Did the American Psychological Association's *Report on Appropriate Therapeutic Responses to Sexual Orientation* Apply Its Research Standards Consistently?

A Preliminary Examination[1]

Christopher H. Rosik

Link Care Center

Fresno Pacific University

1 Address correspondence to Christopher H. Rosik, PhD, Link Care Center, 1734 W. Shaw Ave., Fresno, CA 93711 or e-mail (christopherrosik@linkcare.org).

Abstract

In August 2009, the American Psychological Association Task Force released its report, *Appropriate Therapeutic Responses to Sexual Orientation* (APA, 2009). The report discouraged attempts at changing sexual orientation, asserting that such efforts are unlikely to succeed and involve some risk of harm. The task force further recommended affirmative therapeutic interventions based in part on the conclusion that research has not found developmental influences to be involved in the origin of sexual orientation. In this critical analysis, I identify several methodological limitations cited by the task force in critique of sexual orientation change efforts (SOCE) and apply them to a review of the majority of studies to which the report makes reference regarding devel- opmental theories of sexual orientation. Based on this examination, it appears most of the studies the task force cited in support of its conclusion had similar methodological flaws that led to its dismissal of SOCE research. Thus, it appears the task force applied its methodological critique inconsistently, raising questions about what might give rise to such variation in reviewing standards.

Did the American Psychological Association's *Report on Appropriate Therapeutic Responses to Sexual Orientation* Apply Its Research Standards Consistently?

A Preliminary Examination

The American Psychological Association's recent task force report (APA, 2009), *Appropriate Therapeutic Responses to Sexual Orientation*, concluded that "efforts to change sexual orientation are unlikely to be successful and involve some risk of harm, contrary to the claims of SOCE practitioners and advocates" (p. v). The task force further recommended affirmative therapeutic interventions based in part on the conclusion that research has not found developmental influences to be involved in the origin of sexual orientation.

The task force report contains a major section dedicated to identifying the methodological problems in research on sexual orientation change efforts (SOCE). This section (pp. 26–34) is meticulous in its efforts to identify any and all limitations within SOCE research in order to discredit this literature. At the same time, the report also highlights literature pertinent to developmental theories of sexual orientation. While no body of research is free from limitations, one measure of the degree of thoroughness and objectivity behind scientific critiques of this nature is the extent to which the criticisms are uniformly applied to research affirmed by the reviewers. The current examination seeks to determine if the APA task force scrutinized the limitations of the research supporting its conclusions to the same degree it did the SOCE research.

Method

In order to obtain at least some preliminary assessment of this issue, one claim made by the task force was assessed: "Studies failed to support theories that regarded family dynamics, gender identity, or trauma as factors in the development of sexual orientation" (APA, 2009, p. 23). In support of this claim, the task force cited ten different references (Bell, Weinberg, & Hammersmith, 1981; Bene, 1965; Freund & Blanchard, 1983; Freund

& Pinkava, 1961; Hooker, 1969; McCord, McCord, & Thurber, 1962; Peters & Cantrell, 1991; Siegelman, 1974, 1981; Townes, Ferguson, & Gillem, 1976).

I was able to obtain the source materials for seven of these articles through the EBSCO database and local libraries. Two of these articles (Freund & Pinkava, 1961; Bene, 1965) appeared in relatively obscure or defunct journals, while the other (Siegelman, 1974) was not locally obtainable. Another of these articles, the Hooker (1969) reference, was in fact a review piece and thus not suitable for the present analysis of research methodology. Moreover, a review article does not fit the task force's billing as being a study that "failed to support" the theories in question, since a review article is an interpretation, not an empirical study. The remaining six research studies cited by the task force thus comprise the focus of my analysis.

Table 1 presents the major methodological limitations ascribed by the task force to the SOCE literature along with the frequency of those limitations in the six studies cited by the task force in support of its etiological conclusion. In order for the task force to conclude so unequivocally that the studies cited failed to support developmental theories of sexual orientation, the research it noted should be free from most, if not all, of these limitations. As a check on my objectivity, another psychologist blind to the purposes of this project randomly reviewed three of the six research articles using the same list of limitations. The tabulations indicated agreement in 72% of the ratings, an acceptable degree of reliability.

Results

Before examining the findings, it should be noted that all of these studies are cross-sectional in nature. While one study (McCord, McCord, & Thurber, 1962) did utilize some longitudinal data, it was not analyzed in a manner that took advantage of the cross-sectional character. This lack of prospective data would appear to be an important limitation in considering the task force's utilization of such research to support its position.

