

Modeling Abstinence Education Effectiveness

Thomas E. Smith¹, Burcu Atar², Teresa Ferreira³, Pamela Valentine⁴, M.Graça Pereira^{5*}

¹College of Social Work, The Florida State University, FL 32306-2570, United States

²Department of Educational Measurement and Evaluation, Hacettepe University, Turkey

³The Florida State University, FL 32306-2570, United States

⁴Southeastern Research Institute of Tallahassee, FL 32306-2570, United States

⁵School of Psychology, University of Minho, 4710-057 Braga, Portugal

Abstract

Background: Controversy about the effectiveness of abstinence education has posed troubling dilemmas for everyone involved in this area of study. Strident statements about the lack of efficacy of abstinence education have approached the level of bitter ideology. One remedy to lessen this focus on ideology is to provide a broader analysis of program efforts.

Method: This paper provides an innovative analysis of a community-based abstinence education program that encompassed multiple schools across several counties that includes thousands of students. The design addressed many deficits in published studies; it used hierarchical linear modeling to remedy the design flaws of a simple pretest-posttest analysis.

Results: Pretests were the principal predictors of posttest scores. Gender was also a significant predictor of posttest scores. Age however was not a significant predictor. An interaction between gender and age was a significant predictor although a three-way interaction of gender x age x race was not.

Conclusion: Implications for the findings are stated with recommendations for further research.

Publication History:

Received: July 22, 2014

Accepted: October 30, 2014

Published: January 05, 2015

Keywords:

Abstinence education, HLM, Outcome studies, Modeling

Introduction

Any casual reader of the literature on abstinence education would be bewildered at the acrimony that exists between comprehensive sexual education and abstinence only education proponents. Scores of studies decry the lack of efficacy and costs of abstinence education [1,2]. Other studies support the effectiveness of abstinence education [3,4]. Further, the reports of political involvement in studies of abstinence education are troubling [5,6]. Kirby [7] summarizes the muddle of opinions when he concludes that little can be concluded about the efficacy of abstinence education. As a way of understanding this wide variation in beliefs, some writers [8,9] suggest that this ongoing controversy is akin to a morality play in which religious beliefs are at the heart of adherence to a choice of curriculum. Thus, one conclusion can be drawn by researchers is that ideology trumps methodology.

Authors of reviews uniformly conclude that abstinence-only studies lack credibility because they fail standards of adequate efficacy research methods [10,11,7]. The "gold standards" of efficacy trials are complex and subject to multiple errors. Efficacy research requires carefully specified treatment manuals [i.e., educational curricula] applied by highly skilled educators to a clearly specified student population [12]. Further, the research designs require a high degree of control over the environment to enable randomized control over different treatment conditions. Flexibility in procedures and choice of measures is unlikely. It is not surprising that few psychosocial treatment studies are able to meet CONSORT criteria, commonly used in medical journals [13]. The abstinence-education conditions do not surpass control groups in terms of their effects. In a world of box score summaries, abstinence education has failed to justify its existence. Such victory however may be misleading pyrrhic. Recent data support the effectiveness of abstinence-education programs [14]. It may well be that abstinence educators have altered their programs in response to withering criticism. Further, it is important to consider anecdotes from scores of schools and programs nationwide who extol the virtues of their abstinence-education programs. Thus, successful programs have been reported from a widespread sampling

of educators within diverse settings in diverse educational curricula administered to diverse student populations [15]. Further, the curricula undergo continual specification depending on the needs of students. Finally, many curricula who undergo some degree of quasi-experimental investigation have shown significant effects. Thus, abstinence-education providers can claim that their programs are effective, if not efficacious.

However, there are other considerations. Typically, abstinence education is presented in schools and in classroom settings. There is the possibility that classroom-specific effects might obscure the overall impact of abstinence education. In any long-term study of any intervention modality, it is important to consider plausible threats to conclusions of effectiveness and efficacy. Presumably, this type of study begins the process of unpackaging the black box of abstinence education efforts.

