**Supplementary Online Materials**

|  |
| --- |
| Section S1. Multi-state and Multi-county Study Sensitivity Analyses  References S1. References for Studies with Samples of Majority Girls |
| Section S2. Description of Cultural Sexism Measure and Scores |
| Table S1. Factor Analysis of Cultural Sexism: Item Selection, Response Options, Sources, and Years  Table S2. Factor Scores for All 50 States plus Washington, D.C. |
| Section S3. Addressing Contextual Confounders and Study Characteristics  Table S3: Correlation Matrix of Contextual Variables  Section S4. Examination of Clustering by Study  Section S5. R Analysis Code |
|  |
|  |

This supplementary material has been provided by the authors to give readers additional information about their work.

**Section S1: Multi-State and Multi-County Study Sensitivity Analyses**

In addition to the main analyses, we conducted additional sensitivity analyses. Nine studies included multiple states; for these studies, an aggregate cultural sexism score was created to represent all states in which the study was conducted. In the sensitivity analysis, we removed the 9 studies that occurred in multiple states to ensure that their inclusion did not influence the direction or magnitude of the findings. Results of the sensitivity analysis excluding 9 multistate studies (*k*=29 with aggregated cultural sexism scores) suggested that their inclusion did not impact findings, as the analysis produced almost identical results to the primary analyses (adjusted β=-0.07, CI: -0.13, -0.01, *p*<0.05).

We replicated this analysis at the county level. Among the 93 studies, we identified 17 in which treatment was conducted in multiple counties (e.g., multi-site studies). For these, as in the state-level analyses, we created aggregated scores of cultural sexism across the counties included in the study. We also removed these studies to confirm that their inclusion did not bias the results, and results of this analysis revealed very similar results to the county-level analysis including the full sample (β=-0.08, CI: -0.22, 0.07, *p*=.30), suggesting that the inclusion of studies with aggregated cultural sexism scores did not affect our findings.

**References S1. References for Studies with Samples of Majority Girls\***

Ackerson, J., Scogin, F., McKendree-Smith, N., & Lyman, R. D. (1998). Cognitive bibliotherapy for mild and moderate adolescent depressive symptomatology. *Journal of Consulting and Clinical Psychology*, *66*(4), 685–690. https://doi.org/10.1037//0022-006x.66.4.685

Asarnow, J. R., Baraff, L. J., Berk, M., Grob, C. S., Devich-Navarro, M., Suddath, R., Piacentini, J. C., Rotheram-Borus, M. J., Cohen, D., & Tang, L. (2011). An emergency department intervention for linking pediatric suicidal patients to follow-up mental health treatment. *Psychiatric Services*, *62*(11), 1303–1309. https://doi.org/10.1176/ps.62.11.pss6211\_1303

Asarnow, J. R., Hughes, J. L., Babeva, K. N., & Sugar, C. A. (2017). Cognitive-behavioral family treatment for suicide attempt prevention: A randomized controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *56*(6), 506–514. https://doi.org/10.1016/j.jaac.2017.03.015

Asarnow, J. R., Jaycox, L. H., Duan, N., LaBorde, A. P., Rea, M. M., Murray, P., Anderson, M., Landon, C., Tang, L., & Wells, K. B. (2005). Effectiveness of a quality improvement intervention for adolescent depression in primary care clinics: A randomized controlled trial. *JAMA*, *293*(3), 311–319. https://doi.org/10.1001/jama.293.3.311

Asarnow, J. R., Porta, G., Spirito, A., Emslie, G., Clarke, G., Wagner, K. D., Vitiello, B., Keller, M., Birmaher, B., McCracken, J., Mayes, T., Berk, M., & Brent, D. A. (2011). Suicide attempts and nonsuicidal self-injury in the treatment of resistant depression in adolescents: Findings from the TORDIA study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *50*(8), 772–781. https://doi.org/10.1016/j.jaac.2011.04.003

Auslander, W., McGinnis, H., Tlapek, S., Smith, P., Foster, A., Edmond, T., & Dunn, J. (2017). Adaptation and implementation of a trauma-focused cognitive behavioral intervention for girls in child welfare. *American Journal of Orthopsychiatry*, *87*(3), 206–215. https://doi.org/10.1037/ort0000233

