

REVIEW DRAFT: COMMENTS WELCOME

**Evidence-B(i)ased Policy Deliberation: A Motivated Reasoning
Framework With Applications to the Sex Education Debates**

Norman A. Constantine^{1,2}, Petra Jerman¹, Alice X. Huang^{1,2}

¹ Center for Research on Adolescent Health and Development,
Public Health Institute, Oakland, CA

² School of Public Health, Division of Community Health and Human Development,
University of California, Berkeley, CA

Paper presented at the
29th annual Association for Public Policy Analysis and Management Research Conference
What *Else* Shapes Public Policy Analysis and Management?
November 10, 2007
Washington, DC

Comments and correspondence: nconstantine@berkeley.edu

Acknowledgements

We thank Dawn Richardson and Brent Sugimoto for interview and coding assistance, William Brown for general research assistance, Lynn Bretl for Minnesota document collection, Carmen Nevarez for insight and discussion, and Carol Weiss for her seminal work in this area and for consultation and review. Our research program on *Evidence Use and Appraisal in Public Health Policy* is funded by the W.T. Grant Foundation. Additional support for the parent survey was provided by The California Wellness Foundation.

Evidence-B(i)ased Policy Deliberation: A Motivated Reasoning Framework With Applications to the Sex Education Debates

Norman A. Constantine, Petra Jerman, Alice X. Huang

Abstract

In our research program on evidence use and appraisal in public health policy, we are extending the sociological framework developed by Carol Weiss in the 1980s to further examine the roles of social-cognitive processes in research evidence use and appraisal. Specifically, we seek to understand how implicit (rapid, intuitive, and preconscious) and explicit (slow, logical, and conscious) reasoning processes are employed by policy influentials in appraising and applying research evidence, and how these processes are influenced by (and influence) prior beliefs and information, moral values and judgments, ideology, and interests and obligations. While our model of evidence-b(i)ased policy deliberation is intended to apply more broadly, the sex education debates provide an example where moral values and judgments can be especially salient.

This paper presents a theoretical framework with its foundations in social and cognitive psychology, and also informed by research and theory from moral and political psychology, sociology, and philosophy of science. A series of analyses are reported that examine some of the ramifications of this theory. In particular, findings across cases and across data collection methods were compared to ascertain whether patterns of consistencies and differences are concordant with, or suggestive of modifications to, the theory.

Based on a content analysis of 983 archival documents across five types of sex education debates, we found several theoretically relevant patterns of differences and similarities. Overall, the finding that authors of documents favoring abstinence-only (AO) and those favoring comprehensive sex education (CSE) employed research evidence to a similar extent, while differentially employing information sources and referencing different values at different rates, was consistent with our theoretical expectations. We found that health and safety values prevailed over other values in the arguments made by authors of both types, and further, that CSE-leaning authors make reference to health and safety values about 50% more than did AO-leaning authors. CSE-leaning authors also favored the values of information and science by an even greater degree, although these two values were much less frequently cited by either type of author than were health and safety values. Values about marriage and about moral right and wrong didn't play a big role in the arguments made by either side, and didn't differ significantly by document direction.

A decreasing gradient of use of research evidence on program effectiveness was found from the generic debates down through the debates on federal legislation, state legislation, and dropping off steeply to local school district policy debates. At the same time, debates about local policy were more than three times as likely to include references to unsupported background beliefs about program content. The debate-type gradient of use for research quality heuristics mirrored the gradient of use for research evidence, decreasing from generic to federal to state, and with a steep drop off for local policy debates. Values, on the other hand, did not differ by debate type. Another finding of note was the absence of differences between documents produced by journalists and those produced by advocates. Additional findings from our parent survey of sex education preferences, and from our web-based sex education scenario study are also discussed in light of these findings.

These results provide encouraging initial support for our motivated reasoning model of research evidence use and appraisal. Based on comparisons of patterns of research evidence use relative to uses of background beliefs, values, and information sources, we have illustrated some of the complexity of the research evidence use process. In particular, the complex patterns of interactions seen across debate types, document types, and document direction reinforce the important roles played in sex education policy deliberations by values and background beliefs, together with choice of information sources, in addition to research evidence.

Evidence-B(i)ased Policy Deliberation: A Motivated Reasoning Framework With Applications to the Sex Education Debates

Norman A. Constantine, Petra Jerman, Alice X. Huang

Evidence-based medicine, science-based policy, research-based public health, evidence-based decision making ... the list goes on. These are the calls of the new century, especially appropriate in a time when many in the scientific community believe that the very foundation of science is under attack in the current political and ideological policy and funding environment.

But those who commendably argue for scientific evidence as a basis for public policy have still to fill in some of the details. *What counts as evidence? What makes evidence scientific, and when is evidence compelling? What happens when evidence is in dispute? What role do values and ideology and other social and psychological factors play in selecting and appraising evidence?* These and other related questions about the nature and use of evidence are often overlooked. And when they are addressed, intense disagreement among reasonable experts can result.

Consider, for example, a recent debate within the American Evaluation Association (AEA) over the U.S. Department of Education's priority on "scientifically-based evaluation methods." The heart of the issue was the Department's statement, with funding implications, that "evaluation methods using an experimental design are best for determining project effectiveness" (Federal Register, 2003:62446). AEA submitted a board-approved position statement to the Department, objecting to the blanket nature of this conclusion and discussing other options and contextual considerations to inform the selection of the "best method." Shortly thereafter, a group of prominent evaluation theorists and methodologists, including several former AEA presidents, submitted a competing statement endorsing the Department's priority and its conclusions regarding the superiority of randomized experiments. A sample of the sometimes rancorous debate that resulted can be found in the archives of AEA's EvalTalk listserv (<http://bama.ua.edu/archives/evaltalk.html>) under the heading "NOT the AEA statement on Scientifically Based Evaluation." One of the consequences of this debate was the resignation from the organization of a prominent former president and leading evaluation textbook author, who publicly stated his view that "AEA now has the same relationship to the field of evaluation as the flat earth society has to the field of geology" (Lipsey, 2004). When it comes to standards of scientific evidence, reasonable minds can differ, sometimes strongly.

A quarter century ago Weiss noted, in one of her seminal works on research utilization in policy making, that "Researchers bring not so much discrete findings as their whole theoretical, conceptual, and empirical fund of knowledge into the decision making process." (1980, p. 12). Writing about the power of rigorous case studies as an approach to attempting valid inferences outside the laboratory, Campbell similarly spoke of the extended networks of implications in which scientific evidence must be presented and evaluated (1989). More recently, from her cognitive-developmental psychology work on scientific reasoning and evidence appraisal, Koslowski has emphasized the importance of the network of evidentially-relevant background information to thinking in general and to scientific explanation in particular (Koslowski, 1996; Koslowski & Thompson 2002).

Philosopher of science Susan Haack employs a crossword puzzle metaphor to illustrate her similar view of scientific evidence as a tightly interlocking mesh of reasons well-anchored in experience:

What is scientific evidence, and how does it warrant scientific claims? That honorific usage in which "scientific evidence" is vaguely equivalent to "good evidence" is more trouble than it's worth. When I write of "scientific evidence" I mean, simply, the evidence with respect to

scientific claims and theories. Scientific evidence, in this sense, is like the evidence with respect to empirical claims generally -- only more so: more complex, and more dependent on instruments of observation and on the pooling of evidential resources. ... Scientific evidence, like empirical evidence generally, normally includes both experiential evidence and reasons, and both positive evidence and negative. It is complex and ramifying, structured ... more like a crossword puzzle than a mathematical proof. A tightly interlocking mesh of reasons (entries) well-anchored in experience (clues) can be a very strong indication of the truth of a claim or theory; that is partly why "scientific evidence" has acquired its honorific use. But where experiential anchoring is iffy, or where background beliefs are fragile or pull in different directions, there will be ambiguity and the potential to mislead. (Haack, 2003, chapter 3)

A common theme across these and other analyses of the nature of evidence and evidence use (e.g., Evans, 1989, Chinn and Brewer, 2001) is that “neither theory nor data alone is sufficient to achieve scientific success; each must be evaluated in the context of, and constrained by, the other” (Koslowski, 1996, p. 252).

Policy use of research evidence

The classic knowledge-driven model of research use in policy making is illustrated in Figure 1. It implies a linear sequence in which research generates evidence and knowledge that impel action (Hanney et al., 2003). While intuitively appealing, this model was found lacking by early pioneers in the study of research influence on public policy decision making (e.g., Deitchman, 1976; Weiss, 1977, 1979, 1980; Knorr, 1977; Hayes, 1982). In attempting to come to grips with disappointing findings of little immediate and direct influence of research on policy decisions, social scientists became aware that “other values and interests took precedence over data” (Weiss, 1991, p. 309).



Figure 1. Classic knowledge-driven model of research evidence use in policy making

A consensus among social scientists gradually developed that real-world policy use of research evidence is typically more complex, indirect, and/or socially or politically motivated than implied by the classic knowledge-driven model. As a result, the classic model has been supplanted by a modern typology of research use comprising three primary categories: *instrumental use* (research used in specific direct ways), *conceptual use* (indirect use of research involving the “enlightenment” of policy actors, sometimes referred to as “knowledge creep,” Weiss, 1980), and *symbolic use* (political use of research to legitimate predetermined positions). Research use is now generally seen as falling in one, or some combination, of these and other secondary types of use (Weiss, 1977, 1979, 1980, 1991; Knorr, 1977; Hayes, 1982; Beyer, 1997; Innvaer, et al., 2002; Hanney et al., 2003; Weiss, et al., 2005).

In further developing this typology, Weiss (1991) distinguished among three types of research products: *research as data* (most influential in situations of consensus on values and goals, and with a small number of distinct alternatives under consideration), *research as ideas* (most influential in the early stages of policy discussions, and when uncertainty is high), and *research as argument* (most influential when value conflict is high, and where “research can supply evidence that will reassure supporters, convince the undecided, and weaken rivals’ positions” (p. 309). These three research products loosely correspond to the three policy-use categories of instrumental, conceptual,

and symbolic, and further elucidate the complexity of the research-use process. It is noted, however, that all research use to some extent involves research as argument, if not for or against a specific policy position, then at least for the validity of the research conclusions presented (e.g., Abelson, 1995).

Appraisal of policy-relevant research evidence

Although policy actors often expect, and sometimes demand, that research evidence provide definitive answers to their questions, responsible researchers know better. Rarely do we find simple or conclusive answers to our questions about the effectiveness of a policy or program intervention intended to address a complex social problem. Yet well-designed and skillfully implemented evaluations can and often do provide compelling evidence in this regard. “To make best use of this evidence requires a healthy dose of constructive skepticism, as well as an understanding of common sources of ambiguous results and misleading conclusions that frequently pervade program evaluations” (Constantine & Braverman, 2004, p. 236). Perhaps Carter says it best:

There are no final solutions in the world of social change. ... Failed certainties in social science litter the landscape like so many elephant bones bleaching in the African sun. Honest hard scientists never claim final answers; good social scientists shouldn't either.
(2004, p. xlvi)

Research evidence is frequently assumed to be fundamentally objective -- or at least to be readily appraisable as sound versus unsound. In truth, however, research evidence is often ambiguous, and conclusions based on this evidence – as well as policy implications that might result – can be biased or misleading (Constantine & Braverman, 2004; Faust, 1984, Tesh, 1990). This suggests the need for careful and skillful *appraisal of evidence* by policy influentials and other research consumers, generally thought to involve critical reasoning. This view is represented by the rational model shown in Figure 2.



Figure 2. Rational model of research use involving reasoning (appraisal of evidence)

Few research consumers, however, have the training or inclination – or time – to perform comprehensive critical appraisals of each piece of research they use. A common criteria employed by research consumers to verify research quality is publication in a peer-reviewed journal. Peer review is a process whereby experts help judge the value of a work that they were not part of creating. While peer review does filter out some unpublishable work, much poor quality research survives (Altman, 1994, 2002, Constantine, 2007a, 2007b). This is due in part to a shortage of methodologically skilled reviewers. Even high impact medical journals are only able to provide statistical review for about a third of their published manuscripts, while the situation in the more prolific low-impact journals is likely to be worse (Goodman, et al., 1998). Another potential factor is reviewer bias, which has been found to be based on reviewers’ existing theoretical perspectives and prior beliefs (e.g., Mahoney, 1977).

It is not only public policy research that faces these challenges. Increasingly, medical and epidemiological researchers are openly questioning the research quality in their fields. The recent conclusion by Ioannidis that “most research findings are false for most research designs and for most fields” (2005, p. 699) led to a Wall Street Journal headline stating “Most Science Studies

Appear to be Tainted by Sloppy Analysis” (Hotz, 2007). After conducting a rigorous statistical modeling study of the truth probabilities of common types of studies, Ioannidis noted that:

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. (2005, p.696)

Sterne and Davey Smith put it colorfully:

In many ways the general public is ahead of medical researchers in its interpretation of new 'evidence.' The reaction to 'lifestyle scares' is usually cynicism, which for many reasons, may well be rational. Popular reactions can be seen to reflect a subconscious Bayesianism in which the prior belief is that what medical researchers, and particularly epidemiologists, produce is gobbledygook. (2001, p. 229).

While strong skepticism on the part of research consumers is sometimes found, the other side of the problem can be naiveté and unwarranted faith. Sherman (2003) warns that “increasing demands by government for evidence-led policy raise the risk that research evidence will mislead government rather than leading to an unbiased conclusion” and that ... “few consumers of research understand the statistical biases with which science must always struggle” (p. 6).

