform equally as well as a student who has a comprehensive understanding of evolution if that student a teacher who had little interest in teaching evolution scored them. However, that same student would not perform as well on a standardized test, which does not have the same biases as a teacher might. This study will be using a standardized test administered by the state of Georgia to all students at the completion of a high school biology course.

Studies have also shown that at the high school level, the concepts that most efficiently prepare students for standardized tests are those based more on concrete topics rather than those that require critical thinking (Timmerman, Strickland et al. 2008). Therefore, it may be those classes that focus more on teaching the other aspects of biology and teach neither evolution nor creationism that bring students the most success on tests which measure a student's understanding of biology as a whole. Like the Ingram and Nelson study, the Timmerman et al study argues that the most effective method of teaching evolution is one in which evolution is taught in a way that clarifies students confusions and misconceptions about evolution.

Another study found that students who believe in evolution tend to do better than those who reject evolution for creationism or those who do not know their personal preference between the two (Lawson 1983). This study tested college students and found that those students with prior belief in evolution performed better in a test on evolution than those students with prior belief in creationism. This study, however, does not account for previous knowledge of evolution, which may be a major factor in a students' performance and could greatly vary between those who believe in evolution and those who do not.

The Lawson study, combined with other studies showing that beliefs relate to the curriculum taught (Lawson and Weser 1990), indicate that there will be a correlation between the curriculum taught and a student's performance on standardized tests. While the Lawson and Weser study found that curriculum impacts belief, they also found that other factors, such as scientific reasoning abilities, influenced belief.

Other studies proposed a different approach from the one suggested by Timmerman et al. These studies show that discussion of the debate between science and religion gives students a better understanding of science than if this divide were not discussed (Reiss 2008). This suggests that the students who would perform best on standardized tests are those who are taught to think critically about both evolution and creationism and who engage in the debate between the two.

Methods

The method examining the question of how changes in curriculum policy regarding creationism impact test scores will be a differences-in-differences (DD) statistical regression with the outcome variable of biology test scores and the predictor variable of a school's policy towards creationism. The model also controlled for confounding variables including the year, the county, the percent of students in the school who are disadvantaged, and information on the school's racial makeup. These variables were chosen for two reasons. The county and year control for many factors which would otherwise require a variety of different variables. For example, there are many minor differences in the way each county teaches the courses, such as how old students generally are when they take biology, how the classroom is structured, and other details relating to the curriculum. The study controlled for demographic variables because they are widely accepted to relate to performance. For example, high rates of economic disadvantage generally correlate with worse performance on academic metrics.

The first stage in the data collection process was finding those states, counties, and school districts that will serve as the units of observation in the analysis. To do this, articles and databases were found which track the developments in the evolution vs. creationism debate in specific areas and across the country.

The source used as a starting point is the American Institute of Biological Sciences database of the changes in state policies regarding evolution and creationism (AIBS 2010). This database tracks the major developments in the debate of evolution in the classroom, including changes in state policy and major court cases, over the past 12 years. The states listed by this database which have a school district that deviates from state standards in a direction less favorable to evolution are Georgia, New Mexico, Pennsylvania and Wisconsin. Additionally, the database identifies Arkansas and Maryland as having districts which are promote evolution in their biology curriculum more than the standards established by the state in which the district lies. For a summary of the states which have deviant districts and when and where those deviations occur, refer to Table 1.

Table 1: District Deviations from State Evolution Curricula

State	Deviant District	Direction of Deviation	Year of Deviance
Arkansas	Beebe	Anti-Evolution	Mid 1990s-2005
Georgia	Cobb County	Anti-Evolution	2002-December 2006
Maryland	Cecil	Pro-Evolution	2005
New Mexico	Rio Rancho	Anti-Evolution	2005-2006
Pennsylvania	Dover	Anti-Evolution	2004
Wisconsin	Grantsburg	Anti-Evolution	2004

Of these states and district of interest, Cobb County Georgia provided the most promising case for the study. Cobb County is a large enough population that there are a significant number of observations used for analysis. Also, Cobb County's deviation from state policy was long enough that the results reached are that of a five year period rather than a one year period. Furthermore, the state of Georgia has data accessible for every school from the 2003-2004 school year through the 2009-2010 school year. Therefore, the study contains a large amount of data from both the period in which Cobb County was placing stickers on school books and the period after they removed stickers from their biology books.