Bandura, A., & Menlove, F. L. (1968). Factors determining vicarious extinction of avoidance behavior through symbolic modeling. *Journal of Personality and Social Psychology*, *8*(2), 99–108. https://doi.org/10.1037/h0025260

Beidel, D. C., Turner, S. M., & Morris, T. L. (2000). Behavioral treatment of childhood social phobia. *Journal of Consulting and Clinical Psychology*, *68*(6), 1072–1080. https://doi.org/10.1037/0022-006X.68.6.1072

Bender, N. N. (1976). Self-verbalization versus tutor verbalization in modifying impulsivity. *Journal of Educational Psychology*, *68*(3), 347–354. https://doi.org/10.1037/0022-0663.68.3.347

Bleck, R. T., & Bleck, B. L. (1982). The disruptive child’s play group. *Elementary School Guidance & Counseling*, *17*(2), 137–141.

Block, J. (1978). Effects of a rational–emotive mental health program on poorly achieving, disruptive high school students. *Journal of Counseling Psychology*, *25*(1), 61–65. https://doi.org/10.1037/0022-0167.25.1.6

Brent, D. A., Holder, D., Kolko, D., Birmaher, B., Baugher, M., Roth, C., Iyengar, S., & Johnson, B. A. (1997). A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive therapy. *Archives of General Psychiatry*, *54*(9), 877–885. https://doi.org/10.1001/archpsyc.1997.01830210125017

Carbonell, D. M., & Parteleno-Barehmi, C. (1999). Psychodrama groups for girls coping with trauma. *International Journal of Group Psychotherapy*, *49*(3), 285–306. https://doi.org/10.1080/00207284.1999.11732607

Chemtob, C. M., Nakashima, J., & Carlson, J. G. (2002). Brief treatment for elementary school children with disaster-related posttraumatic stress disorder: A field study. *Journal of Clinical Psychology*, *58*(1), 99–112. https://doi.org/10.1002/jclp.1131

Clarke, G. N., Hawkins, W., Murphy, M., Sheeber, L. B., Lewinsohn, P. M., & Seeley, J. R. (1995). Targeted prevention of unipolar depressive disorder in an at-risk sample of high school adolescents: A randomized trial of a group cognitive intervention. *Journal of the American Academy of Child and Adolescent Psychiatry*, *34*(3), 312–321. https://doi.org/10.1097/00004583-199503000-00016

Clarke, G. N., Hornbrook, M., Lynch, F., Polen, M., Gale, J., Beardslee, W., O’Connor, E., & Seeley, J. (2001). A randomized trial of a group cognitive intervention for preventing depression in adolescent offspring of depressed parents. *Archives of General Psychiatry*, *58*(12), 1127–1134. https://doi.org/10.1001/archpsyc.58.12.1127

Clarke, GREGORY N., Hornbrook, M., Lynch, F., Polen, M., Gale, J., O’connor, E., Seeley, J. R., & Debar, L. (2002). Group cognitive-behavioral treatment for depressed adolescent offspring of depressed parents in a health maintenance organization. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*(3), 305–313. https://doi.org/10.1097/00004583-200203000-00010

Clarke, Gregory N., Rohde, P., Lewinsohn, P. M., Hops, H., & Seeley, J. R. (1999). Cognitive-behavioral treatment of adolescent depression: Efficacy of acute group treatment and booster sessions. *Journal of the American Academy of Child & Adolescent Psychiatry*, *38*(3), 272–279. https://doi.org/10.1097/00004583-199903000-00014

Cohen, J. A., Deblinger, E., Mannarino, A. P., & Steer, R. (2004). A multi-site, randomized controlled trial for children with abuse-related PTSD symptoms. *Journal of the American Academy of Child and Adolescent Psychiatry*, *43*(4), 393–402.

Cohen, J. A., Mannarino, A. P., & Iyengar, S. (2011). Community treatment of posttraumatic stress disorder for children exposed to intimate partner violence: A randomized controlled trial. *Archives of Pediatrics & Adolescent Medicine*, *165*(1), 16–21. https://doi.org/10.1001/archpediatrics.2010.247

Conniff, K. M., Scarlett, J. M., Goodman, S., & Appel, L. D. (2005). Effects of a pet visitation program on the behavior and emotional state of adjudicated female adolescents. *Anthrozoös*, *18*(4), 379–395. https://doi.org/10.2752/089279305785593974

