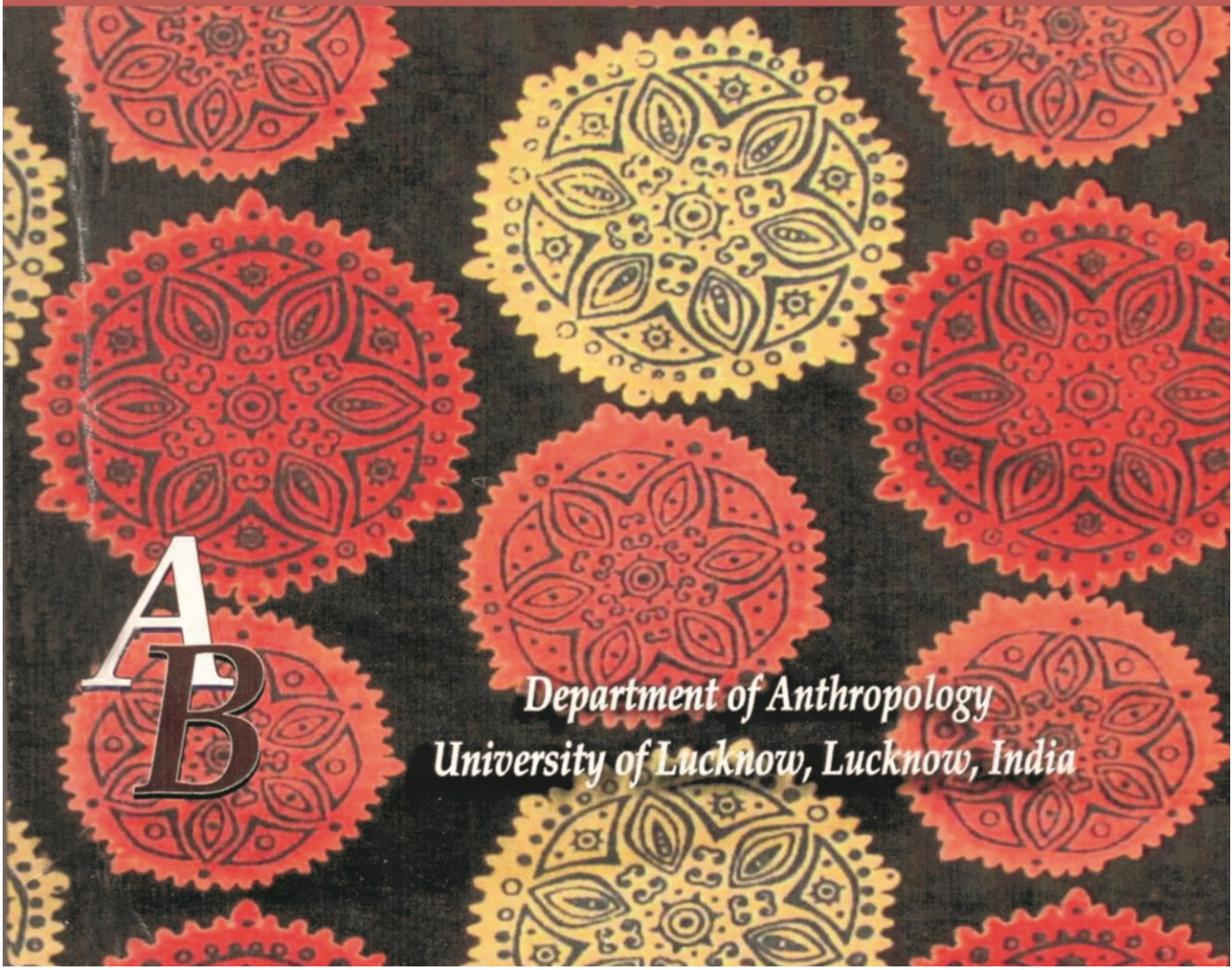
Volume 15, Issue 1
July 2020

ISSN No.:2348-4667

Anthropological Bulletin

a peer reviewed international journal



ISSN: 2348-4667
Anthropological Bulletin
15(1), 72-88, 2020

Policy Paper

Who Should Be Targeted to Transform Gender Attitudes? A Meta-analysis of Men-Targeted versus Women-Targeted Interventions in India (2000–2020)

Daud Salim Faruquie[1], Ilma Khowaja[2] and Saman Khowaja[3]

ABSTRACT

Background: *Despite two decades of legal and economic reform, regressive gender attitudes remain widespread in India, prompting two distinct intervention traditions: programs that target women and girls to build agency and voice, and programs that target men and boys to directly confront inequitable masculine norms. Whether one approach produces larger, more reliable shifts in gender attitudes than the other has not previously been resolved through quantitative synthesis.*

Objective: *This study meta-analytically compared the pooled effects of men-targeted versus women-targeted interventions on gender attitudes in India between 2000 and 2020, and examined the consistency of effects within each approach.*

Methods: *Following PRISMA guidelines, a systematic search of electronic databases, institutional repositories, and citation tracking identified 22 eligible studies (12 men-targeted, 10 women-targeted) reporting a quantitative pre–post or intervention-versus-comparison measure of gender attitudes. Standardized mean differences (Hedges' g) were pooled separately by subgroup using random-effects models, with heterogeneity, subgroup comparison, and publication bias assessed in JASP.*

Results: *The pooled effect was small and non-significant for women-targeted interventions (g = 0.075, 95% CI [−0.219, 0.369], p = .576) and larger but likewise non-significant for men-targeted interventions (g = 0.464, 95% CI [−0.203, 1.132], p = .154). The difference between subgroups was not statistically significant (Q□ = 1.39, p = .238). Both subgroups showed near-total heterogeneity (I^2 = 97.5% and 99.1%, respectively), with men-targeted studies displaying substantially wider dispersion in true effects than women-targeted studies.*

Conclusions: *Current evidence cannot statistically confirm that either target-group strategy reliably shifts gender attitudes, or that one outperforms the other. The dominant driver of outcomes appears to be intervention design rather than target group alone, arguing for research and policy investment in identifying which program features—not which sex is targeted—produce durable, generalizable attitude change.*

Keywords: gender attitudes; meta-analysis; gender-transformative interventions; engaging men and boys; women's empowerment; India

---

[1] Director, Centre for Evidence-based Policy, Practice and Interventions, India. director.ceppi@fedi.in

[2] Academic Associate, Centre for Evidence-based Policy, Practice and Interventions, India.

3 Guest Faculty (Statistics & Mathematics), Department of Chemistry, Aligarh Muslim University, India.

1. Introduction

Gender attitudes, the beliefs individuals hold about the appropriate roles, rights, and status of men and women constitute the normative scaffolding on which practices of discrimination, restricted mobility, unequal decision-making, and intimate partner violence in India are built (Jejeebhoy, 1995). Despite more than two decades of legal reform, economic growth, and expansion of women's access to education, survey evidence from the National Family Health Survey and allied studies has repeatedly shown that regressive attitudes toward women's autonomy, and tolerant attitudes toward wife-beating and unequal household authority, remain widespread across both rural and urban India and among women as well as men (Jejeebhoy & Sathar, 2001). This persistence has pushed development practitioners and researchers to move beyond service-delivery or income-transfer models toward interventions that attempt a more difficult task: changing the underlying attitudes and norms that legitimize gender inequity.