Data

The data collected are from the End of Course Tests administered by the state of Georgia. These tests are administered every year to each student in Georgia who took a high school level course in the relevant subject. To collect data on students' knowledge of biology, this study used observational data from the Georgia End of Course Test in biology. Georgia administered End of Course Tests each year since the 2003-2004 school year. Evolution represents approximately one-fifth of the state standards in biology and of 65 questions on the End of Course Test, the number of questions relating to evolution ranges from 10 to 15 questions (Education 2011).

Rather than collecting data from the entire state of Georgia, this study focused on five counties the northern suburbs of Atlanta. These counties include Cobb County, which instituted stickers on their text books from 2002 to 2006, Cherokee County, DeKalb County, Forsyth County, and Gwinnett County. In all, there are 473 data points in this study. Each data point represents the observation of a single high school for a single year.

Number of Students

The number of students refers to the number of students who took the biology end of course exam at that high school in that year. The analysis used the number of students at each school as a weight for the models. By weighting according to the number of students who took the test, this analysis gives each student the same amount of influence over the analysis, rather than an un-weighted analysis which would have given each high school the same amount of influence in the analysis.

Fail Rate

One of the response variables used for this analysis is the fail rate. The state of Georgia reports three statistics for a high school in a given End of Course Exam, each expressed in percentages, the fail rate, the pass rate, and the pass plus rate. While an average score for each school would have been the ideal response variable, such data could not be collected and this study used the fail rate as a response variable because it represents the portion of students who did not perform acceptably. This is equivalent to using the sum of the pass rate and the pass plus rate, as each student received a result of either fail, pass, or pass plus.

Pass Plus Rate

While the failure rate measures the number of students who performed poorly on the standardized test, the pass plus rate measures the number of students who performed exceedingly well on the standardized test. In addition to running models in which the failure rate was the response variable, models were also run in which the pass plus rate was the response variable. This ensures that measurement of schools not only by how well they were able to keep low performing students from failing the exam, but also by how well they help high performing students to achieve high scores on standardized tests.

Demographic Variables

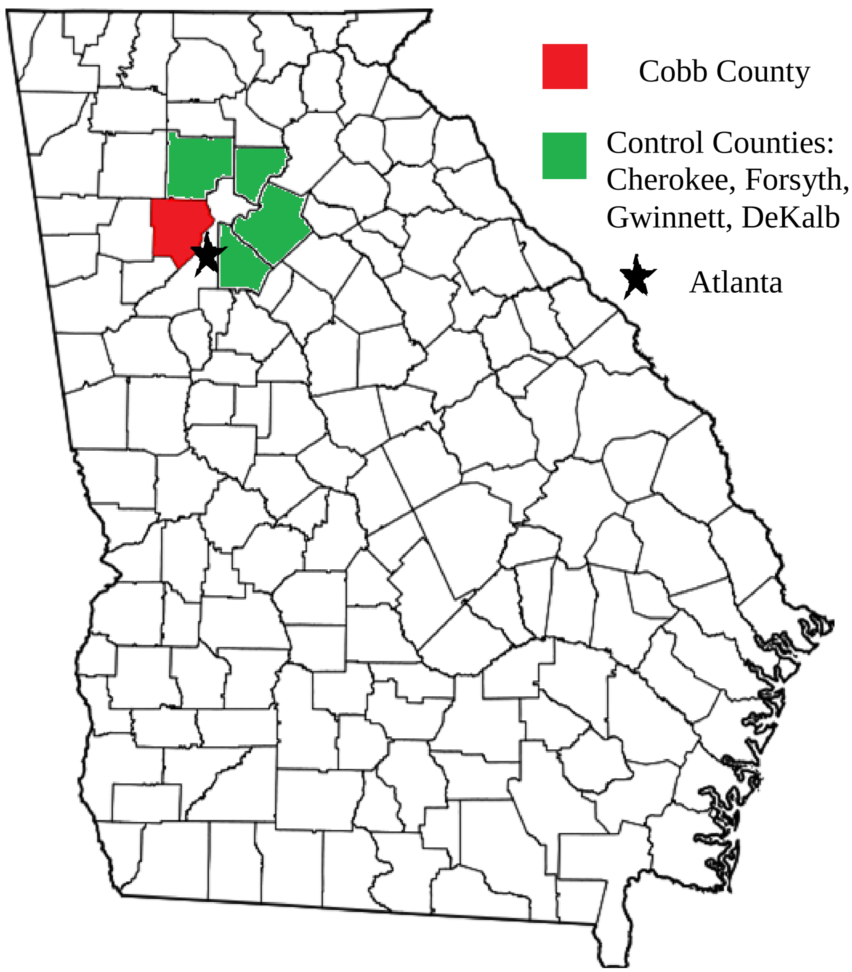
The state of Georgia also collects data on the race and background of those taking each test. The races included in the report are Asian, Black, Hispanic, Native American, Multicultural, and White. In addition to the racial data, Georgia also collects the number of students who were Economically Disadvantaged, Male, Female, Disabled, and Limited English Speakers. Based on these numbers, the portion of test takers who fit in that category at a particular high school was calculated.