Specifically, it would seem to negate the validity of using such studies for making claims for or against any developmental factor in the etiology of same-sex attractions, yet this is not mentioned by the task force. Overall, as Table 1 indicates, all six studies had important methodological weaknesses germane to cross-sectional research designs that the task force used to disqualify SOCE. Below I provide subsections for each limitation to highlight how consistently the SOCE methodological problems were applied to the etiological literature.

Lacks a clear definition of terms. Most of the studies I reviewed generally attempted to provide some clarification in the definition and operationalization of their variables, at least as far as parental relationship or childhood abuse constructs are concerned. However, there were confusing descriptions. For example, terms such as *affectional interaction* or *dependency* in the McCord et al. (1962) study seemed to lack clarity. The authors identified the latter condition as present if the boy "showed an unusually strong desire for adult approval" (p. 363). Boys evidencing high dependency were classified as showing feminine identification, although it was not immediately clear what constituted "strong desire," a qualifier that appears to lend itself to significant subjectivity in interpretation, which is one reason used by the task force to disqualify some SOCE studies.

Reliance on self-report measures. Of the six studies reviewed, all six involved self-report instruments; for five of the studies, self-report measures were the only ones utilized. The McCord et al. (1962) study was partially based on direct observations, but these chart records were reviewed more than a decade later by the researchers, and at that time the observations were categorized into variables of interest to the study. Thus, the task force conclusions regarding the etiological significance of developmental factors for sexual orientation are based almost entirely on respondents' retrospective memory as opposed to observational assessment, placing a serious limitation on definitive conclusions in this regard.

Reliance on measures of unknown validity/reliability. The six studies reviewed were highly inconsistent in reporting the psychometric properties of the instruments they em-

ployed. Most appeared to employ some instruments that presumably had been developed with an eye toward validity and reliability issues; however, statistics such as alphas were rarely reported, so one is left with little or no evidence that psychometrics were considered.

Siegleman (1981; reported in Siegleman, 1978) was the most forthcoming with information about reliability. Freund and Blanchard (1983) reported alphas for two of their scales (p. 14), but it appeared that these alphas were for prior research using the scales with different samples and not for the current study and sample, constituting a major problem (Thompson & Vache-Haase, 2000). Peters and Cantrell (1991) modeled their questionnaire after a preexisting measure but provided no psychometric information for either, even with items described as attitudinal. This is not in line with common practice of ascertaining the reliability and validity anytime a scale is changed or adapted substantially (Thompson & Vache-Haase, 2000). This was also the case for the Townes et al. (1976) questionnaires and the scale variables derived from them. McCord et al. (1962) alluded in a footnote where reliability information can be located (p. 362), but that doesn't address the validity concerns that seem to arise with how some of their variables were operationalized. Bell et al. (1981) reported their composite measures to be reliable but did not provide the specific alphas. These omissions and uncertainties appear to constitute serious psychometric inadequacies when considered in light of the task force standards.

Study participants not blind to study purposes. The frequency of this particular methodological shortcoming is difficult to ascertain from the information provided in the studies. In keeping with the task force sentiment that studies are generally at risk of this problem if they do not explicitly endeavor to address it, I estimated that four of the studies had potentially introduced bias of this nature. In all but the McCord et al. (1962) study, some or all of the participants were recruited by the researchers or their assistants and very little is stated regarding the wording used to encourage involvement in each study. We can infer that participants were often known by the researchers as patients (Freund & Blanchard, 1983), students (Peters & Cantrell, 1991), or social acquaintances

(Bell et al., 1981; Townes et al., 1976). Bell et al. (1981) noted that some field staff who conducted the interviews "attended private parties held for our benefit by individuals supporting the aims of the study" (p. 11) as part of their recruitment strategy.

If the task force applied the same scrutiny to these studies as they did to the SOCE studies, then these contacts could quite conceivably have introduced a bias that left participants far from "blind" to the general aims of the research. Only the Siegleman (1981) research included a measure of social desirability that could provide some check for response bias. The task force observed, "Knowing that one is being studied and what the experimenter hopes to find can heighten people's tendency to self-report in socially desirable ways and in ways that please the experimenter" (APA, 2009, p. 32). This is a most germane concern for four of the six studies examined here and is also applicable to the next problem explored.

Recruiter/selection bias and/or demand characteristics. The present examination suggests that such bias and/or demand characteristics were likely to be present in each of the six studies. McCord et al. (1962) employed data collected by social workers who "would appear unannounced, with a frequency which made it possible to observe families at meals, during their leisure, in the midst of crisis, and during their ordinary daily routines" (p. 362). It is hard to imagine that the sudden presence of an observer in the room would not impact the behavior of parents and children who knew they were in an experimental program aimed at the prevention of delinquency. In another study (Townes et al., 1976), participants were recruited "following 6 months of observation and involvement by the second author in homosexual institutions" (p. 263). These institutions included "homosexual bars" and a "homosexual counseling center." This advanced familiarity with the recruiter could potentially influence responding in an unknown manner.