Cradock, C., Cotler, S., & Jason, L. A. (1978). Primary prevention: Immunization of children for speech anxiety. *Cognitive Therapy and Research*, *2*(4), 389–396. https://doi.org/10.1007/BF01172656

de Anda, D. (1985). Structured vs. nonstructured groups in the teaching of problem solving. *Children & Schools*, *7*(2), 80–89. https://doi.org/10.1093/cs/7.2.80

Deblinger, E., Lippmann, J., & Steer, R. (2016). Sexually abused children suffering posttraumatic stress symptoms: Initial treatment outcome findings: *Child Maltreatment*. https://doi.org/10.1177/1077559596001004003

Diamond, G. S., Wintersteen, M. B., Brown, G. K., Diamond, G. M., Gallop, R., Shelef, K., & Levy, S. (2010). Attachment-based family therapy for adolescents with suicidal ideation: A randomized controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*(2), 122–131. https://doi.org/10.1097/00004583-201002000-00006

Diamond, G., Siqueland, L., & Diamond, G. M. (2003). Attachment-based family therapy for depressed adolescents: Programmatic treatment development. *Clinical Child and Family Psychology Review*, *6*(2), 107–127. https://doi.org/10.1023/A:1023782510786

Dietz, L. J., Weinberg, R. J., Brent, D. A., & Mufson, L. (2015). Family-based interpersonal psychotherapy for depressed preadolescents: Examining efficacy and potential treatment mechanisms. *Journal of the American Academy of Child and Adolescent Psychiatry*, *54*(3), 191–199. https://doi.org/10.1016/j.jaac.2014.12.011

Donaldson, D., Spirito, A., & Esposito-Smythers, C. (2005). Treatment for adolescents following a suicide attempt: Results of a pilot trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*(2), 113–120. https://doi.org/10.1097/00004583-200502000-00003

Ehrenreich-May, J., Rosenfield, D., Queen, A. H., Kennedy, S. M., Remmes, C. S., & Barlow, D. H. (2017). An initial waitlist-controlled trial of the unified protocol for the treatment of emotional disorders in adolescents. *Journal of Anxiety Disorders*, *46*, 46–55. https://doi.org/10.1016/j.janxdis.2016.10.006

Fantuzzo, J., Sutton-Smith, B., Atkins, M., Meyers, R., Stevenson, H., Coolahan, K., Weiss, A., & Manz, P. (1996). Community-based resilient peer treatment of withdrawn maltreated preschool children. *Journal of Consulting and Clinical Psychology*, *64*(6), 1377–1386. https://doi.org/10.1037/0022-006X.64.6.1377

Fantuzzo, J. W., Jurecic, L., Stovall, A., Hightower, A. D., Goins, C., & Schachtel, D. (1988). Effects of adult and peer social initiations on the social behavior of withdrawn, maltreated preschool children. *Journal of Consulting and Clinical Psychology*, *56*(1), 34–39. https://doi.org/10.1037/0022-006X.56.1.34

Fox, J. E., & Houston, B. K. (1981). Efficacy of self-instructional training for reducing children’s anxiety in an evaluative situation. *Behaviour Research and Therapy*, *19*(6), 509–515. https://doi.org/10.1016/0005-7967(81)90077-2

Franklin, M., Foa, E., & March, J. S. (2003). The pediatric obsessive-compulsive disorder treatment study: Rationale, design, and methods. *Journal of Child and Adolescent Psychopharmacology*, *13*(supplement 1), 39–51. https://doi.org/10.1089/104454603322126331

Freeman, J. B., Garcia, A. M., Coyne, L., Ale, C., Przeworski, A., Himle, M., Compton, S., & Leonard, H. L. (2008). Early childhood OCD: Preliminary findings from a family-based cognitive-behavioral approach. *Journal of the American Academy of Child and Adolescent Psychiatry*, *47*(5), 593–602. https://doi.org/10.1097/CHI.0b013e31816765f9

Freeman, J., Sapyta, J., Garcia, A., Compton, S., Khanna, M., Flessner, C., FitzGerald, D., Mauro, C., Dingfelder, R., Benito, K., Harrison, J., Curry, J., Foa, E., March, J., Moore, P., & Franklin, M. (2014). Family-based treatment of early childhood obsessive-compulsive disorder: The Pediatric Obsessive-Compulsive Disorder Treatment Study for Young Children (POTS Jr)--a randomized clinical trial. *JAMA Psychiatry*, *71*(6), 689–698. https://doi.org/10.1001/jamapsychiatry.2014.170