Each data point also contained a variable for the year, labeled 2004 through 2010. The labeling indicates the year in which the school year ended. For example, data points labeled with a 2004 for the year represent the information for the 2003-2004 school year. The study consists of data from the 2003-2004 school year through the 2009-2010 school year for approximately 70 high schools in each year. The models controlled for the year because the distribution of failures and pass pluses greatly varied according to year. For example, the failure rate in the 2004-2005 school year was about 19 percent while the failure rate for the 2005-2006 school year doubled to 38 percent. By controlling for the year, the analysis offset the effects of this variation.

County

Each data point was also labeled by what county that high school is in. In Georgia, the county school district is the unit of educational organization and policy above the individual school. The data is from five counties, Cobb, Cherokee, DeKalb, Forsyth, and Gwinnett, for this analysis. These counties were selected because of their geographical similarity. Each of these counties represents Atlanta's northern suburbs, each of which borders at least one of the other four counties and Fulton County, the County which contains Atlanta, and was excluded due to concerns that the inner city culture of Atlanta would be too different from the educational culture in Cobb County. Of the 80 schools included in the study for 2010, 17 were in Cobb County, 6 were in Cherokee County, 6 were in Forsyth County, 22 were in Gwinnett County, and 39 were in DeKalb County. This puts Cobb County at the median of these five counties in terms of the number of high schools in the county. Cobb County is also near the other counties in terms of the number of students at each high school. A map of Georgia with the studied counties highlighted along with the location of Atlanta can be seen in figure 2.

Figure 2



Policy

The treatment variable for this analysis is the policy used by the school district at the time which the biology course was taught. Schools which used text books with the sticker declaring evolution as a theory, not a fact, were given the "sticker" designation for their policy. Schools in Cobb County from the 2003-2004 school year to the 2006-2007 school year used the sticker on their biology text books while schools outside of Cobb County and Cobb County after the 2006-2007 school year did not use the sticker on biology text books.

Table 2: Data Summary, Continuous Data

Variable	Mean	Weighted Mean	Weighted Standard Error	Weighted Correlation with "Fail"
Students	512	903.48	64.17	.105
Fail	36.15	31.26	.86	1
Disadvantaged	.433	.396	.0111	.843
Migrant	.0001	.0001	.00003	.003
Asian	.051	.057	.0027	.332
Black	.47	.405	.014	.778
Female	.48	.49	.0015	.161
Hispanic	.115	.121	.0055	.105
Limited English	.057	.058	.0029	.184
Native American	.0016	.0017	.0001	.105
Disabled	.099	.093	.0014	.017
Multiracial	.024	.026	.0006	.138
White	.34	.389	.014	.796

Table 3: Data Summary, Nominal Variables

Variable	Proportion of Observations	Weighted Proportion	Weighted Correlation with "Fail"
Policy: Yes	.123	.135	.107
Year: 2010	.169	.150	.057
Year: 2009	.160	.198	.001
Year: 2008	.152	.146	.063
Year: 2007	.148	.147	.166
Year: 2006	.135	.136	.178
Year: 2005	.120	.126	.232
Year: 2004	.116	.097	.149
County: Cobb	.234	.261	.105
County: Cherokee	.070	.073	.194
County: DeKalb	.378	.267	.575
County: Forsyth	.055	.056	.186
County: Gwinnett	.264	.344	.218

Statistical Modeling

The differences-in-differences method used involves creating a statistical regression. The regression run for this project was in the form of a logistic regression. The logistic regression model uses a variety of variables, such as whether or not a sticker policy is in place, to predict the odds that a student at a particular school will receive a particular score on the end of course test. In this case, one student was added to each school who failed the exam and one student was added to each school who received a pass plus on an exam. This ensured that the odds of a student receiving a particular grade on the exam would never be infinite, either in the negative or the positive. While the unit of observation for the study was the individual school, the weighting of the data by the number of students along with the transformation of the failure rate and the pass plus rate into the form of odds, allows interpretation to be done on an individual, student by student, level.