Of particular interest is the decision of Bell et al. (1981) to remove from their sample all participants who were reportedly influenced by psychoanalytic theory regarding the etiology of homosexuality when these individuals differed from the heterosexual

subgroup in a manner dissimilar to the difference found for homosexual participants not exposed to such theories. The clear assumption is that such exposure would bias these participants into responding in a way consistent with the psychoanalytic perspective. However, it is just as plausible that participants whose background was consistent with aspects of psychoanalytic theory sought such information due to their sense of its applicability to their developmental narrative. Whatever one wants to believe about this matter, the decision of Bell et al. most certainly reduced the likelihood that their findings would support the psychoanalytic school of thought. In spite of this, the authors did report support for the nonuniversal application of some aspects of psychoanalytic theory in comprehending pathways to homosexuality, such as a modest role of identification with the same-sex parent and poor relations with father in the development of sexual orientation (pp. 189–191).

It appears probable that the researchers in these studies had little sympathy for the psychoanalytic view of homosexuality; thus, to quote the task force, "It cannot be assumed that the recruiters sought to encourage the participation of those individuals whose experiences ran counter to their own view of these approaches" (APA, 2009, p. 34). This could plausibly introduce "unknown selection biases into the recruitment process" (p. 34).

Small sample size. While there is no strict definition for what constitutes a small sample, most of the studies cited by the task force report total samples with well under 200 subjects, with comparison groups sizes for five of the studies varying from five (McCord et al., 1962) to 147 (Freud & Blanchard, 1983). The vast majority of comparison groups in these studies had sizes in the 30 to 100 range. Small samples limit the reliability and generalizability of subsequent findings. The Bell et al., (1981) study was a welcome exception to this problem with a total sample of nearly 1,500 and comparison groups of nearly 1,000 (homosexual) and 500 (heterosexual). By contrast, only four of the 54 SOCE studies examined by the task force (APA, 2009, pp. 126–130) reported a sample size of 200 or more, with most samples below 50.

An empirical analysis of these sample sizes also brings into question consistency of the task force. The mean sample size of the 54 SOCE-related studies included in Appendix B of the report (M = 46.9, SD = 128.4) was not significantly different than the mean sample size of the seven studies (M = 227.7, SD = 338.7) cited as discounting developmental influences on sexual orientation when equal variances could not be assumed, $t(59) = -1.40$, $p = .21$. When simple case studies and one extreme outlier from both groups (Nicolosi, Byrd, & Potts, 200 [N = 882] and Bell, et. al, 1981 [N = 1456]) were removed from the analysis to allow a more accurate comparison, the SOCE studies (M = 65.0, SD = 70.1) continued to have sample sizes statistically similar to the those found in the developmental studies cited by the task force (M = 102.5, SD = 77.4, $t[29] = -1.25$, $p = .22$).

Thus, despite roughly comparable sample sizes in the respective literatures, the task force chose to level this critique at the SOCE literature but not at its own cited etiological research. The limited sample sizes of these studies clearly make population generalizations an endeavor fraught with uncertainty for both of these literatures, to say the least. Quoting the task force again, "Small samples, sample heterogeneity, weak measures, and violations to the assumptions of statistical tests (e.g., non-normally distributed data) are central threats to drawing valid conclusions" (APA, 2009, p. 32).

In addition, significant sample attrition occurred in the McCord et al. study, with the full sample decreasing from 325 to 255 over the five years of observation. This degree of attrition (22%) was less than the task force reported for many of the early SOCE studies, but in line with the attrition rate (26%) reported in the more recent Jones and Yarhouse (2007) longitudinal research, which the task force summarily dismissed. As the task force noted, "Put simply, dropout may undermine the comparability of groups in ways that can bias study outcomes" (APA, 2009, p. 29). Why such cautions by the task force apply only to the SOCE literature is not readily apparent.

Violations of statistical assumptions. None of the six studies examined provided any statistical information that would allow the reader to assess whether or not applicable

univariate and/or multivariate data assumptions (such as linearity and normality) were met. In fact, none of the researchers mentioned that these assumptions had been checked and confirmed. While it is possible the data assumptions were investigated, absence of comment to this effect creates a degree of doubt as to how confident one can be about the results.