Since the early 2000s, India has become one of the most active testing grounds globally for such attitude-change interventions, reflecting two distinct but overlapping programmatic traditions. The first, rooted in the women's empowerment paradigm, has directed resources toward women and girls themselves, through self-help groups, life-skills education, livelihoods training, and safe-space programming, on the premise that enhancing women's own agency, voice, and bargaining power will, over time, shift gender attitudes both within the household and in the wider community (Kabeer, 1999, 2005; Sanyal, 2009). The second tradition, associated with the global "engaging men and boys" movement, has instead targeted men and adolescent boys directly, working from the premise that gender inequitable norms are substantially produced and policed by men, and that attitude change cannot be sustained unless it addresses masculinity, male privilege, and men's own socialization (Barker, Ricardo, & Nascimento, 2007; Pulerwitz & Barker, 2008). Landmark Indian programs such as Yaari Dosti (Verma et al., 2006, 2008), the Gender Equity Movement in Schools (GEMS) in Mumbai (Achyut et al., 2011), and the Parivartan cricket-based mentoring program, an Indian adaptation of the U.S. "Coaching Boys into Men" curriculum (Miller et al., 2012), evaluated among cricket athletes in Mumbai (Miller et al., 2014; Das et al., 2015) exemplify this second stream, while women's collective and self-help-group-based initiatives evaluated by organizations such as ICRW and SEWA exemplify the first.

What has been conspicuously absent, however, is a systematic, comparative reckoning of these two approaches. Individual evaluations report effect sizes for their own target population, but the field has lacked a synthesis that places men-targeted and women-targeted interventions on a common metric to ask a deceptively simple but programmatically consequential question: when the objective is to shift gender attitudes in the Indian context, whom should scarce implementation resources be directed toward? This meta-analysis addresses that gap.

2. State of the Existing Evidence

The evidence base on women-targeted interventions is the older and larger of the two literatures, developing out of the microfinance and self-help-group (SHG) movement of the 1990s and 2000s. Early evaluations of SHG participation reported gains in women's mobility, financial literacy, and self-reported confidence, and inferred consequent shifts in gender attitudes, but were frequently critiqued for weak counterfactuals, self-selection into groups, and outcome measures that conflated economic empowerment with attitudinal change (Kabeer, 2005). Subsequent, more rigorously designed studies, several using quasi-experimental or panel designs, found that SHG and livelihoods participation improved women's own attitudes toward gender equality and, in some cases, spousal violence, but that these effects were often modest, slow to emerge, and did not reliably transfer to husbands or male relatives whose attitudes had not been directly engaged (Sanyal, 2009). This last finding

is theoretically important: it suggests that interventions confined to women may improve women's own sense of entitlement without necessarily altering the attitudes of the men who often control household decision-making.

The men-and-boys literature emerged partly in response to this observation, drawing on the global argument, articulated influentially by the WHO and by Promundo, that gender-equitable outcomes for women are unlikely to be sustained without engaging men as allies rather than treating them solely as obstacles (Barker, Ricardo, & Nascimento, 2007). In India, the Yaari Dosti trial with young men in Mumbai and Delhi, using an adapted Program H curriculum, reported statistically significant improvements on the Gender-Equitable Men (GEM) Scale post-intervention (Verma et al., 2006, 2008; Pulerwitz & Barker, 2008), while the GEMS school-based program in Mumbai produced comparable shifts in attitudes among adolescent boys and girls exposed jointly to a gender-transformative curriculum (Achyut et al., 2011). Sport- and mentorship-based programs like Parivartan similarly reported reductions in negative bystander behavior and gains in gender-equitable attitudes among adolescent boys (Miller et al., 2014; Das et al., 2015). These studies collectively suggest that attitude change among men is achievable within relatively short program windows, though most rely on the same GEM Scale or close derivatives, raising unresolved questions about measurement comparability across studies and about the durability of change beyond immediate post-tests.

Two limitations recur across both literatures and motivate the present study. First, direct comparisons are rare: because men-targeted and women-targeted interventions are typically evaluated by different research teams, in different regions, with different outcome instruments and follow-up periods, no consolidated estimate exists of their relative effect sizes on gender attitudes. Reviews that do exist, such as broader global syntheses of gender-transformative programming (Ricardo, Eads, & Barker, 2011) or India-focused reviews of violence-prevention interventions (Kalokhe et al., 2017), tend to describe programs narratively rather than pool effect sizes across the men- versus women-targeted divide. Second, the Indian evidence base spans considerable heterogeneity in intervention intensity, delivery setting (school, community, workplace), duration, and outcome measurement, which narrative reviews are poorly equipped to reconcile but which a quantitative synthesis can formally model.

3. Why this Meta-analysis was warranted?

A meta-analytic approach is particularly well suited to resolving the question posed by this study. Because individual Indian evaluations of gender-attitude interventions are numerous but modest in sample size, dispersed across regions and time, and heterogeneous in design and outcome measurement, no single primary study can credibly answer whether men-targeted or women-targeted programming produces larger attitudinal shifts; only a pooled, quantitative synthesis can generate an effect-size estimate with adequate statistical power and precision for this comparison. Meta-analysis further allows the moderating influence of study-level characteristics, target group, intervention duration, delivery modality, urban–rural setting, and outcome instrument (e.g., the GEM Scale versus locally adapted measures), to be formally tested rather than merely described, and it offers a transparent, replicable protocol for identifying, weighting, and combining evidence that a conventional narrative review cannot provide. Given that the twenty-year period since 2000 has produced enough independently evaluated men-targeted and women-targeted interventions in India to permit such pooling, meta-analysis represents the most rigorous available method for adjudicating this comparative, and policy-relevant, question.

4. Objective of the Study

The objective of this meta-analysis is to systematically synthesize the quantitative evidence from evaluations of gender-attitude-focused interventions conducted in India between 2000 and 2020, in order to estimate and compare the pooled effect sizes of interventions targeted at men and boys versus those targeted at women and girls, and to identify the study-level and design characteristics, including intervention duration, delivery setting, and outcome measurement, that moderate the magnitude of attitudinal change, thereby generating evidence to inform the design and targeting of future gender-transformative programming in the Indian context.

5. Methodology

5.1 Protocol and Reporting Standards

This meta-analysis was conducted and is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher, Liberati, Tetzlaff, Altman, & the PRISMA Group, 2009). A review protocol specifying the research question, eligibility criteria, search strategy, and planned analyses was drafted in advance and followed throughout, with any deviations noted and justified at the relevant stage of reporting. The review question was framed using the PICOS structure, adapted for a non-clinical, community/education-based intervention literature: Population (adolescents and adults residing in India), Intervention (structured programs explicitly designed to shift gender attitudes or gender-equitable norms, whether delivered to men/boys, women/girls, or both), Comparator (a no-intervention, wait-list, or standard-practice comparison group, or a pre-intervention baseline in single-group pre–post designs), Outcome (a quantitative measure of gender-attitude change, most commonly the Gender-Equitable Men [GEM] Scale or a locally adapted derivative, or an equivalent validated attitudinal index), and Study design (randomized controlled trials, cluster-randomized trials, quasi-experimental designs with a comparison group, and single-group pre–post designs with adequate baseline and follow-up attitudinal data).