The models also controls for the major confounding variables and gives the strength of the relationship between the stickers placed on text books in Cobb County from 2002 through 2006. The model was run to include a variety of control variables. The first variables controlled for were those variables which are generally thought of as related to educational achievement, such as the racial composition of a school, the percent of the students who are economically disadvantaged, and the percent of students who are disabled. Also, the percent of the test takers who were female was also controlled for as females are generally believed to perform better on standardized tests. Additionally, the year the test was take was controlled for so that the difference in the difficulty of the test the particular year would not have an undue impact on the relationship between the sticker policy and failure rate. Finally, the county was controlled for in the case that some counties produce higher test scores because of other factors related to the county a student is in but unrelated to the other demographic data or the stickers, which could include other curriculum elements or the grade at which students in that county take biology courses.

The models reached are compared to additional models, one in which the sticker is not used to estimate the failure rate and the pass plus rate and one predicting the failure rate and pass plus rate of the history test. Comparing these models to the primary models demonstrates a relationship between the sticker policy and student performance that is not present if the sticker policy is no longer accounted for or if a measure of performance unrelated to evolution is studied.

Results and Analysis

Figure 3 represents the model in which the failure rate is predicted using stickers and other predictor variables. Figure 4 represents the same model without the use of stickers as a predictor. Figure 5 is the falsification test predicting the history fail rate. Figure 6 is the model predicting pass plus rate using stickers and other predictor variables. Figure 7 is the same model without using stickers to predict pass plus rate. Figure 8 is the falsification test predicting history pass plus rate.

Figure 3: Model Estimating Log(Odds(Fail)) by Policy and Control Variables

Estimator	*Estimate* *(Standard Error)*
Sticker	-.06153* (.03681)
Year	
2004	-.2654*** (.04752)
2005	-.5472*** (.04307)
2006	.6108*** (.04198)
2007	.4635*** (.04006)
2008	.1179*** (.04086)
2009	-.03502 (.03974)
2010	omitted
County	
Cherokee	.2369*** (.06080)
Cobb	-.09311**
DeKalb	-.1875*** (.05987)
Forsyth	.3464*** (.06587)
Gwinnett	omitted
Percent Economically Disadvantaged	.02001*** (.002048)
Percent Black	.004766*** (.001402)
Percent Female	-.03255*** (.004475)
Percent Disabled	.03354*** (.007257)
Percent White	-.01132*** (.002292)
Intercept	-.06153 (.2563)
R^2	.8856
N (Sum Weights)	473(234,524)

*Coefficient significant at the 10% level
**Coefficient significant at the 5% level
***Coefficient significant at the 1% level

Figure 4: Model Estimating Log(Odds(Fail)) by Control Variables

Estimator	*Estimate (Standard Error)*
Year	
2004	-.2848*** (.04618)
2005	-.5713*** (.04147)
2006	.5952*** (.04029)
2007	.4483*** (.03829)
2008	.1376*** (.03849)
2009	-.01293 (.03749)
2010	Omitted
County	
Cherokee	.2621*** (.05908)
Cobb	-.1440*** (.03458)
DeKalb	-.1921*** (.05855)
Forsyth	.3615*** (.06433)
Gwinnett	omitted
Percent Economically Disadvantaged	.01954*** (.001378)
Percent Black	.004733*** (.001378)
Percent Female	-.03170*** (.004441)
Percent Disabled	.03022*** (.006992)
Percent White	-.01198*** (.006992)
Intercept	-.05278 (.2505)
R^2	.8745
N (Sum Weights)	473(234,524)

*Coefficient significant at the 10% level
**Coefficient significant at the 5% level
***Coefficient significant at the 1% level

Figure 5: Model Estimating Log(Odds(History Fail)) by Policy and Control Variables

Estimator	*Estimate (Standard Error)*
Sticker	.06123 (.03971)
Year	
2004	.1504*** (.05125)
2005	-.2666*** (.04645)
2006	-.1875*** (.04542)
2007	-.2460*** (.04341)
2008	.3484*** (.04434)
2009	.1597*** (.04298)
2010	omitted
County	
Cherokee	.0657 (.06561)
Cobb	-.003330 (.05024)
DeKalb	-.3053*** (.06483)
Forsyth	.4496*** (.071)
Gwinnett	omitted
Percent Economically Disadvantaged	.01965*** (.002243)
Percent Black	.004204*** (.001524)
Percent Female	-.02596*** (.00523)
Percent Disabled	.009366 (.008013)
Percent White	-.008496*** (.002511)
Intercept	.3555 (.2865)
R^2	.8200
N (Sum Weights)	473(234,524)

*Coefficient significant at the 10% level
**Coefficient significant at the 5% level
***Coefficient significant at the 1% level