Reinforcing the often ambiguous nature of evaluation evidence and the potential for misinterpretation and disagreement, even among experts, Weiss and colleagues recently cautioned:

Although we advocate the use of evidence, we also realize that evaluation rarely settles an issue once and for all. (We examined) the evidence for five programs (Life Skills Training, Project Alert, Midwestern Prevention Project, Project Northland, and CASASTART) that were among the most frequently named school-based drug prevention programs on seven lists of best practice, including the SDFS (Safe and Drug-Free Schools) list ... Our systematic and detailed review of the evaluation evidence for these programs raised serious questions. Most of the approved programs do not have firm scientific support. Evidence about their effectiveness is problematic.” (Weiss et al., 2005, p. 28, commenting on Gruner, et al., 2007, which was in-press at the time).

The potential for conflict of interest is a primary concern in research evidence creation and use. One motivation suggested for biased conduct and interpretation of program evaluations is the real or imagined pressure on grantees and external evaluators to show positive results and definitive conclusions that will please the funders (Braverman & Campbell, 1989; Constantine & Braverman, 2004; Moskowitz, 1993). Conflict of interests can also arise from financial or other connections between program developers, program evaluators, and/or secondary reviewers of program evaluations who create lists of approved programs. These types of connections have been found to be ubiquitous in several fields. For example, Gorman and Conde (2007) analyzed 246 peer-reviewed published evaluations of 34 school-based “model” drug and violence prevention programs

appearing on a federal list of effective and promising programs (including the five programs that Weiss referenced above). They found high potential for conflict of interest, given that most of the evaluations they reviewed involved program developers—78% of the 246 evaluation articles included the program developer as a coauthor, and for only 11% of the 246 articles was the developer totally independent from the evaluation authors (p. 426-427). Similar potential conflicts of interest involving developer-evaluator-reviewer connections can be found across many other areas, ranging from sexual risk behavior program evaluations (e.g., Kirby, 2007, Kirby, et al., 1999) to drug trials conducted by the pharmaceutical industry (Melander, et al., 2003) to federally sponsored consensus reviews of epidemiological and clinical trial evidence in the development of nutritional recommendations (Taubes, 2007).

Values, ideology, and evidence

Values and ideology interact with evidence in complex and fundamental ways in shaping health promotion policy (Weiss, 1983; Tesh, 1990; Raphael, 2000). A *value* can be defined as a principle, standard, or quality considered worthwhile or desirable. Although both scientific values (regarding desirable standards of science) and social values (regarding worthwhile social principles and conditions) can interact with evidence, we primarily focus on social rather than scientific values, with special attention to moral values, i.e., values regarding the rightness or wrongness of an action.

In practice, all health promotion policy stems first from values (Seedhouse, 1997). Values help determine not only the health challenges addressed and the solutions sought, but the extent to which evidence is accepted and the types of evidence embraced (Raphael, 2000). While the effect of values on health promotion policy is not necessarily problematic, problems can arise when values are not explicitly recognized as such, and when the interactions between values and evidence in the policy arguments that are produced are not sufficiently transparent.

A diversity of positions has been presented on the appropriate interplay between facts or evidence and values in program evaluation and public policy. These range from a radical constructivist belief in the fundamental inseparability of values and knowledge claims (e.g., Guba & Lincoln, 1985), to Campbell's (1982) postpositivist commitment to a strict fact-value dichotomy. Yet even Campbell (1982, p 124) acknowledges that “we are such pervasive valuers that almost none of the facts of the world can be appraised without value connotations.”

Much of the above applies to ideology as well, with *ideology* defined according to Bunge (1998, p. 246) as “an explicit system of values and ideals concerning both the prevailing and desirable social order.” As Weiss (1983) notes, “People’s ideologies are sometimes carefully constructed and internally consistent, but they are often haphazard and makeshift. However weakly integrated they may be, they provide an emotionally-charged normative orientation that provides a basis for position taking” (p. 224). Weiss laments the difficulty in disentangling the unique role of research in policy making because of the constant interaction with ideology and interests, further arguing:

Most people – even highly informed and experienced officials – do not have coherent, detailed, comprehensive ideologies. They have something more akin to general predispositions ... which do not cover all situations or give very explicit direction for those they do cover. Thus people work out their ideological positions in dealing with immediate, concrete issues. ... Often people hold a number of conflicting, ideological positions at the same time. ... Ideologies tend to be partial and fluid. They are probably almost as susceptible to being influenced by research as they are to influence the response to research. (pp. 231-232).

In earlier times many sociologists and political scientists dismissed the study of ideology as unworthy. Some argued that the concept of ideology has been overused to the point of questionable utility (Bell, 1961), or at least muddled beyond recognition (Converse, 1964). In contrast, Jost (2006) more recently applied a cognitive-motivational analysis of political ideology and found that meaningful political and psychological differences do covary with ideological self-placement. He concluded that “although ordinary people by no means pass the strictest tests imaginable for ideological sophistication, most of them do think, feel, and behave in ideologically meaningful and interpretable terms.” (p. 667). Caprara and Zimbardo (2004) similarly provide a cognitive-affective perspective, and discuss the importance of values and ideology as political heuristics for voters seeking to assess their congruency with politicians.

Moral values and sexual ideology

Discussion to be added of moral values and moral psychology (Sweeder & Haidt, 1993; Shweder et al., 1997; Haidt, 2001; Haidt & Hersh, 2001; Haidt et al., 1993; Hauser, 2006, etc.), protected values (Baron & Leshner, 2000, Baron & Spranca, 1997), sacred values (Tetlock, 2003; Tetlock et al., 2000), and of sexual values and ideology (Luker, 2006; Constantine, 2007c; Irvine, 2002).

The abstinence-only versus comprehensive sexuality education debates

The current national, state, and local policy debates about funding and other support and mandates for abstinence-only versus comprehensive school-based sexuality education present a compelling illustration of rampant claims and counterclaims based on intertwined fact and value statements. Debates about sex education are largely shaped by national policy rhetoric, and can be especially emotional because they concern children and youth, and because they involve fundamental questions about the nature and purpose of sexuality itself. Indeed the sex education debates provide part of the foundation for society’s broader political and ideological conflicts (Irvine, 2002). This area provides abundant examples -- in research summaries, media reports, advocacy campaigns, and elsewhere -- of ambiguous and inadequately appraised evidence distilled and presented as if it were definitive, often after being filtered through ideological lenses (Constantine & Braverman, 2004). And these issues play out at all levels of policy – national, state, and local (Constantine, Slater, & Carroll, 2007; Irvine, 2002; Luker, 2006). The debate is not just of academic interest – prevalences of adolescent pregnancies, births, and abortions, as well as sexually transmitted infections, some of which are life threatening, are likely to be influenced by the funding and other policy decisions that result. And with prevalence rates still substantially higher in the United States than in other industrialized countries (Singh & Darroch, 2000), it is imperative that rational arguments and evidence-appraisal play a fundamental role in the deliberations and debates about how best to address these value-laden public health challenges.

Do virginity pledges cause virginity?

A compelling illustration of misleading but highly influential research is provided by a widely publicized Add Health study conclusion that virginity-pledge programs effectively delay initiation of sexual intercourse (Bearman and Brueckner, 2001). Complex statistical adjustments were used to reach this conclusion. Several qualifications regarding the program setting were appropriately discussed, most notably that to have an effect, the pledge must occur in a community of other pledgers that is neither too small nor too large relative to the total student population in the school. The authors, however, neglected sufficient consideration of plausible alternative explanations. Foremost among these was the likelihood of a substantial self-selection effect, that is, a pre-existing

disinclination to initiate sex is likely to have been a primary causal factor behind both signing the pledge and delaying intercourse.

The statistical adjustment procedures intended to remove the effect of self-selection were described in an appendix to the article, but this discussion was dense and largely impenetrable, making it develop for a reader to appraise the appropriateness of these adjustments. In any case the adjustments were both logically and statistically inadequate (see Pedhazur and Schmelkin, 1991, pp. 295-296, and Constantine, 2005 for a discussion on the futility of this type of adjustment). Instead, the researchers' conclusions regarding a pledge effect suggest a criterion for causality of *post hoc ergo propter hoc* ("after this, therefore because of this"), a fundamental fallacy of logic, known since classical times, that involves inferring a causal relation on the basis of correlation and temporality alone.

This study, nevertheless, generated extensive media coverage and policy discussion (see, for example, Boyle, 2000; Nesmith, 2001; Schemo, 2001; Willis, 2001) and had a substantial influence on federal sexuality education policy. Prior to this study, the U.S. Department of Health and Human Services had required as performance measures for the evaluation of federally-funded abstinence education programs "the proportion of program participants who have engaged in sexual intercourse" and the birth rate of female program participants (Federal Register, 2000). Two years later, on the heels of extensive media attention to Bearman and Brueckner's (2001) study, these sexual behavior and birth rate measures were replaced with the "proportion of youth who commit to abstain from sexual activity until marriage" (Department of Health and Human Services, 2002). Thus, virginity pledging has become the primary behavioral outcome to be measured.

In considering the question—Do virginity pledges cause initiation of sexual intercourse to be delayed?—the answer remains that they might or might not. This particular study added little or nothing to our knowledge of this wished-for effect, yet somehow became extremely influential in the sex education debates (Constantine & Braverman, 2004).

Reducing the Risk: A program that works?

As another example of fundamentally flawed yet highly influential research on sex education, consider *Reducing the Risk* -- a comprehensive sex education and risk behavior prevention curriculum designed to reduce the incidence of teen pregnancy and sexually transmitted infections (Barth, 1989). This curriculum was recognized by a Centers for Disease Control and Prevention review panel as a "Program That Works." This recognition required that a program had been "proven effective in reducing HIV risk behaviors" (Centers for Disease Control and Prevention, 2001). The "Programs That Work" designation was assigned to *Reducing the Risk*, based on an outcome evaluation (Kirby, Barth, Leland, and Fetro, 1991) that included thirty-two initial significance tests of risk-behavior outcome comparisons (involving nine behavioral outcome measures and various combinations of time of measurement and subgroups based on prior sexual experience levels). Of these, three were significant at the .05 alpha level, involving only the students who were sexually inexperienced at the time of program entry. Thirty-six additional outcome comparisons were conducted to test effects within gender, ethnicity, and risk-level subgroups for the three significant outcomes; of these additional tests two were significant at the .05 alpha level.

Up to this point, sixty-eight significance tests of potential differences between the intervention and control groups were conducted, yielding five statistically significant results. This is not very different from what would be expected by chance in the face of no program effects and given the specified statistical significance criterion of .05. Moreover, a reader who carefully studies the

article will be able to identify numerous additional comparisons that were tested but were not systematically reported, further compromising any conclusions of program effectiveness based on this study.

This is not to say that the evaluation showed Reducing the Risk to be ineffective. Absence of evidence is (usually) not evidence of absence. In fact, the Reducing the Risk evaluation yielded an overall pattern of differences between the intervention and control groups suggestive of potential behavioral effects within some subgroups that would be worthy of further study. Yet these evaluation results certainly do not justify conclusions of “proven effectiveness” according to the standards of evidence associated with the statistical significance testing approach employed, nor by the criteria of the federal review group that certified the program as one that works.

Given the widespread availability of ambiguous, misleading, biased, and poor-quality research, the modern proliferation of research reports and articles¹, and the challenges to most research-consuming policy influentials to fully understand many of the statistical and methodological (and epistemological) issues involved, it should be no surprise that research evidence is often misinterpreted or misused. And in public health areas with intense value and ideology components, such as sex education effectiveness, the potential for biased conclusions and misleading policy implications is likely to be amplified (Constantine & Braverman, 2004; Raphael, 2000; Tesh, 1990).

Heuristics, biases, and evidence

One potential contributor to misleading conclusions and questionable policy applications of research is the use of *judgment heuristics*, or rules-of-thumb, to guide the rapid and often intuitive appraisal of the validity and integrity of research-based evidence and conclusions. These phenomena have been repeatedly demonstrated among lay persons as well as scientists, medical decision makers, and other professionals, often leading to biased or erroneous conclusions, as judged by normative logical systems of rationality (e.g., Tversky & Kahneman, 1974; Nisbett, et al., 1983; Heath, et al., 1994; Gilovich, Griffin, & Kahneman, 2002; Kahneman, 2003).

It has been argued that *implicit filters* selectively focus policy makers’ attention on evidence consistent with their own existing beliefs (Weiss with Bucuvalas, 1980), and that evidence consistent with preferred conclusions tends to be examined less critically than evidence that contradicts preferred conclusions (Ditto & Lopez, 1992). Further, experimental studies have shown that appraisal of one-sided evidence will often lead to bias in favor of that side (e.g., Brenner, Koehler, & Tversky, 1996), yet people tend to implicitly judge one-sided arguments to be better than two-sided arguments, especially for controversial issues such as abortion (Baron, 1995).

Social psychologists studying *motivated reasoning* have found that the desire to arrive at a particular conclusion can bias reasoning through a complex set of cognition processes, such as biased access to beliefs, biased selection of statistical heuristics, and biased appraisal of evidence (Kunda, 1990). Yet this biased motivated reasoning occurs primary at a preconscious level as people attempt to construct a rational justification for their beliefs.

A classic study by Lord, Ross, and Lepper (1977) found that attitude polarization about complex social issues increased in the face of mixed evidence, via a process of “biased assimilation.” According to these authors:

¹ Miser (1999) has calculated that “If physicians would read two articles per day out of the six million medical articles published annually, in one year, they would fall 82 centuries behind in their reading.”

People who hold strong opinions on complex social issues are likely to examine relevant empirical evidence in a biased manner. They are apt to accept "confirming" evidence at face value while subjecting "disconfirming" evidence to critical evaluation, and as a result to draw undue support for their initial position from mixed or random empirical findings. (p. 2098)

As a result, the authors lament that:

... social scientists cannot expect rationality, enlightenment, and consensus about policy to emerge from their attempts to furnish "objective" data about burning social issues. If people of opposing views can each find support for those views in the same body of evidence, it is a small wonder that social science research, dealing with complex and emotional issues and forced to rely on inconclusive designs, measures, and modes of analysis, will frequently fuel rather than calm the fires of debate. (p. 2108).

