



Does Dosage in Abstinence Education Matter?

Será que a Dosagem na Educação para a Abstinência faz a Diferença?

Thomas E. Smith¹, Burcu Atar², Teresa Ferreira³, Pamela Valentine⁴ & M.Graça Pereira⁵

1. Florida State University, FL
2. Hacettepe University, Turkey
3. Florida State University, FL
4. Southeastern Research Institute of Tallahassee, FL
5. University of Minho, Portugal

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

Abstract

Although past reviews uniformly criticized the efficacy and effectiveness of sexual abstinence in adolescents, new studies dispute the earlier findings. Studies that unpackage intervention programs provide one means of understanding why they succeed in some settings and not in others. This study examined 3183 students spread over 35 schools on the number of hours that they received in sexual abstinence education, in a context of health behaviors promotion. A multi-level analysis (HLM) was performed. The number of hours did not appear to make any difference in the outcome scores. Reasons for this finding are presented and their implications are provided.

Keywords: sexual abstinence education, effectiveness, HLM, dosagem

Resumo

Embora a literatura numa forma generalizada tenha uniformemente criticado a eficácia dos programas de abstinência sexual em adolescentes, novos estudos parecem contestar os resultados iniciais. Estudos que descompactam os programas de intervenção fornecem um meio de entender por que os programas de abstinência podem ter sucesso em alguns contextos e não em outros. Este estudo analisou 3.183 alunos distribuídos por 35 escolas sobre o número de horas que eles receberam em educação para a abstinência sexual, num contexto de promoção de comportamentos saudáveis. Foi realizada uma análise multi-level (HLM). O número de horas não pareceu fazer diferença nos resultados obtidos. As razões para esta conclusão são apresentados e suas implicações são fornecidas.

Palavras-Chave: educação para a abstinência sexual, eficácia, HLM, dosagem, processo de investigação

Does Dosage in Sexual Abstinence Education Matter?

Abstinence education continues to be a hotly discussed topic. Recent publications counter the decades of dismal results on the efficacy of abstinence education (e.g., Denny, Young, Rausch, & Spear, 2002; Toups & Holmes, 2002). However, it is unwise to simply dismiss previous studies that found abstinence education to be ineffective (e.g., Trenholm, Devaney, Fortson, Clark, Lisa, & Wheeler, 2008; Sather & Zinn, 2002). Other studies support the effectiveness of abstinence education (e.g., Denny, Young, Rausch, & Spear, 2002; Toups & Holmes, 2002). Proponents and opponents have conducted outcome studies that support their contentions (cf., NAEA, 2011; Kirby, 2002). Recent studies are also examining the mediators of effective abstinence-education programs (i.e., Weed, Erickson, Lewis, Grant, & Wibberly, 2008) as a means of unpacking programmatic

effects. Given the work of diligent scholars and practitioners, it is difficult if not impossible to ascertain the efficacy of abstinence (Kirby, 2000, 2002). Of course, some observers believe that conclusions about abstinence education efficacy have become a latent indicator of the political winds (cf., Santelli, Ott, Lyon, Rogers, Summers, & Schleifer, 2006).

Previous reviews have concluded that abstinence-only studies lack credibility because they fail standards of adequate efficacy research methods (cf., Underhill, Operario, & Montgomery, 2008; Kirby, 2000, 2002). The danger is that program effectiveness is being hidden by stringent research methodology. Thus, one strategy is to depart from traditional “box-score” reductionistic studies that provide little insight into effectiveness or its lack. However, it may well be that abstinence education researchers are faced with a dilemma. One option is to manualize and implement standardized

interventions that are unlikely to be replicated because of stringent methodological requirements. Another option is to attempt process research that examines already existing programs. In either option, the goal is to better understand “what works and why”.

There are a host of features that are of interest to program developers. One such feature is to ascertain the most effective number of sessions to be used in an educational intervention. There are currently a host of programs each with different duration and activities. Further, even when they are manualized, they may be implemented differently at different sites. Clearly, what is of interest is determining the optimal number of sessions.