Figure 6: Model Estimating Log(Odds(Pass Plus)) by Policy and Control Variables

Estimator	*Estimate* *(Standard Error)*
Sticker	.07490* (.04467)
Year	
2004	.4567*** (.05781)
2005	.7104*** (.05237)
2006	-1.0076*** (.05104)
2007	-.8398*** (.04855)
2008	-.01995 (.04961)
2009	.2347*** (.04836)
2010	omitted
County	
Cherokee	-.1951** (.07615)
Cobb	-.004119 (.05492)
DeKalb	.3535*** (.07384)
Forsyth	-.3922*** (.08123)
Gwinnett	Omitted
Percent Economically Disadvantaged	-.02306*** (.002538)
Percent Asian	.02848*** (.004661)
Percent Female	.03113*** (.004861)
Percent Hispanic	.01207*** (.002356)
Percent White	.01626*** (.002199)
Intercept	-2.879*** (.2889)
R^2	.8746
N (Sum Weights)	473(234,524)

*Coefficient significant at the 10% level
**Coefficient significant at the 5% level
***Coefficient significant at the 1% level

Figure 7: Model Estimating Log(Odds(Pass Plus)) by Control Variables

Estimator	*Estimate (Standard Error)*
Year	
2004	0.4722*** (.05718)
2005	0.7287*** (.05131)
2006	-0.9884*** (.04985)
2007	-0.8207*** (.04730)
2008	-0.04400 (.04758)
2009	0.2115*** (.04643)
2010	Omitted
County	
Cherokee	-0.2128*** (.07556)
Cobb	0.05863 (.04027)
DeKalb	0.3368*** (.07331)
Forsyth	-0.4064*** (.08094)
Gwinnett	Omitted
Percent Economically Disadvantaged	-.02273*** (.002536)
Percent Asian	.028732*** (.004668)
Percent Female	.03116*** (.004871)
Percent Hispanic	.01186*** (.002357)
Percent White	.01653*** (.002198)
Intercept	-2.887*** (.2854)
R^2	.8738
N (Sum Weights)	473(234,524)

*Coefficient significant at the 10% level
**Coefficient significant at the 5% level
***Coefficient significant at the 1% level

Figure 8: Model Estimating Log(Odds(History Pass Plus)) by Policy and Control Variables

Estimator	*Estimate (Standard Error)*
Sticker	-.0371773 (.04778)
Year	
2004	0.05889 (.06194)
2005	0.3504*** (.05618)
2006	0.3354*** (.05509)
2007	0.3182*** (.05180)
2008	-0.7300*** (.05326)
2009	-0.2748*** (.05263)
2010	omitted
County	
Cherokee	-0.1195 (.07890)
Cobb	0.006868 (.05879)
DeKalb	0.3613*** (.07760)
Forsyth	-0.4704*** (.08518)
Gwinnett	Omitted
Percent Economically Disadvantaged	-.01719*** (.003052)
Percent Asian	-.01485 (.004114)
Percent Female	.03947*** (.006133)
Percent Hispanic	-.007947 (.005550)
Percent White	.0002912 (.004054)
Intercept	-.08154*** (.3679)
R^2	.7962
N (Sum Weights)	473(234,524)

*Coefficient significant at the 10% level
**Coefficient significant at the 5% level
***Coefficient significant at the 1% level

Strength of Models

The model used to predict the odds of failure at a particular high school given information about that high school's demographics, county, year, and sticker policy produces an RSquare value of .8856. This indicates that the model could explain approximately 88.56

percent of variation in the weighted transformed odds of failure at various high schools in suburban Atlanta in various years. However, the model in which policy is not used to predict failure also produces a high RSquare value of .8745. By contrasting these two models, it is shown that adding policy as a predictor variable brings about a slight increase in the predictive power of the model which already uses year, county and some demographic data.

Similarly, when the model predicting the rate of pass pluses was used, the RSquare value was .8746, indicating that this statistical model accounts for 87.46 percent of the variability in the transformed odds of pass plus. When compared to the model which predicts pass plus without using policy, there is only a slight change in the RSquare value, from .8746 to .8738. This very slight change indicates that adding policy as a predictor variable does not drastically change the effectiveness of the model.

The Effect of Stickers

The most notable result from the model is that removing the sticker policy relates to an increase in the odds of a student failing the test. According to the data, schools that have a policy in which there are no stickers are predicted to have log odds of failing that is .06 higher than that of a school in which the stickers are placed on the books. This indicates that a student at a school without a sticker policy will have 6 percent higher odds of failing than that student would have at a school with stickers on their text book. While this does represent a small difference in how a student will do on the test, it goes against the intuition that those students who did not use biology textbooks that labeled evolution as “not a fact” would perform worse on a test which was based on the theory of evolution.