Method

Sample

This study examined the programs being delivered to 35 schools. Each pregnancy center has a full time county coordinator that schedules schools, teaches classes, organizes and prepares materials, does some of the grading and recording of the grids and supervises the part time facilitators. Data on a little over 3000 [n=3183] participants who received abstinence training during 2008 are reported here. The number of participants had nearly equal numbers

***Corresponding Author:** M.Graça Pereira, School of Psychology, University of Minho, 4710-057 Braga, Portugal, Tel: +351 253 604228; Fax: +351253 604224; E-mail: gracep@psi.uminho.pt

Citation: Smith TE, Atar B, Ferreira T, Valentine P, Pereira MG (2015) Modeling Abstinence Education Effectiveness. Int J Psychol Behav Anal 1: 102. doi: <http://dx.doi.org/10.15344/2455-3867/2015/102>

Copyright: © 2015 Smith et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

males and females. Three-quarters of the participants were Caucasian while the remainder were equally split between African-Americans and Hispanic students.

Procedures

The curriculum included the A-H components and 13 themes that are mandated by federal legislation; the activities are a mix of commercially available curricula; the outputs are the scores on the knowledge and attitudes questionnaire whose items directly measure the A-H components. Although this study did not go to the level of measuring impact, it did provide a methodological argument by which impact can be inferred.

During the first year of funding, the project team hired staff, finalized relationships with site administrators, purchased abstinence education curriculum, created measures, and trained facilitators. All aspects of the project were piloted and the results were examined. Second, the initial curricula were modified based on project staff's observations and participant feedback. Third, the outcome questionnaire also underwent changes to better reflect A-H components and 13 themes. Thus, the first year consisted of an iterative process to prepare for a roll out in the second year that included the current curriculum, activities, and outcome measures.

Facilitators versus classroom teachers delivered the curricula; project staff observed them during development and during each facilitators' training. After being trained, project staff randomly viewed the facilitators' work and gave them feedback. To ensure that there was not observer drift, two staff members were present throughout these fidelity checks. Thus, there was a high level of fidelity in what was presented to students during the second year. In summary, curricula were chosen with an eye towards replicability, manualization, fidelity in implementation, and adherence to federal A-H components and 13 themes.

For each classroom within each school, facilitators and not the classroom teachers administered the outcome measure before and after the training occurred. The measure was developed for the program and consisted items that directly reflected the mandated components and themes. The resulting prepost research design, while not optimal, provided a minimal level of assurance as to the effectiveness of program efforts.

To ascertain whether there was a nesting effect in the curricula being implemented in individual classrooms, a two-way hierarchical linear modeling [HLM] strategy was pursued. HLM was used to control for any nesting effects at the classroom level. If the results are not significant at classroom levels, it can be inferred that the treatment effectiveness was not due to the classroom in which the students received the educational curriculum.

This study was designed to provide access to nested data where the Level-1 were students and the Level-2 were classrooms. The Level-1 predictor variables were pretest scores, hours, age, gender, and race. The first variable [i.e., pretest scores] was interval, grand-mean centered variable. The second variable [i.e., age] was interval, grand-mean centered. The third variable [i.e., gender] was dichotomous, uncentered, which takes on a value of 1 for boys, and 0 for girls. The fourth variable [i.e., ethnicity] was categorical, uncentered, and dummy coded with 1 for Caucasian, 0 for Black; 1 for Caucasian, 0 for Hispanic; 1 for Caucasian, 0 Other Ethnicities.

In addition to these variables, the two-way interaction of age and gender, the three-way interaction of age, gender and ethnicity was also considered. The Level-1 outcome variable was posttest scores. The Level-2 predictor variable was class size, an interval and grand-mean centered variable. This study involved 3,993 students nested in 142 classrooms. The descriptive statistics for the outcome, the student-level and classroom-level variables are presented in Table 1.

Student-Level Variables	Variable Name	Mean	Standard deviation
Posttest scores	Y_{ij}	97.13	13.03
Pretest scores	$(PRE)_{ij}$	78.53	17.48
Age	$(AGE)_{ij}$	14.44	1.38
Gender	$(GENDER)_{ij}$	0.49	0.50
Whites-Blacks	$(RACE1)_{ij}$	0.12	0.32
Whites-Hispanics	$(RACE2)_{ij}$	0.12	0.32
Whites-Others	$(RACE3)_{ij}$	0.04	0.21
Sex x Age Interaction	$(INT2)_{ij}$	7.21	7.38
Class-Level Variables			
Class Size	$(CLSIZE)_{ij}$	28.28	16.27

Table 1: Descriptive Statistics for Outcome and Explanatory Variables at the Student and Classroom Levels.