Although they are by no means synonymous, the number of hours within an educational intervention can be likened dosage with brief therapy. The literature on dose-response is nearly twenty years

old. Howard, Kopta, Krause, and Orlinsky (1986) first discussed the dose-effect phenomenon in psychotherapy. They used a sophisticated methodology in concluding that gains were greater depending on the time spent in therapy. Other researchers have found that a large percentage of patients do not change to any significant degree by leaving services (Hansen, Lambert, & Forman, 2002). Draper and his colleagues (Draper, Jennings, Baron, Erdur, & Shankar, 2000) conducted a study with nearly 1700 students across 42 counseling centers and found that more sessions resulted more benefits.

Thus, the findings from a related area of research suggest that longer educational interventions may result in more benefits. Further, there may be questions on whether gender, ethnicity, or age affected program outcomes. Finally, a question exists on whether the length of programs differed across classrooms despite the implementation of a

manualized intervention. The goal of the study was to examine variables surrounding the amount of impact on adolescents enrolled in an abstinence education program.

Method

Participants

There are currently services being delivered to 35 schools with three full-time county coordinators. There are three full time county coordinators. 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 number of males and females. Three-quarters of the participants were Caucasian while the remainder were equally split between African-Americans and Hispanic students. The students are drawn from four counties in a southeastern state; there is a mix of different ethnic groups, ages, and equal distribution of boys and girls. Table 1 presents the descriptive characteristics of the sample.

Table 1

Descriptive Statistics for Outcome and Explanatory Variables at the Student and Classroom Levels

	Variable	<i>M</i>	<i>SD</i>
Student-Level Variables			
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	$(CLSIZE)_j$	28.28	16.27

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. As a result, training and procedures were developed. Second, the initial curricula were modified as necessary 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, in most instances, 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.

Fully unconditional HLM was used together preliminary information about the reliability estimate of overall classroom means of hours of intervention and the amount of variation in hours of intervention that lies within and between classrooms in the sample. The results of the analysis are given in Table X1. The reliability of the overall classroom means was estimated to be around 0.990. This reliability estimate indicates that the sample classroom means are quite reliable as indicator of the true classroom means. The high reliability justifies for further modeling (table 2).

Results

Fully Unconditional HLM

Table 2
Fully Unconditional HLM Results

Within-class variance ($\hat{\sigma}^2$)	0.528
Between-class variance ($\hat{\tau}_{00}$)	3.143
Reliability of mean post test score ($\hat{\lambda}$)	0.990
Intraclass correlation for between-class variability ($\hat{\rho}$)	0.856
Intraclass correlation for between-class variability, adjusted for reliability($\hat{\rho}_{adj}$)	0.857

The adjusted intraclass correlation, which represents the proportion of

variance in hours of intervention between classrooms, adjusted for reliability was

calculated to be around 0.857 using the following formula,

$$\hat{\rho}_{adj} = \hat{\tau}_{00} / (\hat{\tau}_{00} + (\hat{\sigma}^2_{\epsilon} \lambda))$$

This value indicates that about 86% of variance in hours of intervention is due to differences on mean hours of intervention among classrooms whereas about 14% of variance in hours of intervention is due to individual differences among students. The high intraclass correlation for between-class variability supports the use of HLM.

Unconditional Within-Class HLM

In the unconditional within-class model, the student hours of intervention was estimated as a function of adjusted mean hours of intervention, age, gender, race and two-way interaction of age and gender. While the adjusted mean hoursof intervention was modeled as a randomly varying parameter 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 resultsof the unconditional within-class model are presented in Table 2. The adjusted mean of hours of intervention over classrooms was estimated to be around 5.579 with a standard error of 0.154. It was found that the adjusted mean of hours of intervention vary statistically significantly among classrooms ($p < 0.001$), indicating that there are statistically significant differences on mean hours of intervention among classrooms. The average effect of age on hours of intervention was estimated to be around -0.079 and on average, the effect of age on hours of intervention was found to be statistically significant ($p < 0.001$). However, the effect size for the average age slope issmall ($ES = -0.109$). The average hours of intervention is expected to decrease about 0.109 standard deviation when the age increases 1 standard deviation, controlling for other

variables. The average gender gap in hours of intervention was estimated to be around -0.159 and the effect of gender on hours of intervention was found to be statistically nonsignificant ($p = 0.610$), indicating that there is not difference on the hours of intervention between males and females. Based on the effect size measure, it can be said that the average hour of intervention of males is about 0.219 standard deviation lower than that of females when other variables are controlled, reflecting a small effect.