Narrow sample compositions. This was a problem with all six of the studies in question. The studies included samples that were recruited from "lower-class" boys (McCord et al., 1962); highly educated, younger residents of the San Francisco Bay Area (Bell et al., 1981); university students (Peters & Cantrell, 1991); and psychiatric patients (Freud & Blanchard, 1983). Homophile organizations were sampled in four of the studies. For example, recruitment of participants in the Siegelman (1981) study involved "The Albany Trust," a group "made up of members who support tolerance and freedom of psychosexual expression" (p. 3), and the "Cosmo Group," a college organization that "attempts to reduce censorship on television and radio, and fosters informed and tolerant opinion" (p. 3). For the majority and strongest of these studies, if the findings can be said to be representative of any group, they seem applicable to white persons who are younger, liberal, well educated, and reside in urban settings. Thus, the same criticism of narrowness in sample composition applies to these studies.

Convenience sample. This problem was also present in all six studies. None of the studies utilized a population-based sample, which is another serious obstacle to generalizing these research findings. As Bell et al. (1981) acknowledged, "In our case, we do not claim to have a representative sample of American homosexuals or heterosexuals, or even of those residing in the San Francisco Bay Area" (p. 19). The task force criticized the lack of "population-based probability sampling strategies" (p. 34) found in the SOCE literature, but this was not an obstacle for them when it came to referencing research that purportedly dispelled developmental theories of sexual orientation.

Failure to differentiate between sexual behaviors, attractions, and orientation identity. None of the six studies made all three of these distinctions in its operationalization of

homosexuality. Townes et al. (1976) did not indicate how they defined the homosexual group. McCord et al. (1962) considered boys to have strong homosexual tendencies "if they played with dolls, sometimes wore dresses, frequently expressed the wish to be a girl, or were overtly homosexual" (p. 363), a term that was not defined. Peters & Cantrell (1991) utilized a single item regarding self-reported same-sex versus opposite-sex preference. Of course, the APA task force noted that this distinction has arisen in the past twenty years, after the publication date of all but one of the studies investigated here. This did not prevent the task force from applying the standard to all the SOCE literature dating back to the 1960s.

Failure to differentiate sexual orientation from sexual orientation identity. According to the task force, sexual orientation refers to a person's pattern of sexual, romantic, and affectional arousal and desire, whereas sexual orientation identity refers to one's acknowledgement and internalization of sexual orientation as an identity. Again, none of the studies examined here made this distinction in its methodologies. The APA task force (2009) warning is thus applicable: "Recent research has found that distinguishing the constructs of sexual orientation and sexual orientation identity adds clarity to an understanding of the variability inherent in reports of these two variables" (p. 30). If one were applying the criteria evenhandedly, might this clarity also apply to the study of etiological factors?

Failure to assess for bisexuality. Not surprisingly, none of the six studies assessed participants for bisexuality. Bell et al. (1981) specifically dichotomized their use of the six-point Kinsey scale, where respondents with an average score from two through six "were classified as homosexual" (p. 32).

Discussion

I came to this examination as a psychologist and researcher prepared to agree with the task force and grant that we know very little conclusively about the efficacy of SOCE. I also approached this review not wedded to a one-size-fits-all etiological explanation of same-

sex attractions; rather, I wanted to see how consistently the task force treated the literature it cited to dismiss potential developmental factors such as family dynamics, trauma, and gender identity. Based on the analysis, which surveyed two-thirds of the research studies cited by the task force, including all of the more recent studies, it appears that these studies were not assessed by the task force with the same level of scrutiny or the same standards it applied to the SOCE literature. In fact, the task force cited studies in support of its positions that actually had the same problems as the SOCE studies it criticized.

Regarding SOCE, the task force concluded, "Due to these limitations, the recent empirical literature provided little basis for concluding whether SOCE has any effect on sexual orientation" (APA, 2009, p. 34). Given that many of these same limitations exist in the etiological literature cited by the task force, questions have to be raised as to why it chose to definitively dismiss this literature as "failing to support" developmental theories. It appears, based on the same criteria the task force used to dismiss SOCE, that its own conclusions have little basis in the literature.

A fairer rendering of the etiological literature the task force references would appear to be that this research is so methodologically flawed that we cannot make any conclusive statements concerning the applicability of developmental factors in the origin of homosexuality. Thus, by the task force's own methodological standards, the literature it cites fails to support *or rule out* a role for these potential developmental influences in the genesis of sexual orientation. If such ambiguity exists in the SOCE literature on methodological grounds, then by the task force's own criteria, this ambiguity also is present in the referenced etiological research. It appears that the task force has been inconsistent in the application of its methodological critique to the broader literature on homosexuality, and it may have been willing to offer more definitive conclusions about theories it wishes to dismiss than is warranted by its own standards. In a word, there is the appearance of substantial bias.