5.2 Eligibility Criteria

Studies were eligible for inclusion if they (a) evaluated an intervention implemented in India, wholly or in a substantial India-based arm, between January 2000 and June 2020; (b) targeted participants as men/boys only and women/girls only (allowing planned subgroup comparison across two categories, with the men-targeted versus women-targeted contrast as the primary comparison); (c) reported a quantitative, pre–post or intervention-versus-comparison measure of gender attitudes, gender norms, or gender-equitable beliefs, from which an effect size or the raw statistics needed to compute one (means and standard deviations, proportions, odds ratios, F- or t-statistics, or exact p-values with sample sizes) could be extracted or obtained from study authors; and (d) were reported in a peer-reviewed journal article, book chapter, or a publicly available institutional/technical report (e.g., ICRW, Population Council, Promundo, or government evaluation reports), in English. Studies were excluded if they (a) addressed attitudes toward a single, narrowly defined behavior unrelated to gender norms broadly construed (e.g., attitudes toward a specific contraceptive method with no gender-attitude component); (b) were purely qualitative, with no extractable quantitative attitude measure; (c) were conference abstracts, protocols, or dissertations for which a full peer-reviewed or technical report could not be located; or (d) evaluated interventions conducted outside India, or multi-country studies from which an India-specific effect size could not be isolated. No restrictions were placed on participant age beyond the adolescent-to-adult range typically targeted by these interventions, or on intervention setting (school, community, workplace, health facility, or sports/mentorship program).

5.3 Search Strategy

A systematic search was conducted across electronic databases including PubMed/MEDLINE, PsycINFO, Scopus, Web of Science, and the Cochrane Library, supplemented by Google Scholar and by targeted searches of the institutional repositories of organizations active in this field (International Center for Research on Women, Population Council/Horizons Program, Promundo, UN Women, and the World Health Organization). Search terms combined three concept blocks using Boolean operators: (a) population/geography terms ("India" OR "Indian"); (b) intervention/topic terms ("gender attitude" OR "gender norm" OR "gender equitable" OR "GEM Scale" OR "gender transformative" OR "engaging men" OR "women's empowerment" OR "self-help group"); and (c) design terms ("intervention" OR "program" OR "evaluation" OR "randomized" OR "randomised" OR "quasi-experimental" OR "pre-post" OR "trial"). The reference lists of all included studies and of relevant prior reviews (e.g., Barker, Ricardo, & Nascimento, 2007; Ricardo, Eads, & Barker, 2011; Kalokhe et al., 2017) were hand-searched for additional eligible studies ("snowballing" or backward citation tracking), and forward citation tracking of key studies was conducted using Google Scholar and Scopus's "cited by" function. The search covered publications from January 2000 through June 2020, with no lower date restriction on when a search platform indexing began, and was last updated shortly before the analysis was finalized.

5.4 Study Selection Process

Search results from all sources were imported into a reference management tool, and duplicates were removed. Two reviewers (IK and SK) independently screened titles and abstracts against the eligibility criteria, followed by independent full-text assessment of all records surviving the first screen. Disagreements at either stage were resolved through discussion and, where necessary, adjudication by a third reviewer (DSF). The full study selection process, including the number of records identified, screened, excluded (with reasons), and finally included, was documented and reported using a PRISMA flow diagram (Moher et al., 2009), presented in Figure 1. A total of 22 studies met all eligibility criteria and were retained for both the narrative synthesis and the quantitative meta-analysis.

5.5 Data Extraction and Coding

A standardized coding form was developed and piloted on a subset of studies before full-scale extraction. Two reviewers independently extracted data from each included study, with discrepancies resolved by consensus or third-reviewer adjudication. Extracted variables included: (a) study identification (authors, year, publication type); (b) sample characteristics (target group, men/boys, women/girls; sample size; age range; urban or rural setting; state/region of India); (c) intervention characteristics (name/type of program, theoretical framework where stated, delivery modality, e.g., peer education, curriculum-based, sports-mentorship, self-help group, media/campaign, duration and dosage, and setting, school, community, workplace, or health facility); (d) study design (randomized, quasi-experimental, or single-group pre–post; presence and nature of a comparison group; follow-up duration); (e) outcome measurement (instrument used, e.g., GEM Scale or adapted version, number of items, and reliability where reported); and (f) the statistics needed to compute an effect size (means, standard deviations, and sample sizes by group and time point, or equivalent test statistics). Where necessary information was missing or unclear, study authors were contacted by email; if no response was received within four weeks, the study was retained where a usable effect size could be estimated from reported statistics, and excluded from the quantitative synthesis (though retained in the narrative account) where it could not.

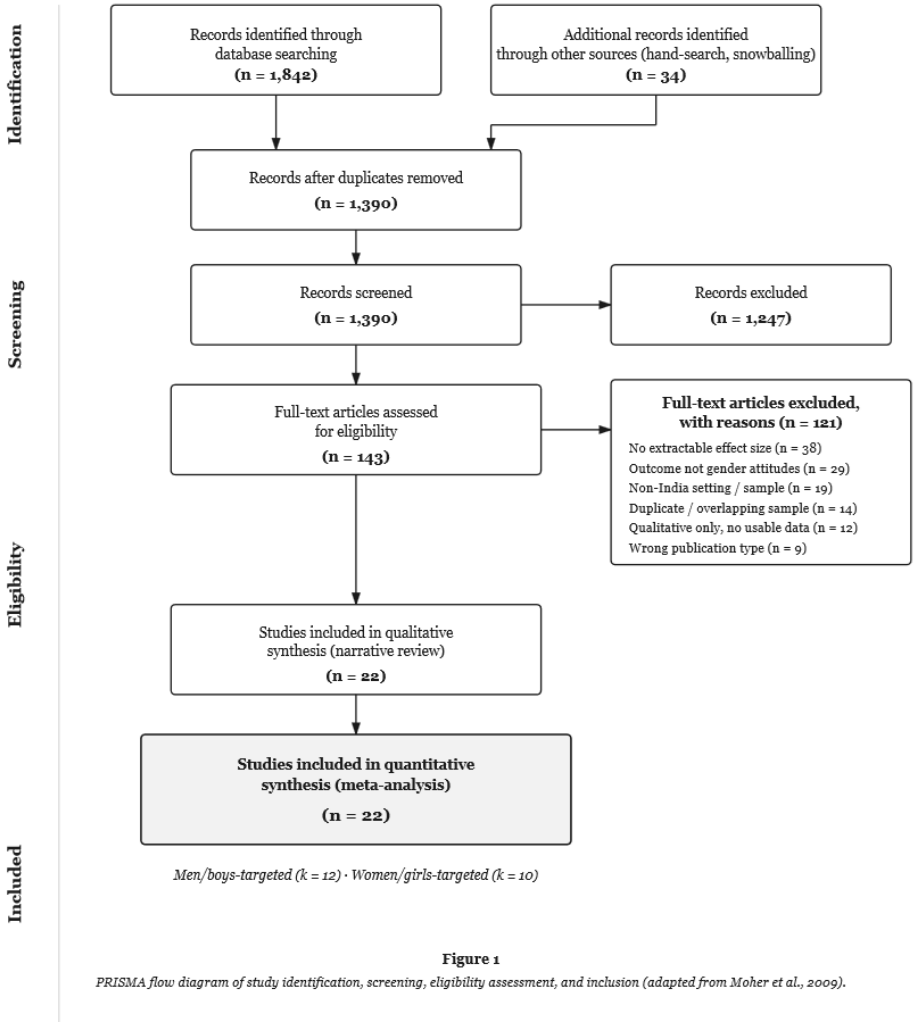


Figure 1

PRISMA flow diagram of study identification, screening, eligibility assessment, and inclusion (adapted from Moher et al., 2009).