Cognitive **dual processing system** theorists such as Evans (1989, 2002) and Stanovich (1999) maintain that much casual reasoning involves the preconscious, implicit, intuitive, and often heuristic reasoning system (involving personal rationality using Evans's current terminology, sometimes referred to as system 1 reasoning), rather than the slower and conscious and more resource-intensive logical system (involving normative rationality, sometimes referred to as system 2 reasoning). Evans emphasizes the fundamental role of prior beliefs in both systems.

Chen and colleagues (1999) examine motivated cognition from a social-psychological dual processing perspective, based in their heuristic-systematic model (Chaiken, 1980, 1987; Chaiken et al., 1989; Chen & Chaiken, 1999). In this theory motivation plays two roles: "**level** of motivation predicts whether heuristic or systematic forms of cognition will predominate in a given judgmental setting," whereas "**type** of motivation predicts the nature or direction of whatever cognition occurs." (Chen et al., 1999, p. 44). Three broad types of motivation are postulated in their multiple motive framework –**accuracy motivation**, the desire to reach an accurate judgment or conclusion through even-handed treatment of judgment-relevant information, **impression motivation**, the desire to reach judgments that promote one's social goals, and **defense motivation**, the desire to protect pre-existing self-definitional beliefs or material interests. A key characteristic of defense-motivated processing is selective use of heuristics:

Heuristics that have judgmental implications congenial to perceivers' existing beliefs are especially likely to be used, whereas incongenial heuristics may be ignored or disparaged. Hence, defense-motivated perceivers may rely on the same heuristics that accuracy-motivated perceivers rely on, but in a selective manner. ... Information that is congruent with one's existing beliefs, such as research supporting one's position on abortion, will be judged more favorably than incongruent information (while) incongruent information may be scrutinized in an effort to derogate its validity. (Chen et al., 1999, p. 45).

Studying early and middle adolescents from a cognitive-developmental dual processing perspective, Klaczynski (2000) reaches similar conclusions about theory-motivated reasoning biases, and the selective use of judgment heuristics versus analytic processing to evaluate evidence that is congruent or incongruent with an adolescent's belief systems. Klaczynski also considers metacognitive dispositions and epistemic beliefs and goals, to distinguish between adolescents who tend to be motivated by theory preservation versus those who tend to be motivated by knowledge acquisition.

Also from a cognitive-developmental perspective, Koslowski (1996) has argued for the importance of prior beliefs and background information in appraising evidence in causal and scientific reasoning. Adolescents as well as college students were found to treat theory and data as interdependent, rely on theory (or explanation) to decide which data to take seriously and which data to search for, and rely on data to reject, refine, and elaborate theory (Koslowski, 1996, Koslowski & Maqueda, 1993; Koslowski & Thompson, 2002). Theoretical background information, such as plausible explanatory mechanisms and plausible rival alternative explanations, influenced responses to evidence and was itself influenced by evidence. Yet participants were willing to search for disconfirming evidence, and when successful to use it to revise or reject theories (Koslowski, et al., 1989). This was in contrast to the accepted view about the ubiquitous nature of confirmation bias (e.g., Klahr & Dunbar, 1988; Lord et al., 1977; Wasson, 1960). Similarly, Petty & Cacioppo (1986) have emphasized that people engaged in motivated reasoning often can be forced to accept undesirable beliefs when their justification attempts are sufficiently constrained by evidence

The motivated reasoning model of research evidence use and appraisal

Considering the diverse bodies of relevant research and scholarship reviewed above, from the fields of cognitive, social, moral, and political psychology, sociology, political science, and philosophy of science, it becomes clear that the classic knowledge-driven model of research evidence use in policy making (Figure 1), together with the rational model of research use involving reasoning and appraisal of evidence (Figure 2), might be lacking some additional important components.

To propose a more complete and realistic model requires a variety of judgments of relevance and importance of different aspects of the various research bases. The model we propose is transdisciplinary², building on and attempting to integrate and apply and extend the foundational work of many others. At its core is social cognition, and more specifically motivated reasoning. We call it the *motivated reasoning model of research evidence use and appraisal* (see Figure 3). It provides a starting point to guide our own research on evidence-b(i)ased policy deliberations in the sex education debates, and perhaps will provide a useful conceptualization to other researchers and policy analysts in public health and other public policy fields. The development and details of this model are described more comprehensively by Constantine and Huang (2007).

This motivated reasoning model represents our understanding that policy influentials engaging in policy deliberations, just like the rest of us, are influenced by their desires to arrive at particular conclusions. These desires, we propose, often lead to biased appraisal of research evidence through a complex set of cognition processes, such as biased access to beliefs, biased selection of statistical heuristics, and biased appraisal of evidence (Kunda, 1990). Level of motivation and type of motivation are both important determinants of the nature of the biased cognitive processing that occurs (Chen & Chaiken, 1999). This motivated reasoning takes place primarily at a preconscious level as policy influentials attempt to construct a rational justification for their beliefs. And it has limits – just as people can be forced to accept undesirable beliefs when their justification attempts are sufficiently constrained by evidence (Koslowski, et al., 1989; Petty & Cacioppo, 1986), so too, we believe, can policy influentials. The fundamental roles of activators and filters such as moral

² We define transdisciplinarity according to Abrams (2006; building on Rosenfield, 1992) as involving “a shared conceptual framework that integrates and extends discipline-based concepts, theories, and methods to address a common research topic (and promotes) the development of novel conceptual models and empirical investigations that integrate and extend the concepts, theories, and methods of particular fields.” (p. 516).

values and social (and sexual) ideologies, perceived social norms, interests and obligations, and perceived importance, are all readily subsumed under a motivated reasoning framework.

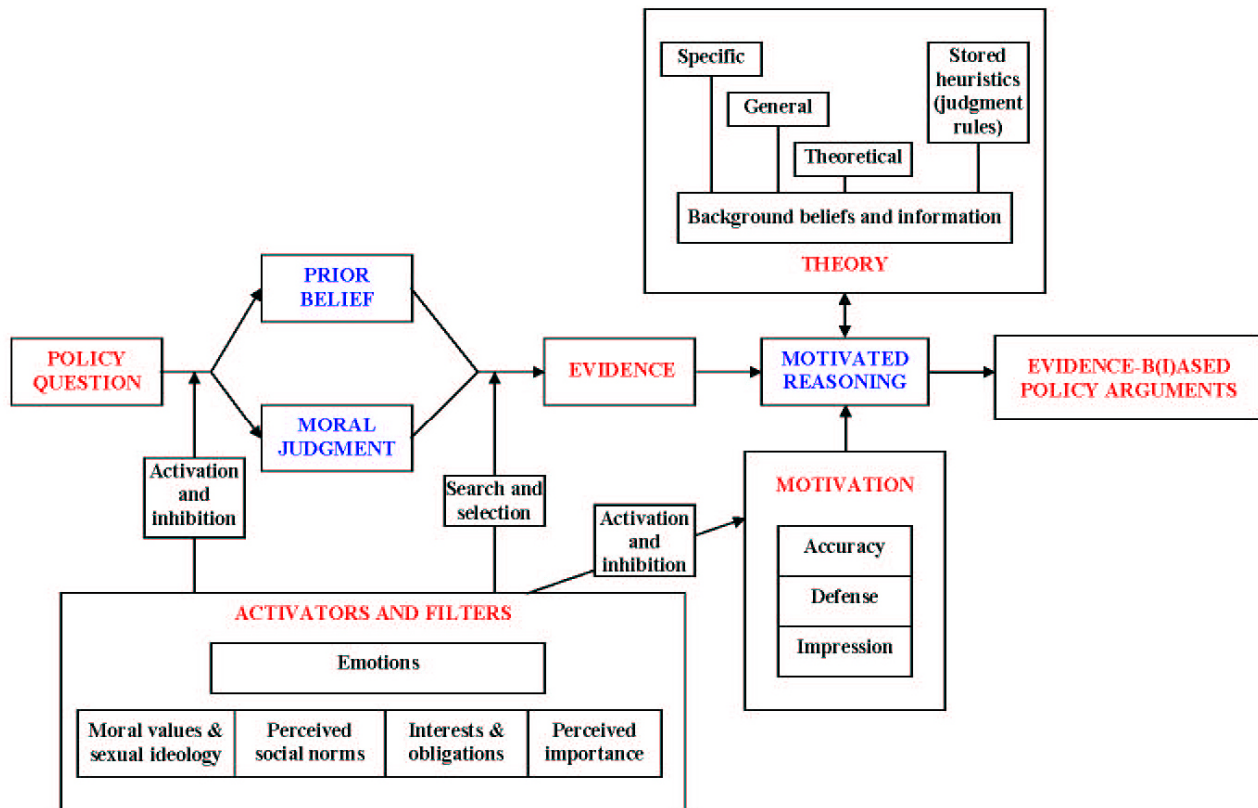


Figure 3. Motivated reasoning model of research evidence use and appraisal.

Goals and approach of this research

The research reported in this paper is intended as a first step in attempting to disentangle the uses of research evidence, background beliefs and information, and values and ideology in the sex education debates, and to provide fundamental insights that will also apply to other areas of public health policy deliberation, and to public policy in general. We approached this endeavor through a theory-based content analysis of archival media and advocacy documents applied across a variety of generic and specific cases in the sex education debates. Its theoretical foundation is our motivated reasoning model of research evidence use and appraisal.

Research questions. Three sets of research questions are addressed by the analyses reported below. **First**, what sources of information were used by journalists and advocates participating in the generic sex education debate and specific debates regarding federal and state legislation and local school district sex education policies? And to what extent were particular information sources differentially favored by document authors on each side of the debates? **Second**, to what extent were different types of research evidence employed in the debates, relative to the use of specific prior beliefs and values? And how did this vary by debate side? **Finally**, in debates about two nationally influential research studies on abstinence-only sex programs (the Bearman and Brueckner (2001) virginity pledge study and the Mathematica federal abstinence-only evaluation), to what extent were different types of research evidence employed in these debates, relative to the use of specific prior beliefs and values? And how did these vary by author's role and direction of argument?

Methods

The media and advocacy documents analyses reported in this paper are part of the larger study of evidence use in the sex education debates. The full study also involves additional quantitative and qualitative analyses of documents, interviews and focus groups with policy influentials, and related data collection and analysis activities. The full study was declared exempt by the Public Health Institute Institutional Review Board.

Document search

For the full study we searched for archival documents across five types of debate cases comprising twelve embedded specific debate cases, and four types of documents comprising eleven specific document types (see Figure 4 and Table 1). Documents were located through a combination of electronic searches using Lexis Nexis, Factiva, Access World News, ProQuest, Google, advocacy organization web sites, the Library of Congress Thomas web site, the GovTrack.us web site, congressional committee web sites, the California State Senate and Assembly web sites, and Minnesota newspaper web site archives.

Document types included (1) *legislative documents* (legislative bills, legislative analysis reports, committee testimony transcripts), (2) *public research reports* (including mainstream research articles or reports that were available to the general public, and researcher press releases and summaries); (3) *advocacy materials* (including media editorial articles, advocate sponsored or conducted research, advocate press releases and media advocacy, and other advocacy materials such as fact sheets and topic overviews); and (4) *media coverage* (newspaper, magazine, and television or radio news transcripts) at the national, state and local level.

Documents were selected that were published between January 1, 1995 and September 15, 2007, and related to sex education in the United States in general or specifically to national sex education legislation, state legislation in California or Minnesota, and sex education policy in two selected local school districts. Documents discussing the preference for or effectiveness of approaches to sex education, the need for some type of sex education, the content of sex education programs, or legislation involving sex education were retained. Documents that focused on the teen births, sexually transmitted diseases, lesbian, gay, or transgender youth issues, reproductive health services, emergency contraception, or abortion, but did not deal with sex education, were excluded. A total of 1,469 documents were retained (see Table 1).

As illustrated in Figure 4, four of the five debate-type cases included two or more embedded specific debate cases. Documents that generally discussed sex education without a focus on one of the embedded cases were categorized into (1) the *generic debates case* with no embedded cases. The *federal legislation debates case* included documents related to (2) the Title V, Section 510 abstinence education grant program debate, and (3) the Family Life Education Act (later reintroduced as the Responsible Education about Life Act). The *state legislation debates case* included documents related to (4) California's Comprehensive sexuality and HIV/AIDS Prevention Education Act (SB71), and (5) Minnesota's Comprehensive Family Life and Sexuality Education Bill (SB581) and various related comprehensive and abstinence-only sex education legislation that went under a variety of other bill numbers and names. The *local school district debates case* included documents from (6) the Mt. Diablo, California school district case, and (7) the Osseo, Minnesota school district case. The *research study debates case* included documents associated with two types of embedded cases – parent survey cases [(8) the Zogby parent survey case, and (9) the Kaiser/PSRA/NPR parent survey case], and influential program evaluation cases [(10) the

Virginity Pledge effectiveness study case, (11) the Mathematica abstinence-only program evaluation case, and (12) the Minnesota ENABL program evaluation case].