The extent to which such a tendency may permeate the APA report is not ascertainable from this examination, but the findings are enough to raise legitimate questions about

the task force's attention to detail and/or its impartiality. Thus, while the report may in some respects be a step forward in the conversation currently occurring over SOCE, it should not be considered as definitive as many who oppose such psychological care may proclaim it to be. Hopefully, the task force's efforts will be a stimulus to much more and sophisticated research on SOCE that includes the active recruitment and participation of diverse perspectives. Such inclusiveness represents the true spirit of our discipline, is essential to understanding human sexual behavior, and may well be the best means to ensure that scientific knowledge is furthered rather than stymied as it pertains to SOCE.

References

American Psychological Association, Task Force on Appropriate Therapeutic Responses to Sexual Orientation. (2009). *Report of the APA Task Force on appropriate therapeutic responses to sexual orientation.* Washington, D.C.: American Psychological Association.

Bell, A. P., Weinberg, M., & Hammersmith, S. K. (1981). *Homosexuality: A study of diversity among men and women*. New York: Simon & Schuster.

Bene, E. (1965). On the genesis of male homosexuality: An attempt at classifying the role of the parents. *Britain Journal of Psychiatry, 3,* 803.

Freund, K., & Blanchard, R. (1983). Is the distant relationship of fathers and homosexual sons related to the sons' erotic preference for male partners, or to the sons' atypical gender identity, or both? *Journal of Homosexuality, 9,* 7–25.

Freund, K., & Pinkava, V. (1961). Homosexuality in man and its association with parental relationships. *Review of Czechoslovak Medicine, 7,* 32–40.

Hooker, E. A. (1969). Parental relations and male homosexuality in patient and nonpatient populations. *Journal of Consulting and Clinical Psychology, 33,* 140–142.

Jones, S. L., & Yarhouse, M. A. (2007). *Ex-gays?* Downers Grove, IL: InterVarsity Press.

McCord, J., McCord, W., & Thurber, E. (1962). Some effects of paternal absence on male children. *Journal of Abnormal and Social Psychology, 64,* 361–369.

Nicolosi, J., Byrd, A. D., & Potts, R. W. (2000). Beliefs and practices of therapists who practice sexual reorientation psychotherapy. *Psychological* report*s, 86,* 689–702.

Peters, D. K., & Cantrell, P. J. (1991). Factors distinguishing samples of lesbian and heterosexual women. *Journal of Homosexuality, 2,* 1–15.

Siegelman, M. (1974). Parental backgrounds of male homosexual and heterosexuals. *Archives of Sexual Behavior, 3,* 3–18.

Siegelman, M. (1978). Psychological adjustment of homosexual and heterosexual men: A cross-national replication. *Archives of Sexual Behavior, 7,* 1–11.

Siegelman, M. (1981). Parental backgrounds of homosexual and heterosexual men: A cross national replication. *Archives of Sexual Behavior, 10,* 505–513.

Thompson, B., & Vache-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational & Psychological Measurement, 60* (2), 174.

Townes, B. D., Ferguson, W. D., & Gillam, S. (1976). Differences in psychological sex, adjustment, and familial influences among homosexual and nonhomosexual populations. *Journal of Homosexualitly, 1,* 261–272.

Table 1

Frequency of the APA Task Force's SOCE Research Methodology Problems among Studies Cited as Disproving Traditional Developmental Theories of Sexual Orientation

Methodological Problem	Number of Studies Containing the Problem[a]
1. Lack clear definition of terms	1 of 6
2. Relies on self-report measures	5 of 6
3. Relies on measures with unknown validity/reliability	5 of 6
4. Participants not blind to study purposes	4 of 6
5. Small sample sizes	5 of 6
6. Violations of statistical assumptions	? of 6[b]
7. Narrow sample compositions	6 of 6
8. Convenience (vs. population-based) sample	6 of 6
9. Potential recruiter/selection bias and demand characteristics	6 of 6
10. Fails to differentiate sexual behavior, attraction, and orientation	6 of 6
11. Fails to differentiate sexual orientation from sexual identity	6 of 6
12. Fails to assess for bisexuality	6 of 6

NOTE. [a]Studies assessed are Bell, Weinberg, & Hammersmith, 1981; Freud & Blanchard, 1983; McCord, McCord, & Thurber, 1962; Peters & Cantrell, 1991; Siegelman, 1981; and Townes, Ferguson, & Gillam, 1976. [b]None of these studies presented data that would enable the reader to evaluate whether these assumptions were met, so the prevalence of this problem cannot be ascertained.