Similarly, the sticker policy relates to an improvement in the pass plus rate of those schools that put stickers on their textbooks. Equation 2 suggests that schools which did not have the stickers on their biology text books had a log odds of pass plus .075 lower than the log odds of those schools that did have the stickers on their text books, even after, the county, the year, and various demographic factors were accounted for. This indicates that the odds of a student receiving a pass plus rating on the test increase by 7 percent if that student is in a school that places the controversial stickers on their textbooks. For example, the model can be run for two hypothetical students, each at a Cobb County school that is 39 percent disadvantaged, 40 percent black, 49 percent female, 9.3 percent disabled, and 39 percent white. The only difference between these two hypothetical students is that one student took the end of course test in 2004, when the policy was in effect, and the other took the test in 2009, after the stickers had been removed from the test. If the model is adjusted to assume that the overall failure rate was the same in 2004 as it was in 2009, then the odds that the 2004 student failed the test are 1.633 while the odds that the 2009 student failed the test are 1.737. That is, there is a 62 percent chance that the student in 2004 failed the test while there is a 63% chance that the student in 2009 failed the test. While this difference is rather minimal, over the average of the 234,524 students studied, one can conclude with some confidence that there is a difference between the failure rate of those students who were subjected to the sticker is lower than the failure rate of those students who were not subjected to the sticker.

Weak Relationship between Stickers and Performance

Even though the evidence suggests that having stickers on the textbooks decreases a student's odds of failing, strong conclusions cannot be made. The p-value for the policy variable in this regression was .095. This indicates that if there were in fact no relationship between the sticker policy and the odds of failure, the data collected would demonstrate a relationship as strong as the one the model suggests 10 percent of the time. By comparison, the p-value for disadvantaged students is less than .0001, indicating that if there were in fact no relationship between the percent of disadvantaged students and the odds of failure, data demonstrating this strong of a relationship would occur less than one in every ten thousand data sets. The p-value for the policy of .095 is considered to be of moderate statistical significance. There seems to be a strong enough relationship in the data to lead us to believe that there is a relationship between the sticker policy and the odds of failure. However, it is difficult to draw any certainty in this relationship.

The model in which the pass plus levels are used to measure a school's performance shows similar results. In this model, the p-value for the policy variable was .0943, indicating that if there were no relationship between policy and the odds of a student receiving a pass plus, the a relationship such as the one observed would be observed less than 10 percent of the time. This again demonstrates moderate statistical significance. The relationship is strong enough to suggest that there is a relationship between the policy used at a particular school and the odds of a student receiving a pass plus score, but it is not strong enough to give certainty as to this relationship.

Falsification Test

Figures 5 and 8 provide models used for a falsification test. That is, these models compare to the original models to discover how valid those original models are. These two models used history test results as the response variable. Because the sticker policy should be completely unrelated to the way students at a school performed on a history test, a model in which the impact that policy has on these history results is minimized is expected.

In the model predicting the failure rate of the history test is predicted, the removal of the stickers from the textbooks are predicted to decrease the odds of a student's failure by 6 percent, at a p-value of .124. This is a statistically insignificant relationship, although there is stronger relationship exhibited in the data than expected from this model. While the relationship between the policy and failure on the history test is stronger than what would was expected, it is still weaker than the relationship between the sticker policy and failure on the biology test. This indicates that the relationship between the sticker policy and the biology test is probably weaker than what is observed in Figure 3, but provides more confidence that some sort of relationship exists.

Figure 8 gives even stronger verification of the relationship between the sticker policy and student performance on the biology test. The model predicting the pass plus rate on the history test indicates that the stickers resulted in only a 3.7 percent decrease in the odds of a student at a particular school receiving a pass plus, with a p-value of .437. This is a very high p-value, indicating that there is almost no relationship between the sticker policy and the odds of a student receiving a pass plus score.

The falsification tests provide additional confidence because the directional values of the predicted impact of the sticker policy on history test scores is the opposite as that for biology test scores. The history test models demonstrate that Cobb County performed better relative to other counties after 2007 than they did before 2007. This could indicate that there is some unknown variable that increases Cobb County scores after 2007. Given this, the fact that biology test scores in Cobb County fell relative to other counties after 2007 demonstrates that the stickers may have an even stronger impact on the biology test scores than the original models suggest.

