tions to have sex, or on five measures of sexual behavior. The intervention had a small, positive impact among some groups on several attitudes related to sexual decision-making, on perceptions about the media’s presentation of sexual images, and on feelings of self-efficacy and intentions to refuse sex. These attitudinal shifts did not translate into positive behavioral changes. Moreover, at 17 months, the intervention had no significant and positive effect upon any mediating variable, upon sexual or contraceptive outcomes or upon pregnancy or STD rates.

Our findings raise an important question: Why did this evaluation reveal no behavioral impact at three months and no impact of any kind at 17 months, when the evaluation of PSI in Atlanta suggested behavioral change? Is it possible that we failed to detect significant positive outcomes? The answer to this question has important implications for recommendations about how to develop effective programs.

**Strengths and Limitations**

This evaluation had several strengths: It employed a strong design with random assignment, short- and long-term follow-up and large sample sizes with sufficient statistical power to detect programmatically meaningful effects. It also allowed for the evaluation of youth-led PSI and adult-led PSI in schools and adult-led PSI in community settings. Moreover, this evaluation accounted for most of the mediating variables that might be affected by educational interventions and, in turn, might affect the initiation of intercourse. Most of the scales employed had acceptable-to-high reliability, and most of the behavioral measures had high internal consistency, both within each survey and between surveys. Finally, we checked extensively for inconsistencies and removed individuals with discrepant data.

However, several limitations are also noteworthy. This study did not have a strict no-treatment control group. While youths in the control groups received whatever program or instruction was otherwise being offered, it typically did not cover human sexuality. In addition, a large majority of the youths in both the treatment and control groups in this study, like those in Atlanta, had previously received some other instruction about aspects of human sexuality at some time during their middle school years. Thus, we could not assess whether PSI was more effective than nothing; all we could evaluate was whether PSI had a significant impact when it was taught in addition to other limited instruction on human sexuality.

There are several measurement limitations that are also noteworthy. Although youths who overreported or underreported sexual activity are likely to have been randomly distributed between the intervention and control groups, some youths who reacted negatively to the program or who were rebelling against its messages may have disproportionately overreported their sexual behavior. On the other hand, it is possible that youths who participated in PSI began to see teenage sexual activity in a less favorable light, and consequently underreported their own sexual activity. In either case, however, it seems likely that youths would overreport or underreport at the three-month posttest rather than at the 17-month follow-up, when any program effects are likely to have diminished. Moreover, our data on rates of sexual behavior are consistent with those from other studies.

**The Measurement of Pregnancy**

The measurement of pregnancy is somewhat problematic. Several weeks or even several months may elapse between conception and the time a young woman receives results from a pregnancy test. Until that time, she may not know of a conception or may incorrectly believe she is pregnant when she is not. Males, on the other hand, may not know that they have caused a pregnancy unless their sexual partner tells them.

Overall, we feel it is unlikely that the interventions produced programmatically important effects that were not detected. In the context of a strong design and methodology, we examined many subgroups of youths and searched at length for significant, positive and consistent behavioral effects. We found insufficient change in the mediating variables to suggest that there could be significant change in behavioral outcomes, and the results were remarkably consistent in demonstrating that PSI did not produce desirable effects upon behavior. Finally, behavioral results frequently were not in the desired direction, were not programmatically significant and were not close to statistical significance.

**Replication Issues**

When programs are replicated and implemented broadly, they are not always replicated with high fidelity. Accordingly, the ways in which the California implementation differed from that of the Atlanta evaluation should be examined.

The scale of the implementation in California was dramatically larger than that in Atlanta, and contractors had to stretch their resources and capacities in order to deliver PSI to large numbers of youths in relatively short periods of time. This raises the possibility that elements of the program may not have been implemented with the same fidelity as in Atlanta. There are several ways in which the California implementation of...