Furman, W., Rahe, D. F., & Hartup, W. W. (1979). Rehabilitation of socially withdrawn preschool children through mixed-age and same-age socialization. *Child Development*, *50*(4), 915–922. JSTOR. https://doi.org/10.2307/1129315

Gaesser, A. H., & Karan, O. C. (2017). A randomized controlled comparison of emotional freedom technique and cognitive-behavioral therapy to reduce adolescent anxiety: A pilot study. *The Journal of Alternative and Complementary Medicine*, *23*(2), 102–108. https://doi.org/10.1089/acm.2015.0316

Gallagher, H. M., Rabian, B. A., & McCloskey, M. S. (2004). A brief group cognitive-behavioral intervention for social phobia in childhood. *Journal of Anxiety Disorders*, *18*(4), 459–479. https://doi.org/10.1016/S0887-6185(03)00027-6

Garber, J., Clarke, G. N., Weersing, V. R., Beardslee, W. R., Brent, D. A., Gladstone, T. R. G., DeBar, L. L., Lynch, F. L., D’Angelo, E., Hollon, S. D., Shamseddeen, W., & Iyengar, S. (2009). Prevention of depression in at-risk adolescents: A randomized controlled trial. *JAMA*, *301*(21), 2215–2224. https://doi.org/10.1001/jama.2009.788

Gillham, J. E., Hamilton, J., Freres, D. R., Patton, K., & Gallop, R. (2006). Preventing depression among early adolescents in the primary care setting: A Randomized Controlled Study of the Penn Resiliency Program. *Journal of Abnormal Child Psychology*, *34*(2), 195–211. https://doi.org/10.1007/s10802-005-9014-7

Ginsburg, G. S., & Drake, K. L. (2002). School-based treatment for anxious African-American adolescents: A controlled pilot study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*(7), 768–775. https://doi.org/10.1097/00004583-200207000-00007

Ginther, L., & Roberts, M. (2008). A test of mastery versus coping modeling in the reduction of children’s dental fears. *Child & Family Behavior Therapy*, *4*, 41–52. https://doi.org/10.1300/J019v04n04\_04

Guerra, N. G., & Slaby, R. G. (1990). Cognitive mediators of aggression in adolescent offenders: II. Intervention. *Developmental Psychology*, *26*(2), 269–277. https://doi.org/10.1037/0012-1649.26.2.269

Herbert, J. D., Gaudiano, B. A., Rheingold, A. A., Moitra, E., Myers, V. H., Dalrymple, K. L., & Brandsma, L. L. (2009). Cognitive behavior therapy for generalized social anxiety disorder in adolescents: A randomized controlled trial. *Journal of Anxiety Disorders*, *23*(2), 167–177. https://doi.org/10.1016/j.janxdis.2008.06.004

Jouriles, E. N., McDonald, R., Rosenfield, D., Stephens, N., Corbitt-Shindler, D., & Miller, P. C. (2009). Reducing conduct problems among children exposed to intimate partner violence: A randomized clinical trial examining effects of Project Support. *Journal of Consulting and Clinical Psychology*, *77*(4), 705–717. https://doi.org/10.1037/a0015994

Kahn, J. S., Kehle, T. J., Jenson, W. R., & Clark, E. (1990). Comparison of cognitive-behavioral, relaxation, and self-modeling interventions for depression among middle-school students. *School Psychology Review*, *19*(2), 196–211.

Kass, R. G., & Fish, J. M. (1991). Positive reframing and the test performance of test anxious children. *Psychology in the Schools*, *28*(1), 43–52. https://doi.org/10.1002/1520-6807(199101)28:1<43::AID-PITS2310280108>3.0.CO;2-I

Kerr, D. C. R., Leve, L. D., & Chamberlain, P. (2009). Pregnancy rates among juvenile justice girls in two randomized controlled trials of multidimensional treatment foster care. *Journal of Consulting and Clinical Psychology*, *77*(3), 588–593. https://doi.org/10.1037/a0015289

King, C. A., Gipson, P. Y., Horwitz, A. G., & Opperman, K. J. (2015). Teen options for change: An intervention for young emergency patients who screen positive for suicide risk. *Psychiatric Services (Washington, D.C.)*, *66*(1), 97–100. https://doi.org/10.1176/appi.ps.201300347