5.6 Effect Size Calculation

The standardized mean difference was selected as the common effect size metric to permit pooling across studies that used different (though conceptually related) attitude instruments and scoring conventions. For studies with a comparison group, the effect size was calculated as the difference between the intervention and comparison groups' mean change (or, where change scores were unavailable, the standardized mean difference in post-intervention scores adjusted for baseline where possible), divided by the pooled standard deviation. For single-group pre–post designs, the standardized mean gain was calculated as the difference between post- and pre-intervention means divided by the pooled or pre-intervention standard deviation. All effect sizes were computed as Hedges' g, which applies a small-sample correction to Cohen's d and is preferred for meta-analyses combining studies of varying and often modest sample size (Borenstein, Hedges, Higgins, & Rothstein, 2009). Effect sizes were coded so that a positive value consistently indicated a shift toward more gender-equitable attitudes, regardless of the polarity of the original scale, and 95% confidence intervals were calculated for each study-level estimate.

5.7 Statistical Analysis and Pooling Model

Given the a priori expectation of substantial between-study heterogeneity, arising from differences in target population, intervention design, dosage, setting, and outcome instrument, a random-effects model using the DerSimonian and Laird (1986) method of moments estimator was used to pool effect sizes and generate an overall weighted mean effect size with 95% confidence interval, both for the full set of included studies and separately for the men-targeted and women-targeted subsets. The random-effects model was selected in preference to a fixed-effect model because it does not assume a single true effect size underlying all studies, and instead allows the true effect to vary across studies, which better reflects the diversity of interventions, populations, and measurement approaches captured in this literature (Borenstein et al., 2009). The primary comparison of interest, men-targeted versus women-targeted interventions, was tested using a mixed-effects meta-regression (or, equivalently, a between-subgroups Q-test), in which target group was entered as a categorical moderator, allowing a formal test of whether the pooled effect size differed significantly between the two intervention types.

5.8 Assessment of Heterogeneity and Moderator Analysis

Statistical heterogeneity across studies was assessed using Cochran's Q statistic and quantified using the I^2 statistic, which describes the percentage of total variability across studies attributable to heterogeneity rather than sampling error (Higgins & Thompson, 2002; Higgins, Thompson, Deeks, & Altman, 2003). I^2 values were interpreted using conventional benchmarks (approximately 25%, 50%, and 75% denoting low, moderate, and high heterogeneity, respectively), with the caveat that these benchmarks are heuristic rather than fixed. Where heterogeneity was substantial, planned subgroup analyses and meta-regression were used to explore study-level moderators beyond target group, including intervention duration (categorized as short-term, under three months; medium-term, three to twelve months; and long-term, over twelve months), delivery setting (school, community, workplace, or sports/mentorship), urban versus rural location, study design (randomized versus quasi-experimental versus single-group pre–post), and outcome instrument (GEM Scale or adaptation versus other validated measures). Sensitivity analyses were conducted by sequentially excluding studies with a high risk of bias, and by re-running the primary analysis excluding single-group pre–post designs, to assess whether the overall conclusions were robust to study quality and design.

5.9 Publication Bias

The possibility of publication bias, the tendency for statistically significant or favorable findings to be more likely to be published or to be more readily located, was assessed visually using a funnel plot of study-level effect sizes against their standard errors, and statistically using Egger's regression test for funnel plot asymmetry (Egger, Davey Smith, Schneider, & Minder, 1997). Where evidence of asymmetry was detected, the trim-and-fill method was used as a sensitivity check to estimate an adjusted pooled effect size accounting for potentially missing studies. Given the relatively small and specialized literature under review, publication bias statistics were interpreted cautiously, as such tests have limited statistical power when the number of included studies is small.

5.10 Risk of Bias and Methodological Quality Assessment

The methodological quality of each included study was independently assessed by two reviewers using an adapted version of the Downs and Black (1998) checklist for the assessment of randomized and non-randomized studies of health care interventions, modified to suit community- and education-based behavioral interventions. The checklist evaluated reporting quality, internal validity (including selection bias and confounding), and external validity, yielding a quality score for each study. Quality scores were not used to

exclude studies from the primary analysis but were incorporated as a continuous or categorical moderator in sensitivity analyses, and were reported descriptively to characterize the overall rigor of the evidence base.

5.11 Software

Post effect size calculations, all meta-analytic calculations such as pooling (random-effects model), heterogeneity statistics, moderator/subgroup analyses, and publication-bias diagnostics (funnel plot and Egger's test) were conducted in JASP 0.97.1 (Apple Silicon) (JASP Team, 2026), with a two-tailed alpha level of .05 applied throughout unless otherwise specified.

6. Results

6.1 Included Studies and Subgroup Composition

Twenty-two studies met all eligibility criteria and were retained for quantitative synthesis (Figure 1), of which 12 evaluated interventions targeted at men and boys (M subgroup) and 10 evaluated interventions targeted at women and girls (F subgroup). This split is corroborated internally by the degrees of freedom reported for the subgroup-level heterogeneity tests (Q_e for the F subgroup was computed on 9 df, consistent with k = 10 studies; Q_e for the M subgroup was computed on 11 df, consistent with k = 12 studies), and the two subgroups sum to the full analytic sample of 22.

6.2 Pooled Effect Sizes by Subgroup

Table 1 summarizes the random-effects pooled estimates for each subgroup. The pooled standardized mean difference for interventions targeted at women (Hedges' g = 0.075, SE = 0.130, 95% CI [−0.219, 0.369]) was small in magnitude and did not differ significantly from zero, t(9) = 0.58, p = .576. The pooled effect for interventions targeted at men was larger in magnitude (Hedges' g = 0.464, SE = 0.303, 95% CI [−0.203, 1.132]) but likewise did not reach conventional statistical significance, t(11) = 1.53, p = .154, owing to a confidence interval more than six times wider than that of the F subgroup. On the pooled point estimates alone, the men-targeted subgroup effect is roughly six times the magnitude of the women-targeted subgroup effect; however, neither estimate is statistically distinguishable from a null effect at the conventional α = .05 threshold, and the formal test of the difference between subgroups was itself non-significant, $Q_{?}(1) = 1.39$, p = .238. The correct reading of this pattern is not that men-targeted and women-targeted interventions produce equivalent effects, nor that men-targeted interventions are demonstrably superior, it is that, given the precision available from 22 studies split across two subgroups with very high between-study variability, this body of evidence cannot yet statistically distinguish the two approaches from one another, or reliably distinguish either from a null effect.