The effect of six student-level predictor variables [i.e., pretest scores, age, gender, race, age*gender interaction, and age*gender*race interaction] on the outcome variable [posttest scores] within classrooms was studied. In addition, an effect size was performed the effect of class size on the posttest scores obtained by the students in each class room.

With a hierarchical linear model, each level in this structure is formally represented by its own sub-model. These sub-models express relationships among variables within a given level, and specify how variables at one level influence relations occur at the other level. Thus, HLM was used in this particular study to help improve the estimation of individual effects, to formulate and test hypotheses about how variables measured at one level affect relations occurring at another level, and to estimate the variance and covariance components with nested data.

A one-way ANOVA with random effects provided useful preliminary information about how much variation in the outcome [i.e., posttest scores] lies within and between classrooms and about the reliability of each classroom's sample posttest scores as an estimate of true population posttest scores. The following is the level-1 or student-level model: $Y_{ij} = B_{0j} + r_{ij}$, where Y_{ij} is the posttest score of student i in classroom j , B_{0j} is the mean posttest score in classroom j , and r_{ij} is the deviation of the posttest score of student i from mean posttest score of classroom j . We assume $r_{ij} \sim$ independently $N[0, \Phi2]$ for $i=1, \dots, n_j$ students in classroom j , and $j=142$ classrooms. $\Phi2$ is the student-level variance.

The following is the level-2 or classroom-level model:

$B_{0j} = G_{00} + u_{0j}$, where G_{00} is the grand-mean posttest score across classrooms and u_{0j} is the deviation of the mean posttest score of classroom j from grand-mean posttest score. We assume $u_{0j} \sim$ independently $N [0, \theta_{00}]$. θ_{00} is the class-level variance.

This yields a combined model: $Y_{ij} = G_{00} + u_{0j} + r_{ij}$ with fixed effect G_{00} and random effects u_{0j} and r_{ij} .

Results

A fully unconditional HLM was used to gather preliminary information about the reliability estimate of overall classroom means of posttest scores and the amount of variation in posttest scores that lies within and between classrooms in the sample. The results of the analysis are given in Table 2. The reliability of the overall classroom means was estimated to be around 0.940. This reliability estimate indicate that the sample classroom means are quite reliable as an indicator of the true classroom means. The high reliability justified further modeling.

Within-class variance ($\hat{\sigma}^2$)	116.98
Between-class variance ($\hat{\tau}_{00}$)	103.87
Reliability of mean post test score ($\hat{\lambda}$)	0.94
Intraclass correlation for between-class variability ($\hat{\rho}$)	0.47
Intraclass correlation for between-class variability, adjusted for reliability ($\hat{\rho}_{adj}$)	0.49

Table 2. Fully Unconditional HLM Results.

The adjusted intraclass correlation, which represents the proportion of variance in posttest scores between classrooms, and adjusted for reliability was calculated to be 0.486 using the following formula, $\hat{\rho}_{adj} = \frac{\hat{\tau}}{\hat{\tau} + (\hat{\sigma}^2 \times \lambda)}$. This value indicates that about 49% of variance in posttest scores was due to differences on mean posttest scores among classrooms whereas about 51% of variance in posttest scores was due to individual differences among students. The high intraclass correlation for between-class variability supported the use of HLM.

Unconditional within-class HLM

In the unconditional within-class model, the student posttest score was estimated as a function of adjusted mean posttest score, pretest score, age, gender, race and two-way interaction of age and gender. While the adjusted mean posttest scores and pretest score slopes were modeled as randomly varying parameters over classrooms at level-2, age, gender, race, and two-way interaction of age and gender slopes were modeled as fixed parameters at level-2. The results of the unconditional within-class model are presented in Table 3. The adjusted mean of posttest scores over classrooms was estimated to be around 97.021 with a standard error of 0.535. It was found that the adjusted mean of posttest scores significantly among classrooms [$p < 0.001$], indicating that there are significant differences on mean posttest scores among classrooms. The average effect of pretest scores on student posttest scores was estimated to be 0.178 and on average, the effect of pretest scores on posttest scores was found to be statistically significant [$p < 0.001$]. However, the effect size for the average pretest score slope is trivial [ES = 0.016]. It was also found that the pretest score slopes statistically significantly vary among classrooms [$p < 0.001$]. The average effect of age on posttest