Analysis of Other Factors

Beside the main focus of this study, the two models produced interesting findings about the control variables. For example, high schools in DeKalb County received the highest rate of pass pluses and the lowest rate of fails, followed by Cobb County and Gwinnett County. Additionally, the fail rates were the lowest and the pass plus rates were highest in 2004, 2005, and 2009. Disadvantaged students, disabled students, and black students seem to lower the pass plus rate and increase the fail rate. However, having many white and female students seems to relate to higher pass plus rates and lower failure rates.

Limitations

Limitations to the analysis limit the strength and types of the conclusions which can be drawn from the study. These limitations are caused by both the nature of the analysis used, such as the number of observations, the type of data and by confounding factors outside the scope of the study.

The first limitation of this model is that there is a small amount of data. In this case, there were 473 points of data. This low sample size limits the standard errors of the coefficients to relatively high values, which in turn add difficulty to drawing statistical significance to any relationships which are not extremely strong. While a sample size of 473 would normally be enough to make strong conclusions about an analysis, because the analysis required many control variables, it could be an over-fit model of the data in which the use of too many control variables produces strong predictive power, when in fact there could be little predictive power.

Even these relatively high standard errors were likely lower than the standard errors which should have actually been used for the study. Bertrand, Duflo, and Mullainathan argue that using DD models often results in standard errors which are too low, and thus a tendency to draw conclusions about relationships which do not exist. While the method of separating out the years into a before-sticker-removal period and after-sticker-removal period helped to create accurate standard errors, because of the low number of treatment and control counties studied, even these standard errors were likely too low (Bertrand, Duflo et al. 2004).

The nature of the data collected limits the analysis. Because each school was measured only in two numbers, the fail rate and the pass plus rate, it is difficult to draw conclusions about how well a school truly performed. It would have been preferable to have some sort of numerical score based on the number of questions students answered correctly for each school. That way, instead of dividing students into three categories, students' scores could be weighted based on the strength of their score. A school with many students who had fantastic scores could differentiate itself from a school whose students barely overcame the pass plus threshold.

The test used to measure a student's understanding of evolution is a biology test. While this is the test with the largest evolution component, it is not directly related to a student's understanding of evolution. Instead, the test examines students on a variety of topics under the broader umbrella of biology. Therefore, it is possible that the students in Cobb County received better scores in the years in which the stickers were on their textbooks because they understood some aspect of biology unrelated to evolution better than the students who they were being tested against, even if they performed poorly on the one-fifth of the test pertaining towards evolution.

The study does not take into account the degree to which the sticker policy was implemented. The sticker policy could have been implemented slowly, only switching between books with stickers and books without stickers when books needed to be replaced, or it could have been implemented immediately. In the model, it was assumed that the effects of the stickers would be seen in the student who took biology classes from the 2003-2004 school year to the 2006-2007 school year. However, if there were some schools that did not place the stickers on their textbooks until after the 2003-2004 school year or removed the sticker after the 2006-2007 school year, then the policy variable would have been miscoded for those schools.

There is also a limitation in the generalizability of the conclusions of this study; the conclusions found in the data studied may not hold if the sticker policy were put in place in another school district or at another time. To draw conclusions which could be generalized, this study would have had to pull data from a wide range of education systems and geographic locations.

Another possible limitation is the fact that data were only collected from the time during which the stickers were on the textbooks and the time after the stickers were removed. This allows for the possibility that the decline in student performance on the biology exam after the stickers were removed were part of a general decline in the performance or the teaching at Cobb County schools. However, the falsification test done using the history test as the response variable provides reason to believe that this limitation does not impact the validity of the conclusions that can be made. Because the falsification test shows that the sticker policy was less related to history scores than it was to biology scores, the difference between the biology scores from the period in which the stickers were placed on the books and the scores from the period in which the stickers were not placed on the books is likely not due to a change in the intellectual climate of Cobb County schools.

Conclusions

As discussed in the theoretical framework section of this paper, there were two possible ways that a sticker policy and a school's performance on standardized tests could be related. First and more obviously, the stickers could prove to harm a school's performance on standardized tests. By this logic, a sticker which leads students to question one of the fundamental theories of biology would cause that student to struggle to gain a full appreciation of the material being taught in their biology classroom. A district that did not hold the same beliefs about evolution in the classroom as the rest of the state would not perform as well as other districts in statewide tests. This theory seems to have been disproven by the empirical data in this study. Rather than seeing an increase in the odds of failure and a decrease in the odds of a pass plus when a student was treated to the sticker policy, the data demonstrate that there is a decrease in the odds of failure and an increase in the odds of receiving a pass plus at those schools which place stickers on their biology text books claiming that evolution is not a fact.