Even though the results show that the average effect of age on hours of intervention is statistically significant, its interaction with gender was not found to have a statistically significant effect on

hours of intervention. For the race variable, the gaps between Whites and Hispanics was found to be statistically significant with small effect size whereas the gaps between Whites and Blacks and Whites and others in hours of intervention were found to be statistically nonsignificant with a very small effect size. It can be said that the average posttest score of Whites is about 0.078 standard deviation higher than that of Blacks, the average posttest score of Whites is about 0.179 standard deviation lower than that of Hispanics, and the average posttest score of Whites is about 0.015 standard deviation lower than that of others when the other variables are controlled (table 3).

Table 3*Unconditional Within-Class HLM Results*

Fixed Effect		Coefficient	SE	P	Effect Size
Adjusted mean hours	γ_{00}	5.579	0.154	< 0.001	-----
Mean age slope	γ_{10}	-0.079	0.021	< 0.001	-0.109
Mean gender slope	γ_{20}	-0.159	0.312	0.610	-0.219
Mean race (W vs. Blacks) slope	γ_{30}	0.057	0.046	0.209	0.078
Mean race (W vs. Hispanics) slope	γ_{40}	-0.130	0.042	0.002	-0.179
Mean race (W vs. Others) slope	γ_{50}	-0.055	0.064	0.391	-0.076
Mean age-gender interaction slope	γ_{60}	0.011	0.021	0.619	-0.015
<i>Random Effect</i>		<i>Standard Deviation</i>	<i>Variance</i>	<i>p-value</i>	
<hr/>					
Level-2 Error Term					
Adjusted classroom mean	u_{0j}	1.777	3.159	< 0.001	
Level-1 error term	r_{ij}	0.724	0.524		

When the within-class variance in the fully unconditional model ($\hat{\sigma}^2 = 0.528$) was compared with the within-class variance in the unconditional within-class model ($\hat{\sigma}^2 = 0.524$), the proportion reduction in variance or proportion variance explained at level-1 was calculated as to be around 0.008. It can be concluded that adding age, sex, race, and the interaction term as predictors of hours of intervention reduced the within-class variance by only 0.8%.

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 hours among classrooms. The results are given in Table 4. The effect of the class size on the adjusted mean hours was not found statistically significant and practically important.

Table 4*Conditional Between-Class HLM Results*

Fixed Effects		Coefficient	SE	p	Effect Size
Adjusted mean hours model					
Intercept	γ_{00}	5.579	0.153	< 0.001	-----
Class size slope		-0.010	0.007	0.170	-0.006
Random Effects	γ_{01}	SD	VAR	p	
Level-2 Error Term					
Adjusted classroom mean	u_{0j}	1.776	3.154	< 0.001	
Level-1 error term	r_{ij}	0.724	0.524	< 0.001	

Conclusions

Dosage varied among classrooms. Over three-quarters of the variation in hours of intervention (86%) was due to differences among classroom with the remainder (14%) being due to the students. Although this is not surprising, it does make an observers pause since the educational intervention was manualized and monitored for adherence. The results suggest that all precautions aside, dosage effects will need to be included in all study findings to control for the number of sessions. The multi-level analysis did allow inferences about the students

themselves. However, it should be viewed as a caveat for other studies that do not have the sample size to justify the use of HLM.

The remainder of the findings was used to examine how student characteristics accounted for variation in the hours of participation. It is not surprising that increased age resulted in decreased participation in the program. Other studies have documented the difficulty of maintaining participating by high school juniors and seniors. However,

given the small effect size, the importance of the finding is attenuated.