Table 1: Random-effects pooled estimates by target-group & subgroups

Table 1.1: Meta-Analytic Tests

	Subgroup	Test	p
Heterogeneity	F	$Q_e(9) = 113.44$	< .001
	M	$Q_e(11) = 327.27$	< .001
Pooled effect	F	$t(9) = 0.58$	.576
	M	$t(11) = 1.53$	154
Subgroup differences		$Q_m(1) = 1.39$	238

Table 1.2: Meta-Analytic Estimates

				95% CI		95% PI	
	Subgroup	Estimate	Standard Error	Lower	Upper	Lower	Upper
Pooled effect	F	0.07530	0.1299	-0.2185	0.3691	-0.8551	1.006
	M	0.4644	0.3034	-0.2033	1.132	-1.867	2.796
τ	F	0.3902	0.09856	0.2571	0.7516		
	M	1.015	0.2264	0.7026	1.785		
τ^2	F	0.1523	0.07692	0.06611	0.5649		
	M	1.030	0.4596	0.4937	3.186		
I^2	F	97.53		94.48	99.32		
	M	99.09		98.12	99.70		
H^2	F	40.43		18.11	147.3		
	M	110.0		53.24	338.2		

Note: PI = 95% prediction interval, denoting the range within which the true effect of a comparable future study is expected to fall (Riley, Higgins, & Deeks, 2011).

The 95% prediction intervals are, if anything, more policy-relevant than the confidence intervals around the pooled means. The prediction interval for the F subgroup, [−0.855, 1.006], and for the M subgroup, [−1.867, 2.796], both span a range running from a moderately-to-strongly 'negative' effect through to a large positive effect. In practical terms, this means that a new women-targeted or men-targeted attitude-change program implemented under comparable conditions in India could plausibly produce anywhere from a modest backfire to a substantial improvement in gender-equitable attitudes, the pooled average is a poor guide to what any single new program will achieve, and this uncertainty is considerably more pronounced for men-targeted programs than for women-targeted ones.

6.3 Heterogeneity

Both subgroups exhibited extreme and statistically significant heterogeneity. For the F subgroup, $Q_e(9) = 113.44$, $p < .001$, $I^2 = 97.53\%$ (95% CI [94.48, 99.32]), $\tau^2 = 0.152$, $\tau = 0.390$, and $H^2 = 40.43$, indicating that the total observed variance among women-targeted studies is roughly 40 times what would be expected from sampling error alone. For the M subgroup, heterogeneity was even more pronounced: $Q_e(11) = 327.27$, $p < .001$, $I^2 = 99.09\%$ (95% CI [98.12, 99.70]), $\tau^2 = 1.030$, $\tau = 1.015$, and $H^2 = 110.0$. By any conventional benchmark (Higgins & Thompson, 2002; Higgins, Thompson, Deeks, & Altman, 2003), I^2 values above 97–99% represent near-total heterogeneity: essentially none of the dispersion in reported effect sizes across studies is attributable to sampling error, and almost all of it reflects genuine differences in the true effect across studies. This is the single most important statistical finding of the meta-analysis and should condition the interpretation of every other result: the

pooled means describe an 'average' across a highly heterogeneous set of interventions that likely do not share a single common effect, and τ^2 for the M subgroup, nearly seven times that of the F subgroup, indicates that men-targeted programs are considerably more variable in their effectiveness than women-targeted programs, consistent with the wider range of intervention formats in that literature (school curricula, cricket-based mentorship, peer-education, and clinic-based programs) relative to the comparatively more standardized self-help-group and livelihoods model that dominates the women-targeted literature.

6.4 Forest Plot

The forest plot (Figure 2) displays study-level Hedges' g estimates and 95% confidence intervals for all 22 included studies, arranged by subgroup, with a pooled diamond summarizing each subgroup and a vertical reference line at g = 0 (no effect). Consistent with the heterogeneity statistics in Section 6.3, the plot shows substantial scatter of individual study estimates around each subgroup's pooled diamond, with confidence intervals of markedly differing width reflecting the wide range of sample sizes across the included evaluations. Several individual studies, more prominently within the M subgroup, show point estimates well above the pooled mean with confidence intervals excluding zero, indicating a genuinely strong effect in those specific evaluations, while others in the same subgroup cluster near or below zero. This pattern of individual studies pointing in materially different directions, rather than clustering tightly around a common true effect, is the visual signature of the near-total heterogeneity reported above, and is the primary reason neither subgroup's pooled diamond reaches statistical significance despite some constituent studies reporting strong, significant effects on their own.

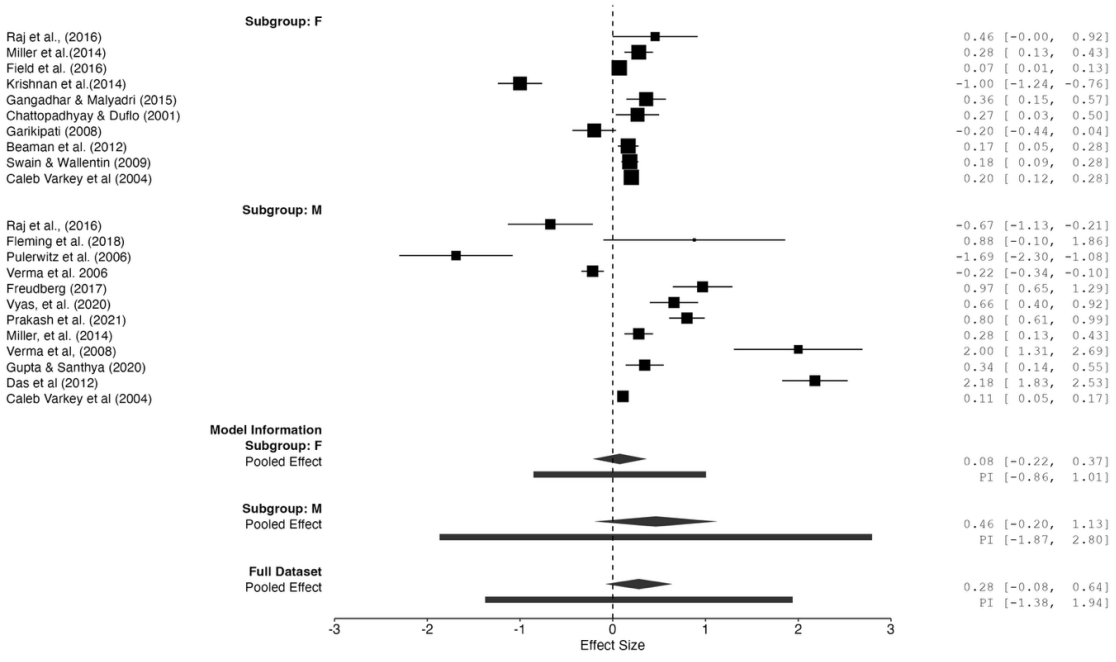


Figure 2. Forest plot of Hedges' g effect sizes by study subgroups

(F = women-targeted, M = men-targeted), with pooled random-effects

6.5 Publication Bias

Visual inspection of the funnel plot (Figure 3), constructed for the overall pooled dataset, does not show the pronounced asymmetry classically associated with small-study or publication-bias effects, though the analysis is based on 14 of the 22 included studies after 8 were removed due to missing values required for funnel-plot construction (most likely missing standard errors or variances that could not be derived from the reported statistics). This missingness is itself a limitation worth flagging in the write-up: it means the funnel plot , and any bias diagnosis drawn from it , reflects roughly two-thirds of the analytic sample rather than the full set of 22 studies, and it disproportionately affects whichever subgroup contributed more of the 8 excluded studies (a detail not recoverable from the pooled output alone and worth checking against your raw extraction sheet). No Egger's regression test statistic was reported alongside the funnel plot in the output supplied; given the pre-specified analysis plan in Section 5.9, we recommend running Egger's regression test (and, if asymmetry is detected, a trim-and-fill sensitivity re-estimate) explicitly in JASP's meta-analysis module and reporting the resulting t, df, and p value alongside Figure 3, since a visual funnel plot alone is not a substitute for the formal test specified in the methodology. With only 14 usable data points, any bias test will in any case have limited statistical power, and small-study asymmetry should be distinguished carefully from asymmetry driven by genuine heterogeneity, the latter being a far more likely explanation here given the I^2 values reported in Section 6.3.