King, C. A., Kramer, A., Preuss, L., Kerr, D. C. R., Weisse, L., & Venkataraman, S. (2006). Youth-Nominated Support Team for suicidal adolescents (Version 1): A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *74*(1), 199–206. https://doi.org/10.1037/0022-006X.74.1.199

Last, C. G., Hansen, C., & Franco, N. (1998). Cognitive-behavioral treatment of school phobia. *Journal of the American Academy of Child & Adolescent Psychiatry*, *37*(4), 404–411. https://doi.org/10.1097/00004583-199804000-00018

Leitenberg, H., & Callahan, E. J. (1973). Reinforced practice and reduction of different kinds of fears in adults and children. *Behaviour Research and Therapy*, *11*(1), 19–30. https://doi.org/10.1016/0005-7967(73)90065-X

Leve, L. D., Chamberlain, P., & Reid, J. B. (2005). Intervention outcomes for girls referred from juvenile justice: Effects on delinquency. *Journal of Consulting and Clinical Psychology*, *73*(6), 1181–1184. https://doi.org/10.1037/0022-006X.73.6.1181

Lewinsohn, P. M., Clarke, G. N., Hops, H., & Andrews, J. (1990). Cognitive-behavioral treatment for depressed adolescents. *Behavior Therapy*, *21*(4), 385–401. https://doi.org/10.1016/S0005-7894(05)80353-3

Mann, J., & Rosenthal, T. L. (1969). Vicarious and direct counterconditioning of test anxiety through individual and group desensitization. *Behaviour Research and Therapy*, *7*(4), 359–367. https://doi.org/10.1016/0005-7967(69)90065-5

Masia Warner, C., Colognori, D., Brice, C., Herzig, K., Mufson, L., Lynch, C., Reiss, P. T., Petkova, E., Fox, J., Moceri, D. C., Ryan, J., & Klein, R. G. (2016). Can school counselors deliver cognitive-behavioral treatment for social anxiety effectively? A randomized controlled trial. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *57*(11), 1229–1238. https://doi.org/10.1111/jcpp.12550

Masia Warner, C., Fisher, P. H., Shrout, P. E., Rathor, S., & Klein, R. G. (2007). Treating adolescents with social anxiety disorder in school: An attention control trial. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *48*(7), 676–686. https://doi.org/10.1111/j.1469-7610.2007.01737.x

Masia-Warner, C., Klein, R. G., Dent, H. C., Fisher, P. H., Alvir, J., Marie Albano, A., & Guardino, M. (2005). School-based intervention for adolescents with social anxiety disorder: Results of a controlled study. *Journal of Abnormal Child Psychology*, *33*(6), 707–722. https://doi.org/10.1007/s10802-005-7649-z

McCarty, C. A., Violette, H. D., Duong, M. T., Cruz, R. A., & McCauley, E. (2013). A randomized trial of the Positive Thoughts and Action program for depression among early adolescents. *Journal of Clinical Child and Adolescent Psychology: The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*, *42*(4), 554–563. https://doi.org/10.1080/15374416.2013.782817

Milos, M. E., & Reiss, S. (1982). Effects of three play conditions on separation anxiety in young children. *Journal of Consulting and Clinical Psychology*, *50*(3), 389–395. https://doi.org/10.1037/0022-006X.50.3.389

Mufson, L., Weissman, M. M., Moreau, D., & Garfinkel, R. (1999). Efficacy of interpersonal psychotherapy for depressed adolescents. *Archives of General Psychiatry*, *56*(6), 573–579. https://doi.org/10.1001/archpsyc.56.6.573

Murphy, C. M., & Richard R., B. (1973). Active and passive participation in the contact desensitization of snake fear in children. *Behavior Therapy*, *4*(2), 203–211. https://doi.org/10.1016/S0005-7894(73)80029-2

Najavits, L. M., Gallop, R. J., & Weiss, R. D. (2006). Seeking safety therapy for adolescent girls with PTSD and substance use disorder: A randomized controlled trial. *The Journal of Behavioral Health Services & Research*, *33*(4), 453–463. https://doi.org/10.1007/s11414-006-9034-2