An interesting finding was that there were no significant differences in how gender affected the hours of participation. Perhaps, it can be concluded that males who participate do so as much as their female counterparts. The findings on ethnicity were intriguing. The study showed that Hispanics were significantly more likely than their White counterparts to participate more fully in the program. Because differences among the other ethnic groups were not statistically significant, it is unclear on how to interpret these results. Although much can be made of cultural differences in regards to sexual abstinence, it is premature to use study findings in that manner. However, the impact of the above variables still accounted for less than 1 percent of variance on the hours of participation. This speaks to the small effect sizes that were seen across the board.

Further, this finding was consistent across classrooms and schools. Possible explanations are not convincing. One explanation is that the first several sessions acted as an inoculation that promoted a change in attitudes. The remaining sessions simply buttressed the results. Yet another explanation was that the sexual abstinence program can truly be delivered in one session. This result flies in the face of published studies that found that single exposures to sexual abstinence messages were insufficient to lead to any lasting benefits. The difficulty then is in determining the just noticeable difference that is discussed in the literature on the dosage effect in psychotherapy.

References

Denny, G., Young, M., Rausch, S., & Spear, C., (2002). An evaluation of abstinence education curriculum series: Sex can wait. *American*

- Journal of Health Behavior*, 26(5), 366-377.
- Draper, M., Jennings, J., Baron, A., Erdur, O., & Shankar, L. (2000). Dose-effect relationships in brief therapy based on a nationwide college counseling center sample. *Research Reports of the Research Consortium of Counseling & Psychological Services in Higher Education*. Austin: The Research Consortium
- Hansen, N.B., Lambert, M.J., & Forman, E.M. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice*, 9, 329-343.
- Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist*, 41(2), 159-164.
- Josephson, J. (2005). Citizenship, same-sex marriage, and feminist critiques of marriage. *Perspectives on Politics*, 3(2), 269-284.
- Kempner, M., & Rodriguez, M. (2005). *Talk About Sex*. Sexuality Information and Education Council of the United States (SIECUS).
- Kirby, D. (2000). What does the research say about sexuality education? *Educational Leadership*, 58(2), 72-76.
- 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.

- McNeilly, C. (1991-07). The effects of psychotherapy: A reevaluation based on dosage. *Psychotherapy Research, 1*(1), 74-78.
- Moore, L., & Moore, G. F. (2011). Public health evaluation: which designs work, for whom, and under what circumstances? *Journal of Epidemiology and Community Health, 65*(7), 596-597.
- Nathan, P. E., Stuart, S. P., & Dolan, S. L. (2000). Research on psychotherapy efficacy and effectiveness: Between Scylla and Charybdis? *Psychological Bulletin, 126*, 964-981.
- National Abstinence Education Association (2011). *Abstinence works: 2011*. Washington, D. C.: NAEA.
- Santelli, J., Ott, M. A., Lyon, M., Rogers, J., Summers, D., & Schleifer, R. (2006). Abstinence and abstinence-only education: A review of U.S. policies and programs. *Journal of Adolescent Health, 38*, 72-81.
- Sather, L., & Zinn, K. (2002). Effects of abstinence-only education on adolescent attitudes and values concerning premarital sexual intercourse. *Family & Community Health: The Journal of Health Promotion & Maintenance, 25*(2), 1-15.
- Toups, M. L. & Holmes, W. R. (2002). Effectiveness of abstinence-based sex education curricula: A review. *Counseling and Values, 46*, 237-240

Trenholm, C., Devaney, B., Fortson, K., Clark, M., Quay, L., & Wheeler, J. (2008). Impacts of abstinence education on teen sexual activity, risk of pregnancy, and risk of sexually transmitted diseases. *Journal of Policy Analysis and Management*, 27(2), 255-276.

Underhill, K., Montgomery, P., & Operario, D. (2007). Sexual abstinence only programmes to prevent HIV infection in high income countries. *Cochrane Database Systemic Review*, 4(9).

Weed S, Erickson I, Lewis A, Grant G, & Wibberly K. (2008). An abstinence program's impact on cognitive mediators and sexual initiation. *American Journal of Health Behavior*, 32(1), 60-73.