scores was estimated to be -0.469 and it was found that age is not significantly related to posttest scores [$p = 0.088$]. The average gender gap in posttest scores was estimated to be around 7.148 and the effect of gender on posttest scores was found to be statistically significant [$p = 0.038$]. Based on the effect size measure, it can be said that the average posttest score of males is about 0.661 standard deviations higher than that of females when other variables are controlled, reflecting a large effect. Even though the results show that the average effect of age on posttest scores is not statistically significant, its interaction with gender was found to have a statistically significant effect on posttest scores. For the race variable, the gaps between Whites and Blacks and Whites and Hispanics in posttest scores were found to be statistically significant whereas the gap between Whites and others in posttest scores was found to be statistically nonsignificant with a negligible effect size. The average posttest score of Whites is about 0.192 standard deviations higher than that of Blacks; the average posttest score of Whites is about 0.123 standard deviations higher than that of Hispanics, and the average posttest score of Whites is about 0.003 standard deviations higher than that of others when the other variables are controlled.

Fixed Effect		Coefficient	SE	<i>p</i>	Effect Size
Adjusted mean posttest score	γ_{00}	97.02	0.54	< 0.001	-----
Mean pretest slope	γ_{10}	0.18	0.02	< 0.001	0.02
Mean age slope	γ_{20}	-0.47	0.28	0.09	-0.04
Mean gender slope	γ_{30}	7.15	3.44	0.04	0.66
Mean Whites vs. Blacks slope	γ_{40}	-2.08	0.77	0.01	-0.19
Mean Whites vs. Hispanics slope	γ_{50}	-1.33	0.65	0.04	-0.12
Mean Whites vs. Others slope	γ_{60}	-0.04	0.67	0.96	-0.003
Mean age-gender interaction slope	γ_{70}	-0.52	0.22	0.02	-0.05
Random Effect	γ_{80}	SD	Variance	<i>p</i>	
Level-2 Error Term					
Adjusted classroom mean	u_{0j}	5.78		33.46	< 0.001
Pretest score effect	u_{1j}	0.21		0.04	< 0.001
Level-1 error term	r_{ij}	9.78		95.60	

Table 3: Unconditional Within-Class HLM Results.

When the within-class variance in the fully unconditional model [$\hat{\sigma}^2 = 116.98$] was compared to the within-class variance in the unconditional within-class model [$\hat{\sigma}^2 = 95.6$], the proportion reduction in variance or proportion variance explained at level-1 was calculated to be 0.182. It can be concluded that adding pretest scores, age, sex, race, and the interaction term as predictors of posttest scores reduced the within-class variance by 18%. In other words, pretest scores, age, sex, race, and interaction term accounted for 18% of the student-level variance in the posttest scores.

Conditional between-class HLM

In the conditional between-class HLM, class size was included into the level-2 model to explain the variation on the adjusted mean posttest scores and on the pretest score slopes among classrooms. The results are given in Table 4. The effect of the class size on both the adjusted mean posttest scores and the pretest score slopes was not

found statistically significant.

Fixed Effect		Coefficient	SE	<i>p</i>	Effect Size
Adj mean posttest score model					
Intercept	γ_{00}	97.02	0.54	< 0.001	-----
Class size slope	γ_{10}	0.03	0.03	0.24	0.00
Pretest slope model					
Intercept	γ_{11}			< 0.001	-----
Class size slope	γ_{12}	0.18	0.02	0.18	0.00
Random Effect					
		SE	Variance	<i>p</i>	
Level-2 Error Term					
Adjusted classroom mean	u_{0j}		5.78	33.46	< 0.001
Pretest score effect	u_{1j}		0.21	0.04	< 0.001
Level-1 error term	r_{1j}		9.78	95.60	

Table 4: Conditional Between-Class HLM Results.

This 2-way HLM analysis confirmed that classroom variables were not a determining factor in the significant scores that indicated success of the educational curriculum. The results were however significant at the individual student level. The results at this level suggest that the program effectiveness could be explained by the change in individual students and not by county or classroom membership. Although there are many other reasons that could explain the change in scores, a common sense analysis of the evaluation results suggested that it is likely that program services were the principal reason why positive results occurred.