Pfiffner, L. J., Jouriles, E. N., Brown, M. M., Etscheidt, M. A., & Kelly, J. A. (1990). Effects of problem-solving therapy on outcomes of parent training for single-parent families. *Child & Family Behavior Therapy*, *12*(1), 1–11. https://doi.org/10.1300/J019v12n01\_01

Pincus, D. B., May, J. E., Whitton, S. W., Mattis, S. G., & Barlow, D. H. (2010). Cognitive-behavioral treatment of panic disorder in adolescence. *Journal of Clinical Child & Adolescent Psychology*, *39*(5), 638–649. https://doi.org/10.1080/15374416.2010.501288

Porter, B., & Hoedt, K. (1985). Differential effects of an Adlerian counseling approach with preadolescent children. *Individual Psychology: Journal of Adlerian Theory, Research & Practice*, 372–385.

Randell, B. P., Eggert, L. L., & Pike, K. C. (2001). Immediate post intervention effects of two brief youth suicide prevention interventions. *Suicide and Life-Threatening Behavior*, *31*(1), 41–61. https://doi.org/10.1521/suli.31.1.41.21308

Redfering, D. L. (1972). Group counseling with institutionalized delinquent females. *American Corrective Therapy Journal*, *26*(6), 160–163.

Reynolds, W. M., & Coats, K. I. (1986). A comparison of cognitive-behavioral therapy and relaxation training for the treatment of depression in adolescents. *Journal of Consulting and Clinical Psychology*, *54*(5), 653–660. https://doi.org/10.1037/0022-006X.54.5.653

Riemann, B. C., Kuckertz, J. M., Rozenman, M., Weersing, V. R., & Amir, N. (2013). Augmentation of youth cognitive behavioral and pharmacological interventions with attention modification: A preliminary investigation. *Depression and Anxiety*, *30*(9), 822–828. https://doi.org/10.1002/da.22127

Ritter, B. (n.d.). *Effect of contact desensitization on avoidance behavior, fear ratings, and self-evaluative statements.* *3*(527S40–528).

Ritter, B. (1968). The group desensitization of children’s snake phobias using vicarious and contact desensitization procedures. *Behaviour Research and Therapy*, *6*(1), 1–6. https://doi.org/10.1016/0005-7967(68)90033-8

Rohde, P., Stice, E., Shaw, H., & Brière, F. N. (2014). Indicated cognitive behavioral group depression prevention compared to bibliotherapy and brochure control: Acute effects of an effectiveness trial with adolescents. *Journal of Consulting and Clinical Psychology*, *82*(1), 65–74. https://doi.org/10.1037/a0034640

Rosenfarb, I., & Hayes, S. C. (1984). Social standard setting: The achilles heel of informational accounts of therapeutic change. *Behavior Therapy*, *15*(5), 515–528. https://doi.org/10.1016/S0005-7894(84)80053-2

Santucci, L. C., & Ehrenreich-May, J. (2013). A randomized controlled trial of the child anxiety multi-day program (CAMP) for separation anxiety disorder. *Child Psychiatry and Human Development*, *44*(3), 439–451. https://doi.org/10.1007/s10578-012-0338-6

Sheslow, D. V., Bondy, A. S., & Nelson, R. O. (1983). A comparison of graduated exposure, verbal coping skills, and their combination in the treatment of children’s fear of the dark. *Child & Family Behavior Therapy*, *4*(2–3), 33–45. https://doi.org/10.1300/J019v04n02\_03

Silverman, W. K., Kurtines, W. M., Jaccard, J., & Pina, A. A. (2009). Directionality of change in youth anxiety treatment involving parents: An initial examination. *Journal of Consulting and Clinical Psychology*, *77*(3), 474–485. https://doi.org/10.1037/a0015761

Southam-Gerow, M. A., Weisz, J. R., Chu, B. C., McLeod, B. D., Gordis, E. B., & Connor-Smith, J. K. (2010). Does cognitive behavioral therapy for youth anxiety outperform usual care in community clinics? An initial effectiveness test. *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*(10), 1043–1052. https://doi.org/10.1016/j.jaac.2010.06.009

Stein, B. D., Jaycox, L. H., Kataoka, S. H., Wong, M., Tu, W., Elliott, M. N., & Fink, A. (2003). A mental health intervention for schoolchildren exposed to violence: A randomized controlled trial. *JAMA*, *290*(5), 603–611. https://doi.org/10.1001/jama.290.5.603