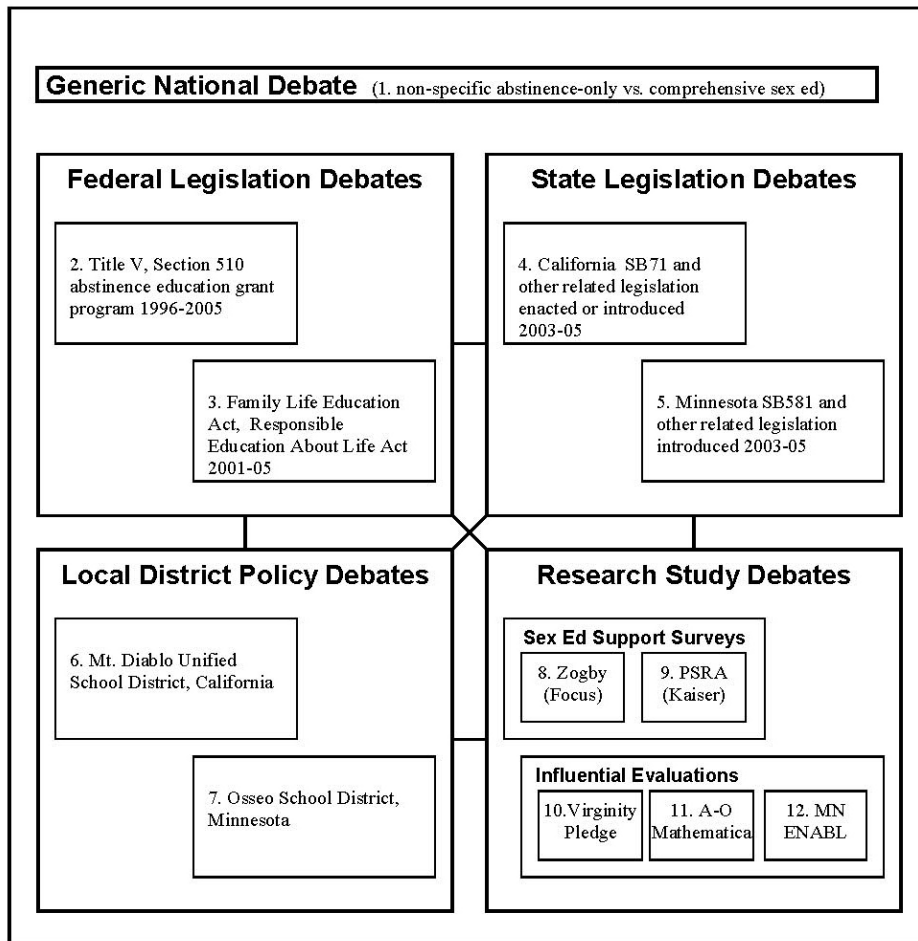


Figure 4. Multiple embedded case design.

Content Analysis

A computer assisted content analysis was conducted, employing a theory-guided coding dictionary with additional development emergent from the data (Neuendorf, 2002; Hart, 2001). The GREP expression-based coding system within ATLAS-ti, V5.0 (Muhr, 2004, pp. 144-190) was used to allow for context dependent content coding. The process of developing and testing the coding dictionary is described below.

An iterative consensus-development process among the three authors was employed to operationalize constructs representing components of arguments in the sex education debates, as informed by our transdisciplinary theoretical framework. This involved developing over time a comprehensive index and dictionary of codes and subcodes, and a series of manually conducted pilot coding tests beginning with a sample of twenty documents and progressing to 80 documents. During this period we were conducting interviews with policy influentials, and the interview results were informing the construct and code development decisions, while the documents analyses and coding specifications were informing the evolving interview protocols. Eighteen team meetings were held over a nine month interval to review and revise progressively more specific and reliable

construct codes and justifications. Appendix A contains the current coding index (the full 26-page coding dictionary is available upon request from the first author).

For the computer assisted content analyses reported in this article, we conducted a second consensus development process among the three authors and two additional graduate student research associates to further operationalize specific theoretical constructs. This involved building from the initial theory-based pilot coding results to specify and select (1) key words and phrases, and (2) detailed contextual conditions to allow for inclusion or exclusion of the key word or phrase. Over a series of six meetings these were developed and progressively refined, and results of a second series of pilot analyses involving a sample of 229 documents were reviewed and critiqued, with appropriate modifications made to the key words and phrases and the specified contextual conditions. One additional pilot test was then conducted using the full sample of 983 advocacy and media documents, to fine tune the inclusion and exclusion details, and to eliminate constructs with fewer than 100 total occurrences. At the conclusion of this process we retained three clusters of theoretical argument constructs (*evidence*, *background beliefs and information*, and *values*), and a final cluster of *sources of information* categories. As shown in Table 2, these comprised a total of twelve theoretical argument constructs and three sources of information categories, which in turn included 50 key words and phrases with appropriate contextual inclusion and exclusion criteria.

Four types of variables were created. *Code count variables* contain the raw key word and phrase code count within a document. *Summed construct code variables* represent the summed total of code counts within a document for each of the twelve theoretical constructs and three sources of information categories. *Standardized code variables* were designed to represent the rate of a key word or phrase per thousand words at the document level, i.e., these were computed by dividing each key word or phrase code count by the word count for the document, and then multiplying by 1,000. Finally, *summed standardized code variables* were computed for each of the twelve constructs and three categories, by summing all component standardized code variables within each construct or category. These were the primary variables used in the analyses.

In addition, each document was coded to represent *case* (12 cases as defined above), *case type* (5 types) *document type* (12 types), *document subtype* (5 subtypes), and *document direction* (abstinence-only, comprehensive sex education, or indeterminate or neutral). The analyses reported below included documents with either an abstinence-only (AO, n=368) or a comprehensive sex education (CSE, n=625) leaning position, within the advocacy (n=660) and media (n=323) document types. This involved 983 of the 1,426 total documents in the database (69%). As described below, most specific analyses employed smaller subsets of this total.

Statistical Analyses

Content coding results were first exported from ATLAS.ti into SPSS, V12.0. Next, summed code variables across key words were computed to represent the twelve theoretical constructs and three information source categories. These were then divided by document word count and multiplied by 1,000 to yield 15 standardized code count variables, representing rate of word use per thousand words.

SPSS was used to compute means and standard deviations for all variables within all categories, for the total sample of documents and for several subsets. A series of three-way univariate analyses of variance were conducted for each of the 15 standardized code count variables, with case type, document type, and document direction as between-document factors.

Results

Overall, across all five types of debates (generic, federal legislation, state legislation, local policy, and influential studies), twelve argument clusters, and three categories of information sources, *research evidence on program effectiveness* was the most frequently used of the 15 clusters, with 6.08 uses per 1,000 words. This was followed by *health and safety values* (4.97/1,000 words), *information sources supportive of comprehensive sex education* (3.27/1,000 words), *information sources supportive of abstinence-only sex education* (2.03/1,000 words), and *background beliefs and information on program content* (1.18/1,000 words). Figure 5 illustrates the use rates for each of the 15 clusters, in order of size.

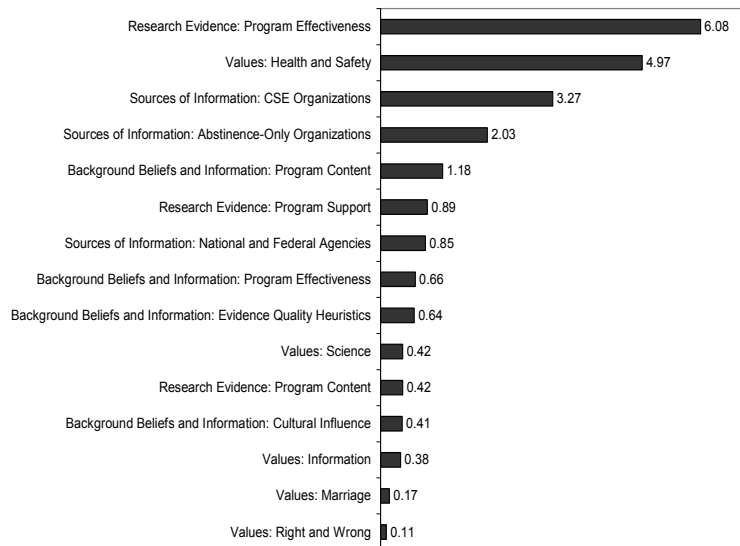


Figure 5. Total argument cluster use rates in order of size, across all debate types (number of documents=983).

We separately conducted more detailed analyses for the subset of documents representing the generic, federal legislative, state legislative, and local policy cases (number of documents=519, analyses 1 and 2), and for the subset of documents associated with the two influential research study cases (n=225, analysis 3). We also separated the analyses of the twelve theoretical argument clusters (analyses 1 and 3) from the analyses of the three sources of information categories (analysis 2). For all analyses, differences between media documents and advocacy documents, and differences between AO-leaning documents and CSE-leaning documents, were examined.

Analysis 1: Use of information sources in the generic and legislative sex education debates

These analyses addressed questions about sources of information used by journalists and advocates participating in the generic, federal legislative, state legislative, and local policy debates (number of documents=519), and the extent to which particular information sources were differentially favored by document authors on each side of the debates.

Within these debates the most frequently referenced category of sources of information overall was CSE-supportive organizations and persons (3.67 references per 1,000 words). This was followed by national and federal agencies (0.83), and by AO-supportive organizations and persons (0.72). Seven of the ten most highly cited specific sources were from the CSE-supportive category (Congressman Henry Waxman [.69], Advocates for Youth [.63], American Civil Liberties Union [.51], SIECUS

.50], Planned Parenthood [.47], Alan Guttmacher Institute [.28], and National Campaign to Prevent Teen Pregnancy [.20]), two are from the abstinence-only supportive persons category (Heritage Foundation [.32] and Focus on the Family [.18]) and one is a federal agency (Centers for Disease Control and Prevention [.44]); see Figure 6.

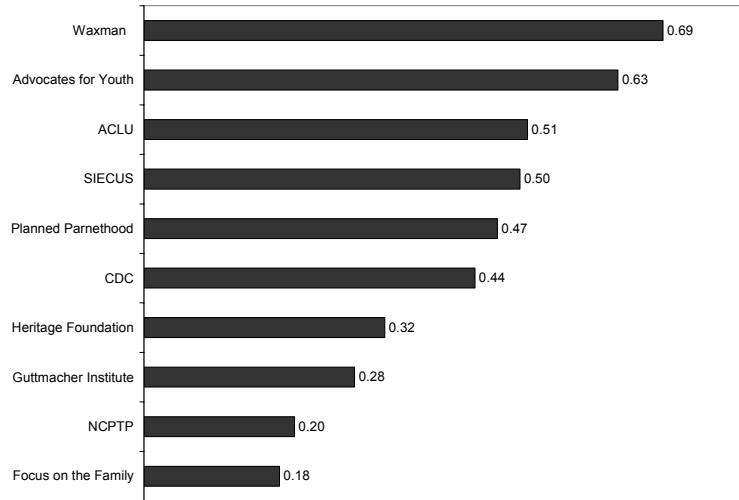


Figure 6. Most frequently cited specific sources of information.

The analyses yielded several theoretically relevant patterns of differences and similarities across the three cluster categories. A detailed examination of the between-category differences is provided below via the three-category analyses of variance for each of the twelve argument clusters. Bar charts are provided for each cluster with one or more argument clusters differing significantly by debate type, document type, or document direction.

Reference to AO-supportive organizations varied by debate type and also by document direction. In addition, references to national and federal agencies, and to CSE-supportive organizations varied by debate type. Clusters of information sources did not vary by document type.

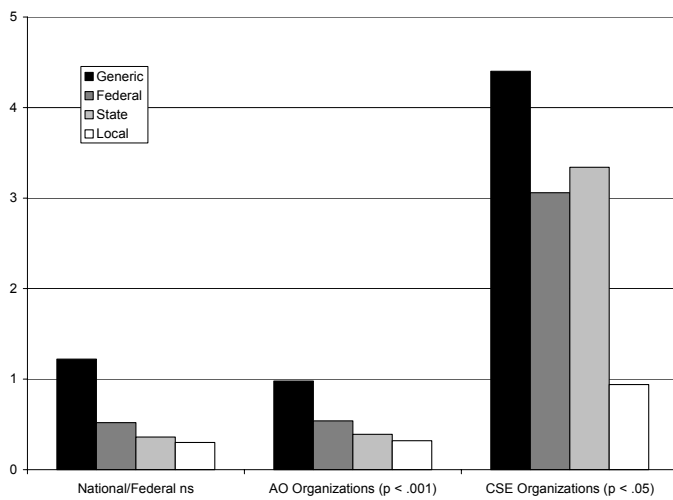


Figure 7. Sources of information by debate type, use rate per 1,000 words, within generic and legislative sex education debates.

Figure 7 illustrates that CSE-supportive sources were most frequently referenced in all debate types, and were referenced most in the generic debates and least in the local school district debates. Both national and federal agency sources, and AO-supportive sources, were also referenced most in the generic debate and least in the local school district debates.

Only one information source differed significantly by document direction (see Figure 8). AO-supportive sources were referenced more than six times as frequently in AO-leaning documents (3.31/1,000 words) than in CSE-leaning documents (0.28). On the other hand, CSE-supportive organizations were referenced as frequently in both AO-leaning and CSE-leaning documents (3.27 vs. 3.82, respectively).

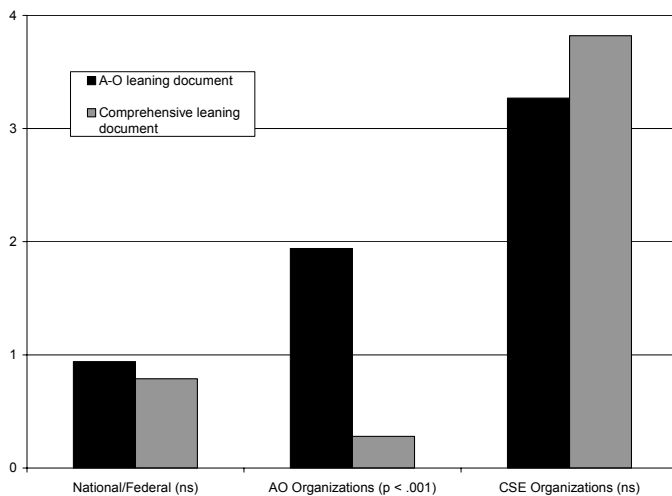


Figure 8. Sources of information by document direction, use rate per 1,000 words, within generic and legislative sex education debates.