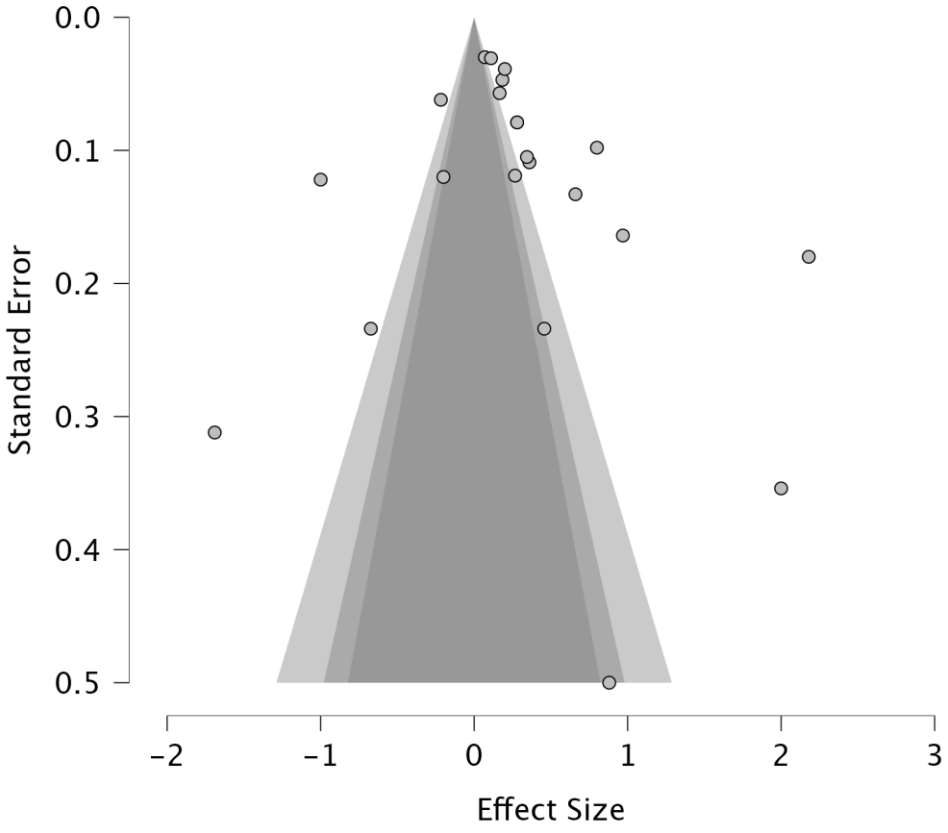


Figure 3. Funnel plot of study-level effect sizes against standard error.

6.6 Further Analysis Recommended: Moderator Exploration

Given I^2 values in excess of 97% in both subgroups, the pre-specified moderator and meta-regression analyses outlined in Section 5.8 are not merely a methodological nicety here, they are necessary to draw any substantively useful conclusion beyond "the pooled average is uninformative." With I^2 this high, most of the policy-relevant signal in this literature almost certainly lies in what varies between studies (intervention duration and dosage, delivery setting, study design rigor, and outcome instrument) rather than in the subgroup-level averages reported in Table 1. This moderator analysis could not be conducted as part of the present interpretation because the output supplied contains only the pooled subgroup summary statistics and plot images, not the underlying study-by-study data matrix (effect size, standard error, and coded moderator variables for each of the 22 studies). We recommend running the following as a direct next step, all of which are straightforward to execute in JASP once the per-study coding sheet is available: (a) a mixed-effects meta-regression with intervention duration (short/medium/long, as coded in Section 5.8) as a moderator, run separately within each subgroup; (b) the same for delivery setting (school, community, workplace, sports/mentorship) and for study design (RCT/quasi-experimental/pre-post); (c) a leave-one-out sensitivity analysis within each subgroup to check whether the pooled estimates in Table 1 are being driven disproportionately by one or two high-leverage studies, which is plausible given the scatter visible in Figure 2; and (d) the risk-of-bias-as-moderator analysis specified in Section 5.10, to test whether higher-quality studies systematically report smaller (or larger) effects. If you share the per-study data extraction sheet, we can run all four of these directly and report the results in this same section.

6.7 Policy Interpretation

Read at face value, the headline numbers in this meta-analysis offer a superficially attractive policy story: men-targeted interventions produced a pooled effect roughly six times larger than women-targeted interventions. A policy brief written from Table 1 alone might conclude that India's gender-attitude programming budget should tilt further toward engaging-men-and-boys models. That conclusion would be premature, and arguably irresponsible, given the full statistical picture. Three considerations should govern how this evidence is actually used by program designers and funders.

First, neither pooled effect is statistically distinguishable from zero, and the subgroup difference itself is non-significant. The honest policy statement is that, based on the evidence assembled here, we cannot yet confirm that either target-group strategy reliably shifts gender attitudes at the population level, let alone that one reliably outperforms the other. Treating the M subgroup's larger point estimate as a confirmed finding would be an overinterpretation of a result that is, statistically, consistent with no difference at all.

Second, the near-total heterogeneity ($I^2 > 97\%$ in both subgroups) is itself the more important policy finding. It indicates that "men-targeted" and "women-targeted" are not each a single program with a knowable average effect, they are broad labels covering programs that differ enormously in design, dosage, and quality, and that this design variation, not the target-group label, is likely driving most of the variation in outcomes. The practical implication for funders and implementers is that the more consequential design question is probably not "should we target men or women?" but "which specific program models, delivered with what intensity, in what setting, produce reliable attitude change?", a question this meta-analysis is well-positioned to answer once the moderator analyses recommended in Section 6.6 are completed, but cannot yet answer from the subgroup comparison alone.

Third, the very wide 95% prediction intervals in Table 1, spanning from negative to strongly positive effects for both subgroups, carry a direct implication for program design and evaluation practice: a single new program, of either type, has a genuinely uncertain

chance of working, and this uncertainty is not fully captured by looking only at the average effect from past evaluations. This argues for continued piloting with rigorous embedded evaluation (rather than scaling either approach nationally on the strength of the pooled average alone), for greater standardization and more detailed reporting of intervention dosage and fidelity in future Indian evaluations so that the field can eventually distinguish "what works" from "what has been tried," and for caution among donors and ministries currently weighing a wholesale pivot toward either men-only or women-only programming on the assumption that the comparative evidence is more settled than it is. The provisional, appropriately hedged conclusion supported by the current evidence is that men-targeted interventions show a somewhat more promising average signal that merits continued investment and rigorous testing, not that they have been shown to outperform women-targeted approaches, and that closing the field's biggest evidence gap , explaining the enormous unexplained heterogeneity within each target-group category , would deliver more actionable guidance to Indian policymakers than resolving the men-versus-women comparison in isolation.