Discussion

Controversy about the effects of abstinence education will undoubtedly continue in professional journals and in political arenas. In this study, we sought to dispel the problems of intraclass correlation that would undermine assertions that the curriculum was effective across time. The study design controlled for the effects of classroom, county, school variables; further, it controlled for the effects of gender, age, and the ethnicity. Age in this study acted as a latent indicator of student development. The purpose of the study was to further study the effects of abstinence education curricula.

The results are intriguing. There were nesting effects that provide cautionary notes for large sample analyses across classrooms. Not surprisingly, pretests were the principal predictors of posttest scores. Gender was a significant predictor of posttest scores. Age however was not a significant predictor. An interaction between gender and age was a significant predictor although a three-way interaction of gender x age x race was not.

The authors began with a discussion of effectiveness versus efficacy. Even this discussion is rife with controversy. Because there is a continuum of methodologies ranging from the “gold standard” of efficacy trials to the cloudiness of service research, the methodology of this study will likely be seen as falling somewhere in between a sole focus on internal validity as compared to one on external validity.

There can be no doubt that there were flaws with study design. The lack of a comparison group, behavioral measures, and long-term follow-up are significant threats to internal validity. This study surely

could not make any claim to efficacy study using the “gold standard” of experimental design. Nor was it designed to be so. Rather, this study was designed as a modest addition to the literature on mediators and moderators of abstinence education. It supports weed and his colleagues’[14] recent study that examined mediators of abstinence education. It is this type of examination that will increase understanding of abstinence education.

References

1. Trenholm C, Devaney B, Fortson K, Clark M, Bridgespan LQ (2008) Impacts of abstinence education on teen sexual activity, risk of pregnancy, and risk of sexually transmitted diseases. *J Policy Anal Manage* 27: 255-276.
2. Sather L, Zinn K (2002) Effects of abstinence-only education on adolescent attitudes and values concerning premarital sexual intercourse. *Fam Community Health* 25: 1-15.
3. Denny G, Young M, Rausch S, Spear C (2002) An evaluation of abstinence education curriculum series: Sex can wait. *Am J Health Behav* 26: 366-377.
4. Touns ML, Holmes WR (2002) Effectiveness of abstinence-based sex education curricula: A review. *Counseling and Values* 46: 237-240.
5. Santelli J, Ott MA, Lyon M, Rogers J, Summers D, et al. (2006) Abstinence and abstinence-only education: A review of U.S. policies and programs. *J Adolescent Health* 38: 72-81.
6. Dworkin SL, Santelli J (2007) Do Abstinence-Plus Interventions Reduce Sexual Risk Behavior among Youth? *PLoS Med* 4(9): e276.
7. Kirby D (2002) Do Abstinence-Only Programs Delay the Initiation of Sex Among Young People and Reduce Teen Pregnancy? Washington, DC: National Campaign to Prevent Teen Pregnancy.
8. Kempner M, Rodriguez M (2005) Talk About Sex. Sexuality Information and Education Council of the United States [SIECUS].
9. Josephson J (2005) Citizenship, same-sex marriage, and feminist critiques of marriage. *Perspect on Politics* 3: 269-284.
10. Underhill K, Operario D, Montgomery P (2007) Abstinence only programmes to prevent HIV infection in high income countries. *Cochrane Database Syst Rev* 4: CD005421.
11. Kirby D (2000) What does the research say about sexuality education? *Educ Leadership* 58: 72-76.
12. Nathan PE, Stuart SP, Dolan SL (2000) Research on psychotherapy efficacy and effectiveness: Between Scylla and Charybdis? *Psychol Bull* 126: 964-981.
13. Moore L, Moore GF (2011) Public health evaluation: which designs work, for whom, and under what circumstances? *J Epidemiol Community Health* 65: 596-597.
14. Weed S, Erickson I, Lewis A, Grant G, Wibberly K (2008) An abstinence program’s impact on cognitive mediators and sexual initiation. *Am J Health Behav* 32: 60-73.
15. National Abstinence Education Association (2011) Abstinence works: 2011. Washington, DC. NAEA.