Analysis 2: Use of research evidence in the generic and legislative sex education debates.

These analyses were restricted to the generic, federal, state, and local school district debates (number of documents=519), and to the twelve theoretical construct clusters. We analyzed use of the three types of research evidence, together with the associated theoretical construct clusters of background beliefs (four types) and values (five types).

Overall, research evidence on program effectiveness together with values regarding health and safety were the two most frequently used types of argument in these debates (see Figure 9). These were followed distantly by background beliefs about program content, program effectiveness, and research quality heuristics.

The analyses yielded several theoretically relevant patterns of differences and similarities across the three cluster categories. A detailed examination of the between-category differences is provided below via the three-category analyses of variance for each of the twelve argument clusters. Bar charts are provided for each cluster with one or more argument clusters differing significantly by debate type, document type, or document direction.

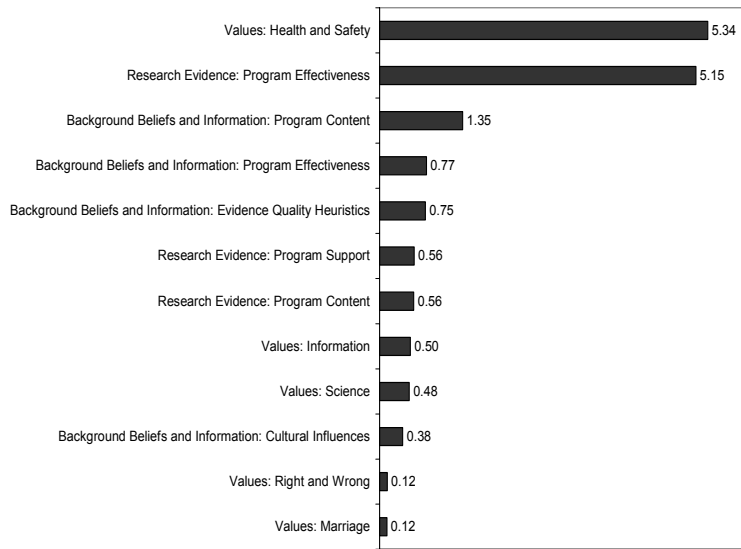


Figure 9. Theoretical argument cluster use rates in order of size, within generic and legislative sex education debates.

Research Evidence. Use of research evidence did not vary by document direction (AO-leaning or CSE-leaning) or by document type (media vs. advocacy), but did vary by debate type (debates about federal legislation, state legislation, local school district policy, and non-specific or generic sex education debates).

As shown in Figure 10, use of research evidence regarding program effectiveness and program content varied by debate type. Research evidence regarding program effectiveness was the most common, with a mean of 5.15 uses per 1,000 words across all types of documents. Documents associated with generic sex education debates were the most likely to use research evidence, while documents associated with debates about federal, state, and local sex education legislation or policies were less likely to do so. Documents associated with local school district policy debates were the least likely to use research evidence, showing 0.98 uses per 1,000 words.

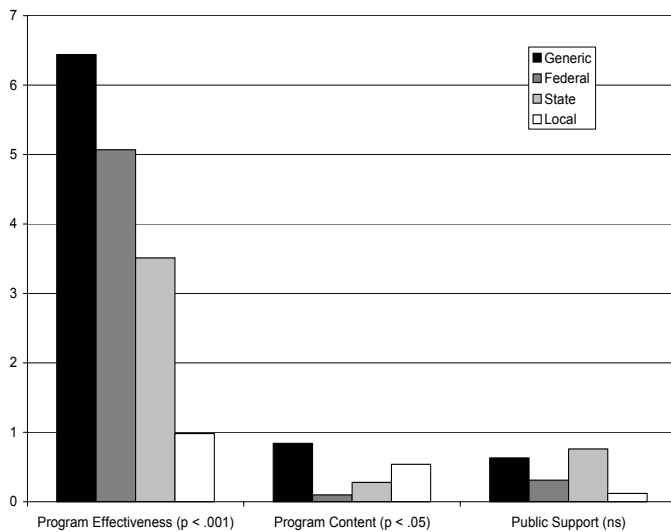


Figure 10. Research evidence use by debate type, use rate per 1,000 words, within generic and legislative sex education debates.

Research evidence regarding program content and public support were both equally common, with an overall mean of 0.56 uses per 1,000 words. Documents associated with the generic sex education debates, and local school policy debates, were the most likely to use program content evidence, while documents associated with debates about federal and state legislation or policies were less likely to do so. Documents associated with federal legislation debates were the least likely to use program content evidence, showing 0.10 uses per 1,000 words.

Background Beliefs and Information. Indications of background beliefs and information also did not vary by document type or document direction. In addition, program effectiveness beliefs and beliefs about cultural influences did not vary by debate type, but as is shown in Figure 11, program content beliefs and evidence quality heuristics did vary by debate type.

Program content beliefs were the most common beliefs indicated, with a mean of 1.35 uses per 1,000 words across all types of documents. Local school district policy debates were most likely to indicate program content beliefs, with 3.56 uses per 1,000 words, and federal legislation debates were the least likely to do so, with 0.65 uses per 1,000 words. Evidence quality heuristics were virtually non-existent within the local school district debate cases (0.04/1,000 words), and were most frequently indicated within the generic debate case (1.07/1,000 1,000 words).

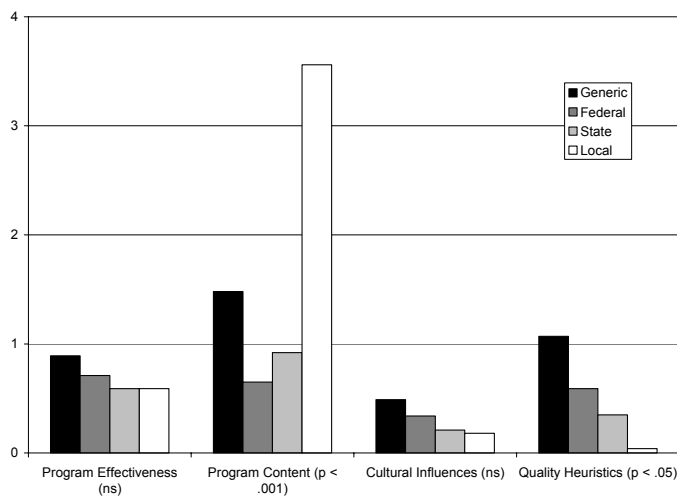


Figure 11. Background beliefs and information by debate type, use rate per 1,000 words, within generic and legislative sex education debates.

Values. References to values varied by document direction for the pragmatic values of health and safety, information, and science, but not for the traditional religious values of marriage, and right versus wrong. References also varied by document type for the health and safety value, but they did not vary by case type for any values.

The most frequently referenced value was health and safety, with an overall mean of 5.34 uses per 1,000 words (see Figure 12). Although this value was commonly found in all debate types, document types, and document directions, it was found significantly ($p < .001$) more within CSE-leaning documents (5.95/1,000 words) than within AO-leaning documents (3.67). The other four tracked values were all much less frequent, although the information and science values (both pragmatic values, together with health and safety) were also significantly more prevalent in CSE-leaning documents. The two traditional religious values of marriage and right versus wrong were the least common and did not differ significantly by document direction.

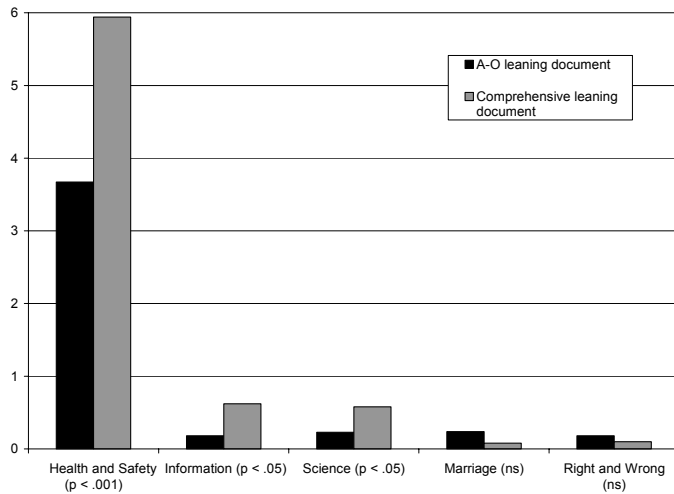


Figure 12. Values by document direction, use rate per 1,000 words, within generic and legislative sex education debates.

Analysis 3: Use of research evidence in the debates about influential AO research studies.

The third set of analyses was restricted to the debates about two nationally influential research studies on abstinence-only sex programs—the Bearman and Brueckner (2001) virginity pledge study (n=177), and the Mathematica federal abstinence-only evaluation (n=48). Both of these studies involved two major releases, the first being more favorable to abstinence-only, and the second less favorable to abstinence-only. As a result, earlier documents in the database for these cases were reacting to the AO-favorable releases, and later studies to the AO-unfavorable releases. For the present analyses the earlier and later documents were not differentiated within each of the two research study debates. We analyzed use of the three types of research evidence, together with the associated theoretical construct clusters of background beliefs (four types) and values (five types).

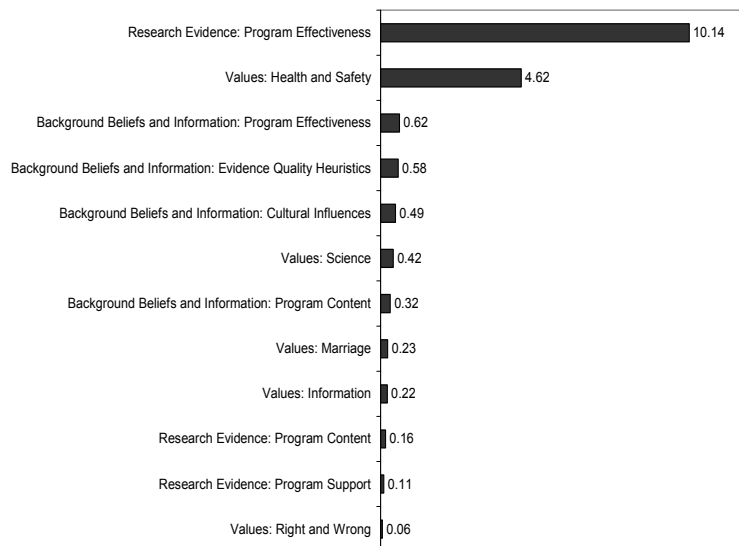


Figure 13. Theoretical argument cluster use rates in order of size, within nationally influential research study debates.

Overall, research evidence on program effectiveness was the most frequently used type of argument in these debates, with more than twice the use of the second most frequent cluster, values about health and safety (see Figure 13). These were followed distantly by research quality heuristics, background beliefs about program effectiveness, and values about science.

Research Evidence. As shown in Figure 14, only research evidence related to program effectiveness varied by debate type, with twice as much use in the Mathematica AO study debates as compared to the virginity pledge study debates (16.26 vs. 8.49).

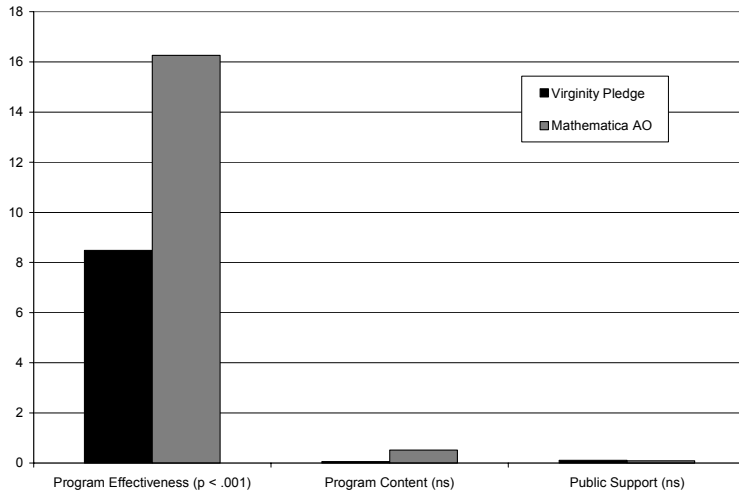


Figure 14. Research evidence use by debate type, use rate per 1,000 words, within nationally influential research study debates.

Use of research evidence on program content varied by document type, although this type of use by both document types was very low (0.26 for advocacy documents and 0.03 for media documents). No other research evidence use varied by document type (see Figure 15). No type of research evidence use varied by document direction.

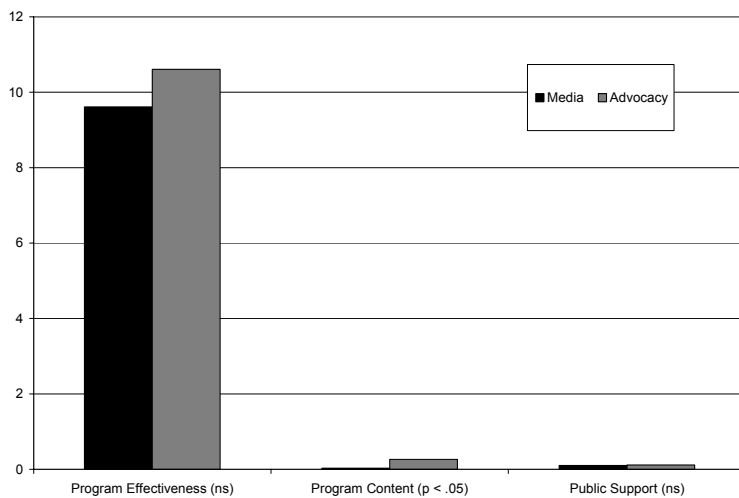


Figure 15. Research evidence use by document type, use rate per 1,000 words, within nationally influential research study debates.