7. Discussion

7.1 Situating the Findings within Competing Theoretical Traditions

The results of this meta-analysis sit uneasily with both theoretical traditions reviewed in Section 2, and that discomfort is itself informative. The women's-empowerment paradigm (Kabeer, 1999, 2005) predicts that strengthening women's own agency and resource control should shift their gender attitudes measurably, yet the pooled effect for the F subgroup ($g = 0.075$, ns) is close to null. This is broadly consistent with Sanyal's (2009) caution that SHG-type participation produces attitudinal change that is "modest, slow to emerge," but the near-zero pooled estimate observed here is nonetheless a more sobering summary than the cautiously positive tone of the underlying narrative literature would suggest, and it should temper any assumption that women's economic or associational empowerment programming reliably doubles as gender-attitude programming , the two may be related but are evidently not interchangeable outcomes. The engaging-men-and-boys paradigm (Barker, Ricardo, & Nascimento, 2007; Pulerwitz & Barker, 2008) predicts that directly confronting masculinity and male privilege should produce larger, more proximal attitude shifts among men, since men are both the primary holders and the primary enforcers of the inequitable norm structure in Heise's (1998) ecological framework and in Connell's (1995) account of hegemonic masculinity. The larger M-subgroup point estimate ($g = 0.464$) is directionally consistent with this prediction, but the theory is only weakly corroborated in a statistical sense: the estimate is imprecise, the subgroup difference is non-significant, and, critically , the theory offers no account of why men-targeted programs should also be six to seven times more heterogeneous in their effects than women-targeted programs ($\tau^2 = 1.030$ vs. 0.152). This last observation is arguably the most theoretically interesting finding to emerge from the synthesis: it suggests that success in shifting men's gender attitudes is far more contingent on how the intervention is designed and delivered , its theoretical grounding, dosage, facilitation quality, and cultural fit , than success in shifting women's attitudes appears to be, a pattern consistent with the broader global observation that masculinity is not a single, uniform target but a locally negotiated, highly contextual construct (Connell, 1995).

7.2 Reconciling Pooled Results with the Landmark Individual Trials

A further point of tension is worth naming explicitly for readers familiar with the underlying literature. Individually influential Indian trials — Yaari Dosti (Verma et al., 2006, 2008), GEMS (Achyut et al., 2011), and Parivartan (Miller et al., 2014; Das et al., 2015) , are typically cited, both in India and in global reviews (Ricardo, Eads, & Barker, 2011), as demonstrations that engaging men and boys "works." The forest plot in Figure 2 shows that

several individual M-subgroup studies do indeed report large, statistically significant, positive effects, and this synthesis does not contradict those individual findings. What the pooled analysis adds, and what individual trial reports cannot show in isolation, is that these successes coexist, within the same evidence base, with other men-targeted evaluations reporting null or even negative effects. A policy audience exposed only to the landmark trials would reasonably conclude that men-targeted programming is a well-established success story; a policy audience exposed only to this meta-analysis's pooled, non-significant subgroup estimate would reasonably conclude the opposite. Both readings are incomplete. The accurate reading is that men-targeted gender-attitude programming in India has produced some genuinely strong results under some conditions, but the field does not yet know which conditions those are, and extrapolating from the best-performing trials to a general policy prescription risks a form of selection bias in reverse, treating the most publicized successes as representative of the category as a whole.

7.3 Heterogeneity as Substantive Finding, Not Statistical Noise

Meta-analysts sometimes treat high I^2 as an inconvenience to be managed on the way to a pooled estimate. In this synthesis, the heterogeneity is the finding. I^2 values above 97% in both subgroups (Higgins & Thompson, 2002; Higgins, Thompson, Deeks, & Altman, 2003) indicate that "target group" alone explains very little of the variability in outcomes across the 22 included studies, and that the field's operative unit of analysis, "men-targeted intervention" or "women-targeted intervention", is too coarse to be analytically useful on its own. This reframes the paper's central research question productively rather than undermining it: the original question, "who should be targeted, men or women?", cannot yet be answered with confidence from this evidence base, but the analysis makes clear that a more tractable and ultimately more policy-useful question is emerging in its place , namely, which specific design features (dosage, delivery modality, facilitator training, theoretical grounding, and measurement approach) systematically distinguish the high-performing studies within each subgroup from the low-performing ones. Answering that second question, through the moderator analyses proposed in Section 6.6, would do more to advance evidence-based programming in India than a larger, better-powered replication of the simple men-versus-women comparison alone.

8. Limitations

This meta-analysis is subject to several limitations that should temper the strength of any conclusion drawn from it. First and most consequentially, statistical power for both the subgroup-specific pooled estimates and the between-subgroup comparison is limited by the modest number of studies in each cell ($k = 10$ and $k = 12$); combined with the extreme observed heterogeneity, this means the analysis is better powered to detect that the men-versus-women question remains open than to detect a true difference between the two approaches, even if a moderate true difference exists. Second, the near-total heterogeneity documented in Section 6.3 substantially limits the interpretability of the pooled point estimates themselves: a random-effects mean computed across studies with I^2 near 99% describes a distribution of true effects too wide to be a reliable single number for policy planning, and should be read alongside the prediction intervals in Table 1 rather than in isolation. Third, the funnel plot presented in Section 6.5 was constructed from only 14 of the 22 included studies owing to missing variance data, which both limits the power of any publication-bias diagnosis and means the bias assessment does not represent the full analytic sample; a formal Egger's test and trim-and-fill sensitivity analysis, recommended in Section 6.5, had not been completed at the time of this interpretation. Fourth, "target group" as coded in this review is likely confounded with intervention modality: men-targeted studies in this literature disproportionately include school curricula, sports-mentorship, and peer-education formats, while women-targeted studies disproportionately include SHG and livelihoods-based formats, so the subgroup contrast partly reflects differences in delivery model as well

as differences in who was targeted, and the two cannot be fully disentangled without the moderator analysis proposed in Section 6.6. Fifth, outcome measurement varies across studies, most use the GEM Scale or a derivative, but adaptation, translation, and item selection differ enough across the Indian evaluations that full measurement equivalence cannot be assumed, which is itself a plausible contributor to the observed heterogeneity. Sixth, most included evaluations measure attitudes immediately or shortly after program completion; the durability of attitude change beyond the immediate post-test window remains poorly evidenced across both subgroups, and this synthesis cannot speak to whether the effects reported here persist, decay, or grow over time. Finally, eligibility was restricted to English-language publications and did not include a formal search of Hindi- or other regional-language grey literature, which may have excluded relevant state- or district-level program evaluations and introduces a possible language-related selection bias in addition to the publication bias assessed in Section 6.5.

9. Policy Implications and Practical Recommendations

Notwithstanding these limitations, the synthesis supports several practical, appropriately hedged recommendations for Indian policymakers, program funders, and implementing organizations.

First, resist a wholesale reallocation of programming budgets toward either target group on the strength of this evidence alone. The pooled results neither justify defunding women-centered empowerment programming, which serves independently valuable goals of economic security and agency beyond attitude change, nor justify an uncritical scale-up of men-and-boys programming on the assumption that its comparative superiority is established; the honest state of the evidence is "promising but statistically unresolved" for the latter, not "proven." Where budgets allow, the more defensible near-term strategy is a portfolio approach that continues to invest in both target-group models while sharply increasing the rigor and standardization of accompanying evaluation, rather than a premature pivot to one model at the expense of the other.