The other plausible theory for the relationship between a sticker policy and the test results of a particular school would be that the controversy surrounding evolution leads students to take more of an interest in it, which in turn could result in better test performance. The stickers, widely disapproved of by science teachers, some parents, and the courts, would inspire students to take an interest in the material that they were learning in the classroom. Rather than viewing evolution as a standard scientific theory, students would view evolution as a controversy, something that they could form their own opinion on and that their opinion mattered, rather than having to trust the expertise of their teachers. Rather than teaching evolution as one of many units in the biology curriculum, Cobb County teachers could present evolution as a current event. Teachers could reference the status of the ongoing court cases, or recent decisions made by the board of education. This would lead students to seek out answers about evolution, what it is, and why scientists hold it as a fundamental theory. This could even have spillover effects on the entire biology course. By seeking out answers about evolution, students would then pay more attention to the discussion of evolution in the classroom and therefore perform better on tests which measure their knowledge of evolution.

In both the model which predicts the odds of a student failing and the model which predicts the odds of a student receiving a pass plus, the sticker has a minimal effect on student performance. Even if the models underestimate the effect that the sticker had on student performance, the data suggest that removing the sticker will change the odds of failing or receiving a pass plus by no more than 15 percent, only about a quarter of the effect of county to county variation and a tenth of the effect of the year to year variation. Because student test scores are largely unaffected by the sticker policy, then factors other than academic success should figure more prominently in future debate over these stickers.

The data observed in this study support the hypothesis that stickers would inspire inquisition about evolution, leading to higher test scores, rather than inspiring students to doubt evolution in a way that impacts their understanding of evolution negatively. Students at schools with the stickers on their biology text books have both lower odds of failing the end of course test in biology and higher odds of receiving a pass plus on the test.

The acceptance of this hypothesis has interesting policy implications for educators. Rather than focusing on making sure that students receive an education which espouses values that are politically correct, schools should seek to stir up controversy. They should seek out opportunities to get their students interested in school by encouraging them to take sides on issues that haven't been decided. Schools should allow for debates to be had between differing opinions rather than teaching one accepted opinion. If students believe that they are learning things straight out of a textbook, then they won't be interested. However, if they are learning about a subject that has been the subject of national debate and has caused local dispute, they are more likely understand the material. Therefore, rather than concluding that schools should or shouldn’t adopt a sticker policy, this study concludes that schools should adopt policies which will grab the attention of students and encourage them to form their own views.

References

AIBS. (2010). "AIBS State News on Teaching Evolution." Retrieved 3/25/2011, 2011, from http://www.aibs.org/public-policy/evolution_state_news.html.

Bertrand, Marianne, Duflo, et al. (2004). "How Much Should We Trust Differences-In-Differences Estimates?" The Quarterly Journal of Economics Volume 119(Issue 1): Pages 249-275.

Darwin, Charles (1859). On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. London, John Murray.

Education, Georgia Department of (2011). "Report Card."

Ingram, Ella L. and Craig E. Nelson (2006). "Relationship between achievement and students' acceptance of evolution or creation in an upper-level evolution course." Journal of Research in Science Teaching 43(1): 7-24.

Isaak, Mark (1995) "Five Major Misconceptions About Evolution." TalkOrigins.Org.

Lawson, Anton E. (1983). "Predicting science achievement: The role of developmental level, disembedding ability, mental capacity, prior knowledge, and beliefs." Journal of Research in Science Teaching 20(2): 117-129.

Lawson, Anton E. and John Weser (1990). "The rejection of nonscientific beliefs about life: Effects of instruction and reasoning skills." Journal of Research in Science Teaching 27(6): 589-606.

Lerner, L.S. (2000). "Good science, bad science: Teaching evolution in the states."

Mead, L.S. and A. Mates (2009). "Why science standards are important to a strong science curriculum and how states measure up." Evolution: Education and Outreach 2(3): 359-371.

Reiss, M. J. (2008). "Should science educators deal with the science/religion issue?" Studies in Science Education 44(2): 157-186.

Timmerman, B. E., D. C. Strickland, et al. (2008). "Curricular reform and inquiry teaching in biology: where are our efforts most fruitfully invested?" Integrative and Comparative Biology 48(2): 226-240.

Acknowledgements:

I would like to acknowledge Professor Jacob Vigdor, the advisor on this project. He provided valuable insight as to the scope and method of the study in addition to reading countless drafts of this project.

I would also like to acknowledge Professor Judith Kelley, the thesis program director. Without her support, this project would never have happened.

Finally, I would like to thank Professor Mark Jeuland. Professor Jeuland suggested the idea of using the history test data for a falsification test.