Values. References to values did not vary by debate type for any values. By document type, advocacy documents showed more reference to valuing information (0.32) than did media documents (0.11). No other types of values differed by document type (see Figure 16).

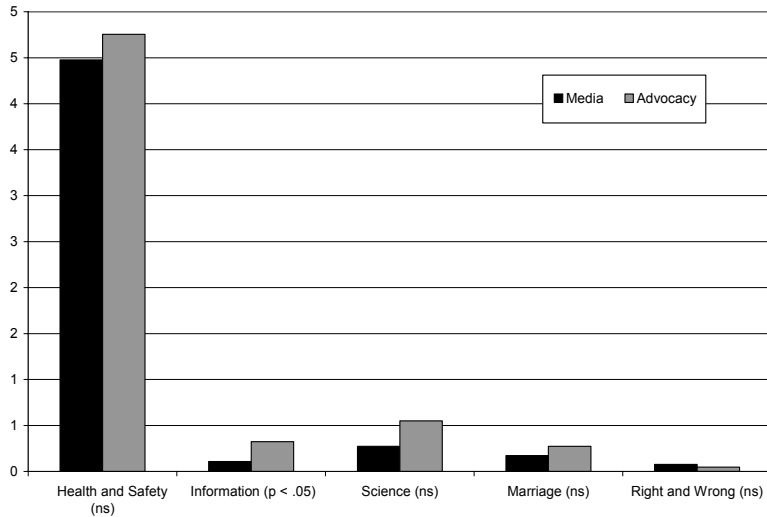


Figure 16. Values by document type, use rate per 1,000 words, within nationally influential research study debates.

As illustrated in Figure 17, only one value varied by document direction, with AO-leaning documents more frequently referencing the marriage value (0.40) than CSE-leaning documents (0.13).

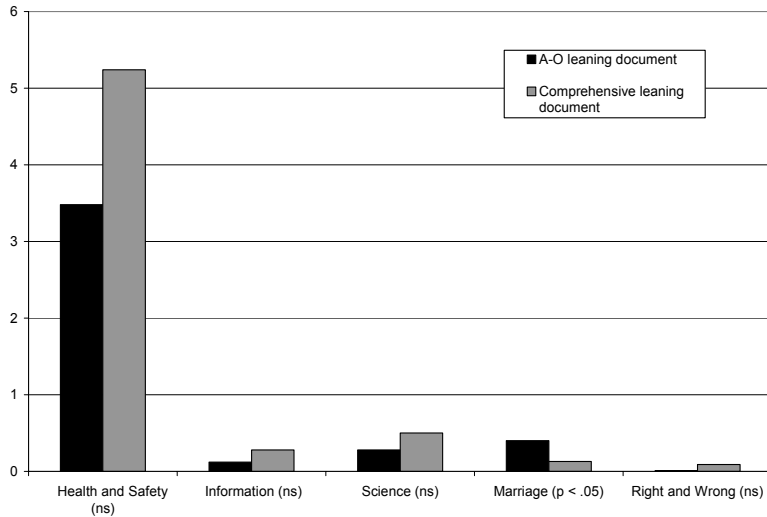


Figure 17. Values by document direction, use rate per 1,000 words, within nationally influential research study debates.

Discussion

The primary findings that authors of AO-leaning documents and CSE-leaning documents employed all three types of research evidence to a similar extent, while differentially employing information sources and referencing different values at different rates, was consistent with our theoretical expectations. Indeed, both sides in the debates were consistently appealing to research evidence, especially evidence regarding program effectiveness, and were taking advantage of the wide range of available research evidence. This evidence, albeit of variable quality, can be gleaned from numerous sources with sympathies to either CSE or AO, or from national and federal agencies. Authors writing AO-leaning documents were more than four times as likely to cite AO-supportive organizations and persons than were CSE-leaning authors, yet both sides cited CSE-supportive organizations and persons at about the same rate. Authors of CSE-leaning documents cited CSE supportive organizations more than three times as frequently as they cited either AO-supportive organizations, or national or federal agencies.

While health and safety values prevailed over other values in the arguments made by authors of both types, CSE-leaning authors made reference to health and safety values about 50% more than did AO-leaning authors. CSE-leaning authors also favored the other two pragmatic values of information and science by an even greater degree, although these two values were much less frequently cited by either type of author than were the health and safety values. Surprisingly, values about marriage and about moral right and wrong didn't play a big role in the arguments made by either side, and didn't differ significantly by document direction.

Other interesting difference emerged among debate types. A decreasing gradient of the use of research evidence on program effectiveness was seen from the generic debates down through the debates on federal legislation, state legislation, and dropping off very steeply to local school district policy debates. At the same time, debates about local policy were more than three times as likely to include references to unsupported background beliefs about program content. Furthermore, the debate-type gradient of use for research quality heuristics mirrored the gradient of use for research evidence, decreasing from generic to federal to state, and with a steep drop off for local policy debates. Types of values referenced, on the other hand, did not differ by debate type.

It is perhaps not surprising that the local policy debates were the least likely of the four debate types to involve research on program effectiveness, and the most likely to involve unsupported background beliefs on program content. It is not unreasonable to assume that local advocates and journalists are more limited in their access to and immersion in the research. If local debates are indeed being played out primarily in terms of program content beliefs, together with health and safety values, then they involve a fundamental different content dynamic than do the geographically and politically higher level and higher stakes debates. The similar gradient we found for research quality heuristics reinforces this view that less local debates tend to be more research oriented.

A further finding of note was the absence of differences between documents produced by journalists and those produced by advocates. No differences were found within the generic and policy debates analyses for any of the sources of information categories or the argument construct clusters between these two types of documents. And although there were small but significant differences in the influential research studies analyses for research evidence on program content and values regarding information (both favoring advocates), these differences were in the low frequency clusters.

It appears in this regard that journalists writing articles and other documents with a discernable position on the sex education debates were using similar sources of information and similar clusters of argument constructs to the advocates in this field. Some might argue that this follows from our selection for analysis of only documents with a directional leaning. The main reason for this selection was the dearth of documents available that did not lean in one direction or the other.

Given that CSE has become established as the mainstream position (e.g., it has been formally supported by most large mainstream medical, health, and education associations, such as American Academy of Pediatrics, American College of Obstetrics and Gynecology, American Medical Association, American Psychological Association, American Public Health Association, American School Health Association, National Association of School Nurses, National Association of School Psychologists, National Parent Teachers Association, Society for Adolescent Medicine), it seems appropriate to view journalists who write CSE-leaning documents as mainstream journalists rather than as advocates, and in fact, one could also note that CSE advocates themselves are advocating for the mainstream side of the issue. Given the absence of differences between journalists and advocates, together with the significant differences found between AO-leaning document authors and CSE-leaning document authors in information sources and values referenced, one hypothesis to pursue further would be that AO-leaning authors have a unique moral (values) and strategic (information sources) perspective that differentiates them from both the mainstream journalists and the similarly mainstream CSE advocates.

Related findings from the parent survey

Results from a survey study recently completed by our research group (Constantine, et al., 2007) appear to shed some light on this question. In this random digit dial telephone survey of 1,284 California parents, we found that 89% of parents supported comprehensive sex education and 11% supported abstinence only. If nothing else, this, along with similar findings nationally (Hoff et al., 2000; Bleakley, et al., 2006) and in North Carolina (Ito, 2006) helps establish the mainstream nature of the CSE position. Furthermore, self-identified liberal, moderate, and conservative parents all overwhelmingly supported comprehensive sex education (>70%, see Figure 18), as did self identified evangelical Christians (84%, see Figure 19).

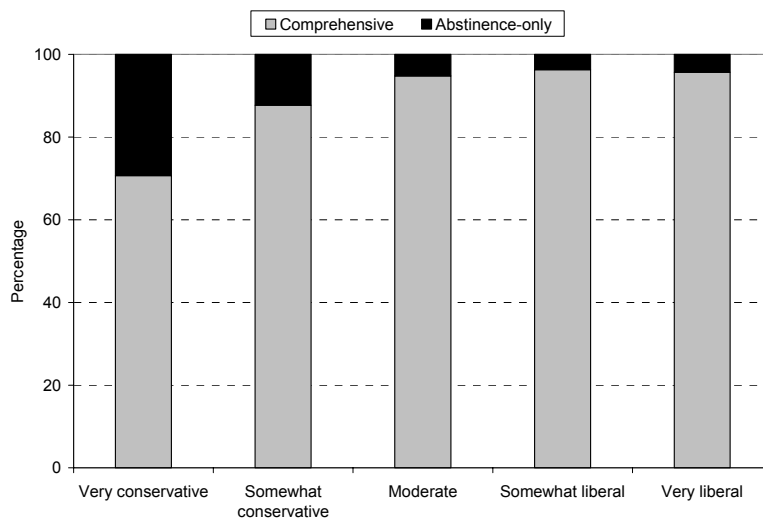


Figure 18. Percentage of support for sexuality education approaches by ideological leaning.

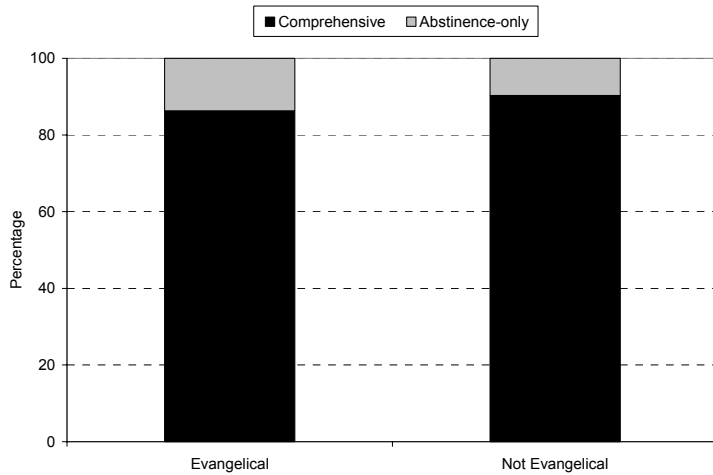


Figure 19. Percentage of support for sexuality education approaches by evangelical or not.

We found, however, one subgroup across all social, demographic, religious, and geographic categories examined that supported abstinence-only – those reporting deontological (absolutist) moral values rather than consequentialist (pragmatic) moral values in this area in response to open-ended questioning about the reasons for their preferences (see Figure 20). Furthermore, two-thirds of those supporting abstinence-only sex education gave moral absolutist reasons for their preference, while only 4% of those supporting comprehensive sex education gave moral absolutist reasons (see Figure 21).

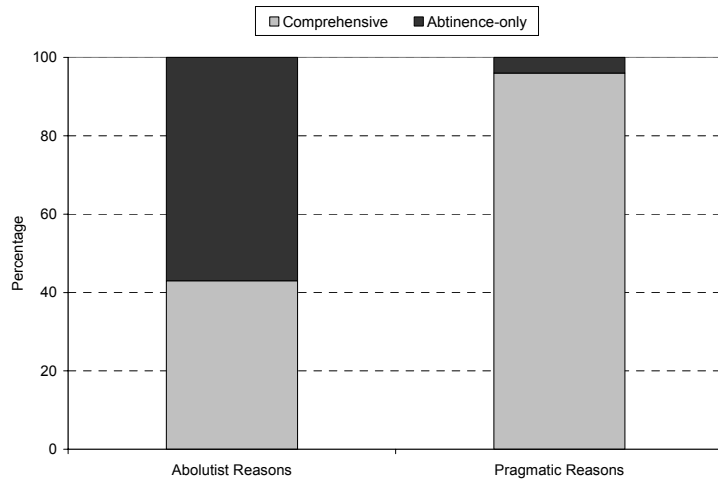


Figure 20. Sex education preference, by type of reason for preference.

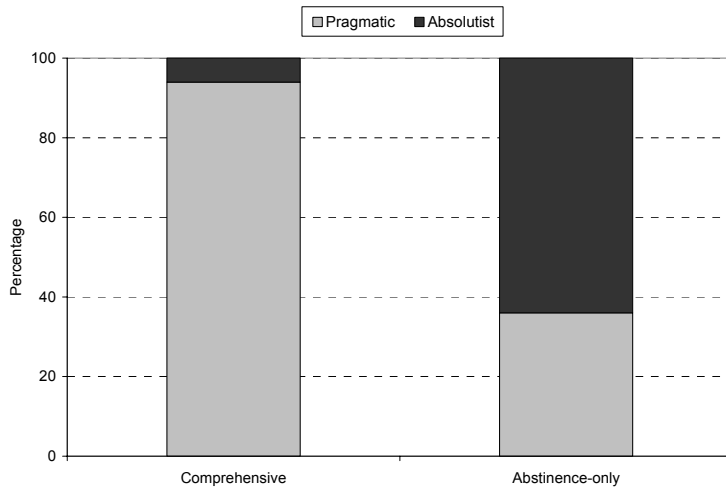


Figure 21. Pragmatic vs. absolutist reasons for policy preference, by type of preference.

These findings of fundamental differences in values orientation related to sex education policy between the small minority of AO supporters and the large majority of CSE supporters reinforce the importance of the differences in values references we found between AO-leaning and CSE-leaning documents. At the same time these differences corroborate our view of a pragmatic-values oriented mainstream majority engaged in debate with a small and uniquely absolutist-values oriented minority. Further analyses with these survey data of open-ended responses to additional questions about program effectiveness beliefs, and evidence in support of these beliefs, are planned to further investigate this issue.

Related findings from the web-based randomized scenario study

An additional ongoing study of our research group is also relevant, involving a web-based randomized scenario study of policy-related graduate students at two universities (Huang & Constantine, 2007). A sample of 232 University of California, Berkeley and Texas A&M University graduate students in public policy, public health, education, business, and engineering provided pragmatic, moral, and emotional judgments to a series of web-based scenarios about adolescent sexual behavior and sex education policy. Key contextual and evidentiary factors were randomly varied across scenarios.