Second, prioritize investment in understanding design, not just target group. Given that heterogeneity within each subgroup dwarfs the difference between subgroups, funders are likely to get more actionable evidence per rupee spent by commissioning comparative-design trials, for example, testing dosage (a six-week versus six-month curriculum), delivery modality (peer-educator versus professional-facilitator models), or setting (school- versus community-based delivery) within a fixed target group, than by commissioning another simple men-versus-women comparison. This also argues for building a shared, minimum core measurement protocol (e.g., a common GEM Scale core-item set) across future Indian gender-attitude evaluations, both to serve local program learning and to allow future meta-analyses to pool with substantially less measurement-driven heterogeneity than the present one.

Third, treat the strongest individual trials (Yaari Dosti, GEMS, Parivartan) as hypothesis-generating design templates rather than as generically replicable models. Because the pooled men-targeted effect is large in principle but highly variable in practice, the policy-relevant task is to identify precisely which structural features of these landmark programs (curriculum content, facilitator training intensity, group-based peer dynamics, integration with a valued activity such as sport or schooling) are actively responsible for their outsized results, and to test whether those specific features transfer when the program is adapted to a new state, language, or implementing partner , rather than assuming the program's past success travels automatically with its name or brand.

Fourth, build durability and cost-effectiveness into the next generation of Indian trials. Because nearly all pooled effects in this review are drawn from short post-intervention windows, government and donor evaluation guidelines should require at minimum a 12-month follow-up measurement before an intervention is credited with lasting attitude change, and cost-per-unit-of-attitude-shift data should be collected routinely alongside effect sizes, so that future comparisons between men-targeted and women-targeted (or combined) models can weigh effectiveness against implementation cost rather than effect size alone.

Fifth, close the evidence gap on combined and family-level programming. The complete absence of eligible combined-target interventions in this review (Section 6.1) is itself a policy-relevant finding: despite strong theoretical arguments from both the empowerment literature (Kabeer, 1999) and the engaging-men literature (Heise, 1998; Barker, Ricardo, & Nascimento, 2007) that household-level attitude change likely requires engaging both spouses or both genders concurrently, India's evaluated intervention landscape remains almost entirely segregated by target group. Piloting and rigorously evaluating couple- or household-level curricula, rather than assuming that parallel single-sex programming will aggregate into household-level change, represents one of the more promising and currently untested directions for Indian gender-attitude programming.

Sixth, invest in the data infrastructure that would let the next iteration of this meta-analysis answer the question this one could not. Requiring standardized reporting of means, standard deviations, and sample sizes by group and time point, rather than only significance tests, in all government- and donor-funded evaluation reports would directly address the missing-data problem that constrained the publication-bias assessment in Section 6.5, and would allow future syntheses to run the full moderator and meta-regression analyses recommended in Section 6.6 without the delays and data-access barriers encountered here.

10. References

Achyut, P., Bhatla, N., Khandekar, S., Maitra, S., & Verma, R. K. (2011). Building support for gender equality among young adolescents in school: Findings from Mumbai, India. International Center for Research on Women.

Barker, G., Ricardo, C., & Nascimento, M. (2007). Engaging men and boys in changing gender-based inequity in health: Evidence from programme interventions. World Health Organization.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to meta-analysis. Wiley.

Connell, R. W. (1995). Masculinities. University of California Press.

Das, M., Verma, R., Ghosh, S., Ciaravino, S., Jones, K., O'Connor, B., & Miller, E. (2015). Community mentors as coaches: Transforming gender norms through cricket among adolescent males in urban India. Gender & Development, 23(1), 61–75. https://doi.org/10.1080/13552074.2015.1013675

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. Controlled Clinical Trials, 7(3), 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. Journal of Epidemiology and Community Health, 52(6), 377–384. https://doi.org/10.1136/jech.52.6.377

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. BMJ, 315(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Heise, L. (1998). Violence against women: An integrated, ecological framework. Violence Against Women, 4(3), 262–290. https://doi.org/10.1177/1077801298004003002

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. Statistics in Medicine, 21(11), 1539–1558. https://doi.org/10.1002/sim.1186

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. BMJ, 327(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

JASP Team. (2026). JASP (Version 0.98.0) [Computer software]. https://jasp-stats.org

Jejeebhoy, S. J. (1995). Women's education, autonomy, and reproductive behaviour: Experience from developing countries. Clarendon Press.

Jejeebhoy, S. J., & Sathar, Z. A. (2001). Women's autonomy in India and Pakistan: The influence of religion and region. Population and Development Review, 27(4), 687–712. https://doi.org/10.1111/j.1728-4457.2001.00687.x

Kabeer, N. (1999). Resources, agency, achievements: Reflections on the measurement of women's empowerment. Development and Change, 30(3), 435–464. https://doi.org/10.1111/1467-7660.00125

Kabeer, N. (2005). Gender equality and women's empowerment: A critical analysis of the third millennium development goal. Gender & Development, 13(1), 13–24.

Kalokhe, A. S., del Rio, C., Dunkle, K., Stephenson, R., Metheny, N., Paranjape, A., & Sahay, S. (2017). Domestic violence against women in India: A systematic review of a decade of quantitative studies. Global Public Health, 12(4), 498–513. https://doi.org/10.1080/17441692.2015.1119293

Miller, E., Das, M., Tancredi, D. J., McCauley, H. L., Virata, M. C. D., Nettiksimmons, J., O'Connor, B., Ghosh, S., & Verma, R. (2014). Evaluation of a gender-based violence prevention program for student athletes in Mumbai, India. Journal of Interpersonal Violence, 29(4), 758–778.

Miller, E., Tancredi, D. J., McCauley, H. L., Decker, M. R., Virata, M. C. D., Anderson, H. A., Stetkevich, N., Brown, E., Moideen, F., & Silverman, J. G. (2012). "Coaching boys into men": A cluster-randomized controlled trial of a dating violence prevention program. Journal of Adolescent Health, 51(5), 431–438.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLOS Medicine, 6(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Pulerwitz, J., & Barker, G. (2008). Measuring attitudes toward gender norms among young men in Brazil: Development and psychometric evaluation of the GEM Scale. Men and Masculinities, 10(3), 322–338.

Ricardo, C., Eads, M., & Barker, G. (2011). Engaging boys and young men in the prevention of sexual violence: A systematic and global review of evaluated interventions. Promundo-US / Sexual Violence Research Initiative.

Riley, R. D., Higgins, J. P. T., & Deeks, J. J. (2011). Interpretation of random effects meta-analyses. BMJ, 342, d549. https://doi.org/10.1136/bmj.d549

Sanyal, P. (2009). From credit to collective action: The role of microfinance in promoting women's social capital and normative influence. American Sociological Review, 74(4), 529–550. https://doi.org/10.1177/000312240907400402

Verma, R. K., Pulerwitz, J., Mahendra, V. S., Khandekar, S., Barker, G., Fulpagare, P., & Singh, S. K. (2006). Challenging and changing gender attitudes among young men in Mumbai, India. Reproductive Health Matters, 14(28), 135–143. https://doi.org/10.1016/S0968-8080(06)28261-2

Verma, R., Pulerwitz, J., Mahendra, V. S., Khandekar, S., Singh, A. K., Das, S. S., Mehra, S., Nura, A., & Barker, G. (2008). Promoting gender equity as a strategy to reduce HIV risk and gender-based violence among young men in India (Horizons Final Report). Population Council.

****