Storch, E. A., Geffken, G. R., Merlo, L. J., Mann, G., Duke, D., Munson, M., Adkins, J., Grabill, K. M., Murphy, T. K., & Goodman, W. K. (2007). Family-based cognitive-behavioral therapy for pediatric obsessive-compulsive disorder: Comparison of intensive and weekly approaches. *Journal of the American Academy of Child and Adolescent Psychiatry*, *46*(4), 469–478. https://doi.org/10.1097/chi.0b013e31803062e7

Strayhorn, J. M., & Weidman, C. S. (1989). Reduction of attention deficit and internalizing symptoms in preschoolers through parent-child interaction training. *Journal of the American Academy of Child and Adolescent Psychiatry*, *28*(6), 888–896. https://doi.org/10.1097/00004583-198911000-00013

Szigethy, Eva, Bujoreanu, S. I., Youk, A. O., Weisz, J., Benhayon, D., Fairclough, D., Ducharme, P., Gonzalez-Heydrich, J., Keljo, D., Srinath, A., Bousvaros, A., Kirshner, M., Newara, M., Kupfer, D., & DeMaso, D. R. (2014). Randomized efficacy trial of two psychotherapies for depression in youth with inflammatory bowel disease. *Journal of the American Academy of Child and Adolescent Psychiatry*, *53*(7), 726–735. https://doi.org/10.1016/j.jaac.2014.04.014

Szigethy, EVA, Kenney, E., Carpenter, J., Hardy, D. M., Fairclough, D., Bousvaros, A., Keljo, D., Weisz, J., Beardslee, W. R., Noll, R., & DeMASO, D. R. (2007). Cognitive-behavioral therapy for adolescents with inflammatory bowel disease and subsyndromal depression. *Journal of the American Academy of Child & Adolescent Psychiatry*, *46*(10), 1290–1298. https://doi.org/10.1097/chi.0b013e3180f6341f

Treatment for Adolescents With Depression Study Team. (2003). Treatment for Adolescents With Depression Study (TADS): Rationale, design, and methods. *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*(5), 531–542. https://doi.org/10.1097/01.CHI.0000046839.90931.0D

Truax, C. B., Wargo, D. G., & Silber, L. D. (1966). Effects of group psychotherapy with high accurate empathy and nonpossessive warmth upon female institutionalized delinquents. *Journal of Abnormal Psychology*, *71*(4), 267–274. https://doi.org/10.1037/h0023590

Warren, R., Deffenbacher, J., & Brading, P. (1976). Rational-emotive therapy and the reduction of test anxiety in elementary school students. *Rational Living*, *11*, 26–29.

Wehr, S. H., & Kaufman, M. E. (1987). The effects of assertive training on performance in highly anxious adolescents. *Adolescence*, *22*(85), 195–205.

Weisz, J. R., Southam-Gerow, M. A., Gordis, E. B., Connor-Smith, J. K., Chu, B. C., Langer, D. A., McLeod, B. D., Jensen-Doss, A., Updegraff, A., & Weiss, B. (2009). Cognitive-behavioral therapy versus usual clinical care for youth depression: An initial test of transportability to community clinics and clinicians. *Journal of Consulting and Clinical Psychology*, *77*(3), 383–396. https://doi.org/10.1037/a0013877

White, W. C., & Davis, M. T. (1974). Vicarious extinction of phobic behavior in early childhood. *Journal of Abnormal Child Psychology*, *2*(1), 25–32. https://doi.org/10.1007/bf00919350

Williams, C. E., & Jones, R. T. (1989). Impact of self-instructions on response maintenance and children’s fear of fire. *Journal of Clinical Child Psychology*, *18*(1), 84–89. https://doi.org/10.1207/s15374424jccp1801\_10

Young, J. F., Mufson, L., & Davies, M. (2006). Efficacy of Interpersonal Psychotherapy-Adolescent Skills Training: An indicated preventive intervention for depression. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *47*(12), 1254–1262. https://doi.org/10.1111/j.1469-7610.2006.01667.x

Young, J. F., Mufson, L., & Gallop, R. (2010). Preventing depression: A randomized trial of interpersonal psychotherapy-adolescent skills training. *Depression and Anxiety*, *27*(5), 426–433. https://doi.org/10.1002/da.20664

*\** Reference list reflects references from all original randomized control trials included in the main analytic dataset (i.e., studies with samples of 50% or more girls).