One set of findings is especially relevant. As with the parent study, we found majority support for CSE among all political ideology groups. Of special interest is that conservatives' judgments of the *appropriateness* of a mandatory statewide comprehensive sex education policy scenario were influenced by hypothetical high-quality research evidence (on effectiveness of different types of sex education), while their *moral-acceptability* judgments about this policy were not. In contrast, both appropriateness and moral-acceptability judgments were influenced by evidence among liberals and moderates (see Figures 22 and 23, note also that for all groups and all conditions, the mean ratings of appropriateness and moral acceptability were at or above the neutral midpoint of the seven point rating scale). Again, this reinforces our postulation of a complex interaction between evidence use and values in the area of judgments on sex education policy. Additional analyses of judgments of other types of adolescent sexual behavior and sexual health policy scenarios, together with qualitative analyses of open-ended judgment justification responses, are currently underway.

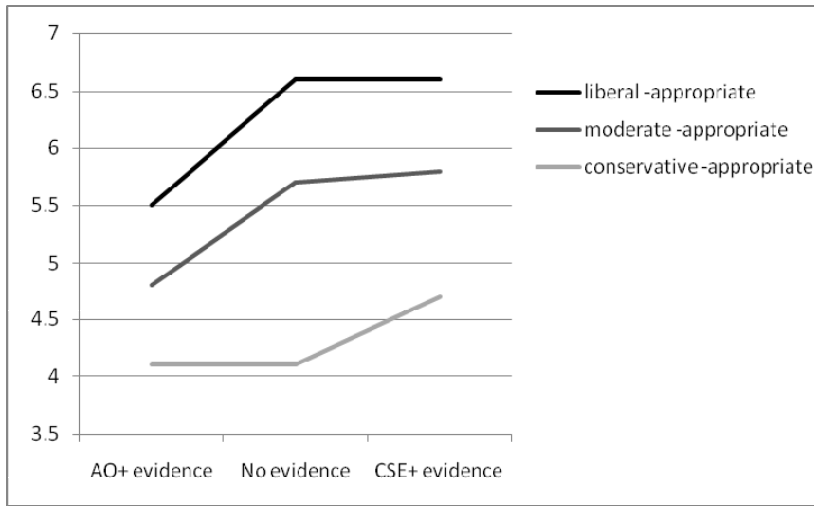


Figure 22. Appropriateness judgments (scale of 1 to 7) of a statewide comprehensive sex education policy, by hypothetical evidence presented

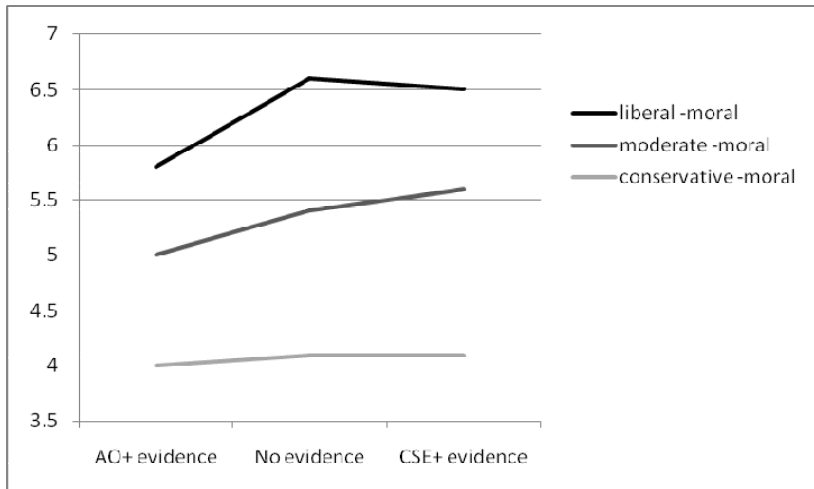


Figure 23. Moral acceptability judgments (scale of 1 to 7) of a statewide comprehensive sex education policy, by hypothetical evidence presented

Conclusion

The results reported and discussed above provide encouraging initial support for our motivated reasoning model of research evidence use and appraisal. Based on comparisons of patterns of research evidence use relative to uses of background beliefs, values, and information sources, we have illustrated some of the complexity of the research evidence use process. In particular, the complex patterns of interactions found across debate types, document types, and document direction reinforce the important roles played in these policy deliberations by values and background beliefs, together with choice of information sources, in addition to research evidence. The motivated reasoning model we propose readily accommodates these patterns. Ongoing and planned research will address additional aspects and ramifications of the model, employing different types and sources of data. The research reported in this paper provides a strong initial foundation for this further investigation.

References

- Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: LEA.
- Abrams, D. B. (2006). Applying transdisciplinary research strategies to understanding and eliminating health disparities. *Health Education & Behavior, 33*, 515-531.
- Altman, D. G. (1994). The scandal of poor medical research. *BMJ, 308*, 283-284.
- Altman, D. G. (2002). Poor quality medical research: What can journals do? *Journal of the American Medical Association, 287*, 2765-2767.
- Baron, J. (1995). Myside bias in thinking about abortion. *Thinking and Reasoning, 1*, 221-235.
- Baron, J., & Leshner, S. (2000). How serious are expressions of protected values? *Journal of Experimental Psychology-Applied, 6*(3), 183-194.
- Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes, 70*(1), 1-16.
- Bearman, P. S., & Brueckner, H. (2001). Promising the future: Virginity pledges and first intercourse. *American Journal of Sociology, 106*, 859-912.
- Beyer, J. M. (1997). Research utilization: Bridging the gap between communities. *Journal of Management Inquiry, 6*, 17-22.
- Bleakley, A. et al. (2006). Public opinion on sex education in US schools, *Archives of Pediatrics and Adolescent Medicine, 160*, 1151-1156.
- Braverman, M. T. & Campbell, D. T. (1989). Facilitating the development of health promotion programs: Recommendations for researchers and funders. In M. T. Braverman (Ed.). *Evaluating health promotion programs*. New Directions for Evaluation, no. 43. San Francisco: Jossey-Bass.
- Brenner, L. A., Koehler, D. J., & Tversky, A. (1996). On the evaluation of one sided evidence. *Journal of Behavioral Decision Making, 9*, 59-70.
- Bunge, M. (1998). *Social science under debate: A philosophical perspective*. Toronto, Ontario, Canada: University of Toronto Press.
- Campbell, D. T. (1982). Experiments as arguments. In E. R. House, S. Mathison, J. A. Pearsol, & H. Preskill (Eds.), *Evaluation studies review annual* (Vol. 7, pp.117-128). Beverly Hills, CA: Sage.
- Campbell, D.T. (1988). Evolutionary epistemology. In E. S. Overman (Ed.), *Methodology and epistemology for social science* (pp. 393-434). Chicago: University of Chicago press.
- Carter, H. C. (2004). Presidential perspective. In M. T. Braverman, N. A. Constantine, & J. K. Slater (Eds.), *Foundations and evaluation: Contexts and practices for effective philanthropy* (p. xlvii). San Francisco: Jossey-Bass

- Centers for Disease Control and Prevention (2001). *Programs That Work*. Atlanta: Author.
- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*, 752-766.
- Chaiken, S. (1987). The heuristic model of persuasion. In M. P. Zanna, J. M. Olson, & C. P. Herman (Eds.), *Social influence: The Ontario symposium* (Vol. 5, pp. 3-39). Hillsdale, NJ: Erlbaum.
- Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212-252). New York: Guilford.
- Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73-96). New York: Guilford Press.
- Chinn, C. A., & Brewster, W. F. (2001). Models of data: A theory of how people evaluate data. *Cognition and Instruction*, *19*, 323-393.
- Constantine, N. A. (2005). *Do virginity pledges cause virginity?* Retrieved October 31, 2007, from <http://crahd.phu.org/VirginityPledgesFAQ.html>
- Constantine, N. A. (2007a). The peer-review process. In S. Boslaugh (Ed.), *Encyclopedia of epidemiology*. Thousand Oaks, CA: Sage.
- Constantine, N. A. (2007b). Publication bias. In S. Boslaugh (Ed.), *Encyclopedia of epidemiology*. Thousand Oaks, CA: Sage.
- Constantine, N.A. (2007c). [Review of the book *When sex goes to school: Warring views on sex—and sex education—since the sixties*]. *Sex Education*, *7*, 441-445.
- Constantine, N. A., & Braverman, M. T. (2004). Appraising evidence on program effectiveness. In M. T. Braverman, N. A. Constantine, & J. K. Slater (Eds.), *Foundations and evaluation: Contexts and practices for effective philanthropy* (pp. 236-257). San Francisco: Jossey-Bass.
- Constantine, N. A. & Huang, A. X. (2007). *A motivated reasoning model of research evidence use and appraisal in public policy deliberation*. (in preparation).
- Constantine, N. A., Jerman, P., & Huang, A. X. (2007). California parents' preferences and beliefs on school-based sexuality education policy. *Perspectives on Sexual and Reproductive Health*, *39*, 167-175.
- Constantine, N. A., Slater, J. K., & Carroll, J. (2007). Motivational aspects of community support for comprehensive school-based sexuality education. *Sex Education*, *7*, 421-439.
- Deitchman, S. J. (1976). *Best-laid schemes: A tale of social research and bureaucracy*. Cambridge, MA: MIT Press.

- Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, *63*, 568-584.
- Evans, J. (1989). *Bias in human reasoning: Causes and consequences*. London: Erlbaum.
- Evans, J. (2002). The influence of prior belief on scientific thinking. In P. Carruthers, S. Stich, & M. Siegal (Eds.), *The cognitive bases of science*. Cambridge, United Kingdom: Cambridge University Press.
- Faust, D. (1984). *The limits of scientific reasoning*. Minneapolis: University of Minnesota Press.
- Federal Register. (2000). *65 Federal Register* 69562–65.
- Federal Register. (2003). *68 Federal Register* 62446.
- Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of human judgment*. Cambridge, United Kingdom: Cambridge University Press.
- Goodman S. N., Altman D. G., & George S. L. (1998). Statistical reviewing policies of medical journals: Caveat lector? *Journal of General Internal Medicine*, *13*, 753-756.
- Gruner, A., Murphy-Graham, E., Petrosino, A., & Weiss, C. H. (2007). The devil is in the details: Examining the evidence for “proven” school-based drug prevention programs. *Evaluation Review*, *31*, 43-74.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Haack, S. (2003). *Defending science – Within reason*. New York: Prometheus.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814-834.
- Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology*, *31*, 191-221.
- Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, *65*, 613-628.
- Hanney, S. R., Gonzales-Block, M. A, Buxton, M. J., & Kogan, M. (2003). The utilization of health research in policy-making: Concepts, examples, and methods of assessment. *Health Research Policy and Systems*, *1*, 1-28.
- Hart, R. P. (2001). Redeveloping diction: Theoretical considerations. In M. D. West (Ed.), *Theory, method, and practice in computer content analysis* (pp. 43-60). Westport, CT: Ablex.
- Hauser, M. (2006). *Moral minds: How nature designed our universal sense of right and wrong*. New York: Harper Collins - Ecco.

- Hayes, C.D. (1982). *Making policies for children: A study of the federal process*. Washington, DC: National Academy Press.
- Heath, L., Tindale, R. S., Edwards, J., Posava, E. J., Bryant, F. B., Henderson-King, E., et al. (1994). *Applications of heuristics and biases to social issues*. New York: Plenum Press.
- Hoff, T. et al. (2000). *Sex education in America: a series of national surveys of students, parents, teachers, and principals*. Kaiser Family Foundation.
- Hotz, R. L. (2007). Most science studies appear to be tainted by sloppy analysis. *Wall Street Journal*. September 14, 2007, p. B1.
- Huang, A. X. & Constantine, N. A. (2007). *Judgments on adolescent sex and sex education policy: the interacting influences of context, evidence, ideology, and moral values*. (in preparation).
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2, 696-701.
- Innvaer, S., Vist, G., Trommald, M., & Oxman, A. (2002). Health policy-makers perceptions of their use of evidence: A systematic review. *Journal of Health Services Research and Policy*, 7, 239-244.
- Irvine, J. M. (2002). *Talk about sex: The battles over sex education in the United States*. Berkeley: University of California Press.
- Ito, K.E. (2006). Parent opinion of sexuality education in a state with mandated abstinence education: Does policy match parental preference? *Journal of Adolescent Health*, 39, 634-641.
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58, 697-720.
- Kirby, D. (2007). *Emerging answers: Research findings on programs to reduce teen pregnancy and sexually transmitted diseases*. Washington DC: National Campaign to Prevent Teen Pregnancy.
- Kirby, D., Barth, R. P., Leland, N., & Fetro, J. V. (1991). Reducing the risk: Impact of a new curriculum on sexual risk-taking. *Family Planning Perspectives*, 23, 253-263.
- Klaczynski, P. A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development*, 71, 1347-1366.
- Klahr, D. & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12, 1-48.
- Knorr, K. D. (1977). Policymakers use of social science knowledge: Symbolic or instrumental. In C. H. Weiss (Ed.), *Using social research in public policy making* (pp. 165-182). Lexington, MA: Lexington Books.

- Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press/Bradford Books.
- Koslowski, B., & Maqueda, M. (1993). What is confirmation bias and when do people have it? *Merrill-Palmer Quarterly invitational Issue on "The Development of Rationality and Critical Thinking,"* 39, 104-130
- Koslowski, B., Okagaki, L., Lorenz, C., & Umbach, D. (1989). When covariation isn't enough: The role of causal mechanism, sampling method and sample size in causal reasoning. *Child Development*, 60, 1316-1327.
- Koslowski, B., & Thompson, S. (2002). Theorizing is important, and collateral information constrains how well it is done. In P. Carruthers, S. Stich, & M. Siegal (Eds.), *The cognitive bases of science*. Cambridge, United Kingdom: Cambridge University Press.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480-498.
- Lipsey, M. W. (2004). *The 2004 Claremont debate: Lipsey vs. Scriven. Determining causality in program evaluation & applied research: Should experimental evidence be the gold standard?* Retrieved October 31, from http://www.cgu.edu/include/SBOS_2004_Debate.pdf
- Lord, C., Ross, L., & Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37, 2098-2109.
- Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer-review system. *Cognitive Therapy and Research*, 1, 161-175.
- Melander, H., Ahlqvist-Rastad, J., Meijer, G., & Beermann, B. (2003). Evidence b(i)ased medicine: Selective reporting from studies sponsored by pharmaceutical industry. *BMJ*, 326, 1171–1173.
- Miser, W. F. (1999). Critical appraisal of the literature. *Journal of the American Board of Family Practice*, 12, 315-333.
- Moskowitz, J. M. (1993). Why reports of outcome evaluations are often biased or uninterpretable. *Evaluation and Program Planning*, 16, 1–9.
- Muhr, T. (2004). ATLAS.ti: The knowledge workbench (V5.0, 2nd ed.) [Computer software]. Berlin: Scientific Software Development.
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.
- Nisbett, R. S., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, 90, 339-363.
- Pedhazur, E. J. & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123-205). New

- York: Academic Press.
- Raphael, D. (2000). The question of evidence in health promotion. *Health Promotion International*, 15, 355-367.
- Rosenfield, P. L. (1992). The potential of transdisciplinary research for sustaining and extending linkages between the health and social-sciences. *Social Science & Medicine*, 35, 1343-1357.
- Seedhouse, D. (1997). *Health promotion: Philosophy, prejudice, and practice*. New York: Wiley.
- Sherman, L. W. (2003). Misleading evidence and evidence-led policy: Making social science more experimental. *The ANNALS of the American Academy of Political and Social Science*, 589, 6-19.
- Shweder, R. A., & Haidt, J. (1993). The future of moral psychology: Truth, intuition, and the pluralist way. *Psychological Science*, 4(6), 360-365.
- Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The "big three" of morality (autonomy, community, divinity) and the "big three" explanations of suffering. In A. M. Brandt & P. Rozin (Eds.), *Morality and health* (pp. 119-169). New York: Routledge.
- Singh, S., & Darroch, J. E. (2000). Adolescent pregnancy and childbearing: Levels and trends in developed countries. *Family Planning Perspectives*, 32, 14-23.
- Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.
- Sterne, J. A. C., & Davey Smith, G. (2001). Sifting the evidence – What’s wrong with significance tests. *BMJ*, 322, 226-231.
- Taubes, G. (2007). *Good calories, bad calories: Challenging the conventional wisdom on diet, weight control, and disease*. New York, NY: Knopf.
- Tesh, S. (1990). *Hidden arguments: Political ideology and disease prevention policy*. New Brunswick, NJ: Rutgers University Press.
- Tetlock, P. E. (2003). Thinking the unthinkable: sacred values and taboo cognitions. *Trends in Cognitive Sciences*, 7(7), 320-324.
- Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, 78(5), 853-870.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- U.S. Department of Health and Human Services. (2002, December). *SPRANS Community-Based Abstinence Education Program, Pre-Application Workshop*. Available at <http://www.mchb.hrsa.gov/programs/adolescents/abedguidetext.htm>. Accessed August 14, 2003.

- Wasson, P.C. (1960). On the failure to eliminate hypotheses on a conceptual task. *Quarterly Journal of Experimental Psychology*, 12, 129-140.
- Weiss, C. H. (1977). Research for policy's sake: The enlightenment function of social science research. *Policy Analysis*, 3, 531-546.
- Weiss, C. H. (1979). The many meanings of research utilization. *Public Administration Review*, 39, 426-431.
- Weiss, C. H. (1980). *Social science research and decision-making*. New York: Columbia University Press.
- Weiss, C.H. (1983). Ideology, interests, and information: The basis of policy positions. In D. Callahan & B. Jennings (Eds.), *Ethics, the social sciences, and policy analysis* (pp. 213-245). New York: Plenum Press.
- Weiss, C. H. (1991). Policy research: Data, ideas, or arguments? In P. Wagner, C. Weiss, B. Wittrock, & H. Wolman (Eds.), *Social science and modern states* (pp. 307-322). Cambridge, MA: Cambridge University Press.
- Weiss, C. H., & Bucuvalas, M. J. (1980). Truth tests and utility tests: Decision makers' frames of reference for social science research. *American Sociological Review*, 45, 302-313.
- Weiss, C. H., Murphy-Graham, E., & Birkeland, S. (2005). An alternate route to policy influence: How evaluations affect D.A.R.E. *American Journal of Evaluation*, 26, 12-30.

Table 1. Archival documents by case and document type, across all cases and documents

		Legislative Documents			Research Reports		Advocacy Materials			Media Coverage			Other
CASE	Total number of documents	Bills and revisions	Legislative analysis reports	Committee discussion and testimony transcripts	Mainstream article or research report	Researcher press releases & summaries	Advocates article or research report	Advocates press releases and media advocacy	Advocates other materials	National media	State or local media	Other media	Other relevant materials
1. Generic Debate	355				3	1	45	141	27	71	50	17	
2. Section 510	166	7		56	3		12	36	2	28	15	7	
3. FLEA/REAL	50	5					11	10	7	12	5		
4. CA SB71	118	13	15	7	2	1	6	26	7	3	35	3	
5. MN SB581	60	14		7			4	8	18	1	8		
6. CA school	53					1	1	1	25	1	7	4	13
7. MN school	25						1	3			8	6	7
8. Zogby Survey	169		1		3	2	13	111	19	3	5	12	
9. Kaiser Survey	66				3	1	5	8	8	14	15	9	3
10. Virginity Pledge	243				21	2	7	79	4	45	45	40	
11. Mathm. Eval.	63				4		4	31	1	12	4	7	
12. MN ENABL	58	1			1		12	21	2	5	9	6	1
TOTAL	1,426	40	16	70	40	8	121	475	120	195	206	111	24

Table 2. Argument clusters, argument constructs/categories, and key words and phrases.

<p>Research Evidence</p> <p>Program Effectiveness</p> <ol style="list-style-type: none"> 1. evaluate, evaluated, evaluation, evaluations [excluding any co-occurrence with: curricula or curriculum] 2. evidence [excluding any co-occurrence with: curricula or curriculum] 3. research [excluding: any co-occurrence with: curricula or curriculum] 4. studies, study [excluding: any co-occurrence with: curricula or curriculum] <p>Program Content</p> <ol style="list-style-type: none"> 1. curricula, curriculum [co-occurring with: evaluate, evaluated, evaluation, evaluations, evidence, found, report, research, publish, published, studies, or study] <p>Public Support</p> <ol style="list-style-type: none"> 1. support [excluding: supported, supporter, financial support] 2. poll [co-occurring with: adults, parents, students, teachers, or young people]
<p>Background Beliefs and Information</p> <p>Program Effectiveness</p> <ol style="list-style-type: none"> 1. effective, effectiveness, ineffective, ineffectiveness [excluding any co-occurrence with: evaluate, evaluated, evaluation, evaluations, evidence, found, report, research, publish, published, studies, study, condom, condoms, contraception, and contraceptive] <p>Program Content</p> <ol style="list-style-type: none"> 1. curricula, curriculum [excluding any co-occurrence with: evaluate, evaluated, evaluation, evaluations, evidence, found, report, research, publish, published, studies, or study] <p>Cultural Influences</p> <ol style="list-style-type: none"> 1. culture 2. media <p>Evidence Quality Heuristics</p> <ol style="list-style-type: none"> 1. peer-review, peer-reviewed 2. systematic 3. independent [excluding: independent of] 4. scientific, scientifically, scientist, scientists [excluding: scientific integrity, scientific objectivity, scientific culture, scientific distortion, scientific tradition, scientific question, scientific realities, scientific reality, scientific fact, scientific facts]
<p>Values</p> <p>Health and Safety</p> <ol style="list-style-type: none"> 1. health 2. protect 3. safety <p>Information</p> <ol style="list-style-type: none"> 1. informed [excluding: informed the, and informed consent] 2. information [co-occurring with: critical and vital] <p>Science</p> <ol style="list-style-type: none"> 1. science [excluding: science adviser]

Marriage

1. faithful
2. for marriage

Right and Wrong

1. morally
2. wrong

Specific Sources

National and Federal Agencies

1. CDC, Centers for Disease Control
2. GAO, General Accounting Office
3. IOM, Institute of Medicine, National Academies, National Academy, and National Research Council
4. NIH, National Institutes of Health, National Institute of Health
5. Surgeon General
6. YRBS

AO organization/person/study

1. Bearman
2. Family Research Council
3. Focus on the Family
4. Heritage Foundation, Rector
5. National Abstinence Clearinghouse
6. Zogby

CSE organization/person/study

1. ACLU, American Civil Liberties Union
2. Advocates for Youth
3. Guttmacher
4. Kirby
5. Kaiser
6. NCPTP, National Campaign to Prevent Teen Pregnancy
7. PHI, Public Health Institute
8. Planned Parenthood
9. SIECUS, Sexuality Information Council of the United States
10. Waxman

Appendix A: Coding Index
(The complete coding dictionary is available from the first author)

<p>1. Research evidence 100 Research evidence (temporary) 11X <u>Program effectiveness</u> 111X Program effectiveness positive 1111 Direct 1112 Indirect 112X Program effectiveness null 1121 Direct 1122 Indirect 113X Program effectiveness negative 1131 Direct 1132 Indirect 12X <u>Program content</u> 121 Direct 122 Indirect 13X <u>Need for an intervention</u> 131 Direct 132 Indirect 14X <u>Public support</u> 141 Direct 142 Indirect</p> <p>2. Reasoning 200 Reasoning (temporary) 211 Logical 212 Intuitive 213 Interpretative 214 Pragmatic 215 Analogy-based</p> <p>3. Values 300 Values (temporary) 31X <u>Consequentialist</u> (moral) 311 Health and safety 312 Information 32X <u>Deontological</u> (moral) 321 Gender roles 322 Information 323 Marriage 324 Obedience 325 Religion 326 Sexual behavior expectations 327 Sexuality</p>	<p>33X <u>Mixed or other</u> 331 Choice 332 Diversity 333 Family role 334 Responsibility 335 Self-actualization 336 Supporting children 337 Science 338 Separation</p> <p>4. Background beliefs and information 400 Background beliefs and information (temporary) 41X <u>Specific</u> 411X Program effectiveness 4111 Program effectiveness positive 4112 Program effectiveness null 4113 Program effectiveness negative 412 Program content 413 Need for an intervention 414 Public support 415 Legislative information 42X <u>General</u> 421 Adolescent characteristics 422 Social influence 423 Environmental context 43X <u>Heuristics</u> 431 Reputation and plausible motivation for or protection from bias 432 Research quality 44X <u>Theoretical explanation</u> 441 Plausible explanatory mechanisms 442 Plausible rival explanations 443 Theoretical information</p> <p>5. Affect 511 Affect</p> <p>6. Social norms and cultural influences</p> <p>7. Interests and obligations</p> <p>8. Motivational goals 800 Motivational goals (temporary) 811 Accuracy 812 Defense 813 Impression</p> <p>9. Temporary codes 911 Uncategorized</p>
--	---

0. Document and argument designations

01XX Document case

Definition: Each document is designated by the case from which it originates.

Code: 0101 Generic, 0102 Sec510, 0103 FLEA/REAL, 0104 SB71, 0105 MN, 0106 MtDiablo, 0107 Osseo, 0108 Zogby, 0109 PSRA, 0110 Virginitiy pledges, 0111 Mathematica, 0112 MNENABL

02XX Document type

Definition: Each document is designated by its classification in the archive, as legislative, research, advocacy, media coverage, or other type of document.

CATEGORIES AND SUBCATEGORIES USED:

Legislative Documents:

- LegislativeB (bills and revisions)
- LegislativeA (analysis reports)
- LegislativeT (discussion and testimony transcripts)

Research Documents:

- ResearchM (mainstream articles or research reports; not published by an advocacy group)
- ResearchP (researcher press releases and summaries)

Advocacy Materials:

- AdvocacyR (any report or article published by an advocacy group)
- AdvocacyM (press release and research summaries by advocacy groups, and media advocacy including editorials and letters to editor)
- AdvocacyO (advocacy materials that did not fit in any other advocacy category)

Media Coverage (non-advocacy):

- MediaN (large national newspapers such as NY Times, Washington Times)
- MediaSL (state and local newspapers)
- MediaO (media coverage that did not fit in other media categories, such as online papers)

Other:

- Other (materials that did not fit in any other category)

Code: 0201 AdvocacyM, 0202 AdvocacyR, 0203 LegislativeA, 0204 LegislativeB, 0205 LegislativeT, 0206 MediaN, 0207 MediaO, 0208 MediaSL, 0209 ResearchM, 0210 ResearchP, 0211 Other

03XX Direction of document

Definition: Each document is designated as having an overall abstinence, comprehensive, or neutral focus.

Codes: 0301 Abstinence, 0302 Comprehensive, 0303 Neutral

04XX Argument attribution and direction

Definition: Each code is designated with an argument attribution and an argument direction, as follows:

<i>Argument Attribution</i>	<i>Argument Direction</i>
→ Author (A)	→ Abstinence (A)
→ Source (S)	→ Comprehensive (C)
	→ Inconclusive (I)

→ *Source vs. author:* Is author making the argument, or is a source used by the author,
or a source of the source, making the argument?

→ *Comprehensive vs. abstinence vs. inconclusive:* Is the argument in support of an abstinence-only or a comprehensive approach to sex education, or can the direction not be ascertained? This should be determined within the context of the broader argument being made by the author.

Codes: 0401 AA, 0402 AC, 0403 AI, 0404 SA, 0405 SC, 0406 SI