**Section S2. Description of Cultural Sexism Measure and Scores**

*State-level cultural sexism*

To measure state-level cultural sexism, we drew from a series of 19 explicit or implicit attitudinal indicators across two data sources, Project Implicit and General Social Survey. Indicators were chosen for model consideration if they included coverage of all 50 states and Washington, D.C with sufficient Ns for all states (minimum *N* = 50). As described in the methods section, we first ran an exploratory factor analysis, then retained the following indicators (based on a 0.50 factor loading threshold) for confirmatory factor analysis to generate state scores:

Project Implicit modules on implicit sexism were obtained from the Gender-Career and Gender-Science versions of the Implicit Association Test (IAT; Nosek et al., 2007). These correspond to items 1 and 2 in Table S2 below. The Gender-Science and Gender-Career versions of the IAT measure strengths of associations between gender concepts (e.g., male and female) and gender norms (e.g., science versus liberal arts; and career versus family) by measuring response latencies during a computer-administered categorization task. Across multiple trials, one gender concept (e.g., male vs female) and one gender norm concept (e.g., career vs family) are presented on the screen. Participants are asked to categorize the concept, by making one of two keystrokes (i.e., “e” or “i”) based on a set of rules assigned for each block of trials. During the first block of trials the rule requires that the concept of “male” and concept “career” are categorized with one key and the concept of “female” and concept of “family” are categorized with the other key. During a second block of trials, the rule requires that the concepts of “male” and “family” are categorized with one key, and the concepts of “female” and “career” are categorized with the other key. To score the task, the average response latency on the two blocks of trials is compared. A faster average response time during the first block compared to the second block reflects a stronger implicit association between male and career and/or female and family. A faster average response time during the second block compared to the first block reflects a stronger implicit association between male and family and/or female and career. A participant’s IAT score is the standardized difference in mean response time on the two blocks of trials described above. Individual IAT scores can range from -2 to +2, with a score of zero reflecting no implicit association and higher scores reflecting a stronger implicit association for traditional gender norms. More information on the Implicit Associations Test can be obtained at the Project Implicit website: https://implicit.harvard.edu/implicit/.

Six of the indicators (see items 3-8 in Table S2), accessed through the Project Implicit Gender-Science Module (<https://osf.io/y9hiq/>), measured cultural sexism by first presenting a prompt describing the gender disparity in science and engineering faculty positions at research universities and subsequently asking participants to rate the *importance of six factors that potentially explain this disparity.* After items 3 and 4 were reverse coded, higher scores reflected explanations of the disparity that downplayed social factors (e.g., gender stereotypes) in favor of other explanations (e.g., innate abilities, differences in work ethic). For example, one explanation offered for the disparity was that “different proportions of men and women are found among people with the very highest levels of math ability,” which suggests that the disparity is driven by a comparatively superior innate math ability in men compared to women. As a result, higher ratings of the importance of this factor in explaining the disparity represent a higher level of sexism. The next item (see item 9 in Table S2), also accessed through Project Implicit, measured participants’ expectations of male participation in STEM majors, with higher values representing higher levels of male participation.

The final two indicators were obtained from the General Social Survey (GSS), an ongoing, nationally-representative survey administered in the United States every two years that asks about attitudes towards socially and politically salient topics. The items we chose had been previously validated as measures of state-level sexism (Charles et al., 2018). These asked about attitudes towards women and women’s stereotypical gender roles. Four items had been routinely assessed in recent years and were considered for inclusion in the final model. Ultimately two of these four questions (“Tell me if you agree or disagree with this statement: Most men are better suited emotionally for politics than are most women,” and “It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family”; items 10 and 11 in Table S2) were included in the final model based on meeting the 0.50 factor loading threshold.

All measures were coded such that higher levels corresponded to higher values of cultural sexism. Measures were standardized to their average response value for all respondents, regardless of state of residence, and then aggregated to the state level so that every state’s measure was the mean of the standardized individual response for respondents residing in that state. While the indicators often represented different survey years (e.g., Project Implicit questions ranged from 2003 to 2018), we aggregated all responses to the state level regardless of year queried. Averaging the responses in this way allowed for all states to have a sizable number of respondents, regardless of sampling variation from year to year. This was deemed appropriate because, while national sexism has declined overall, state-level sexism has been shown to be consistent over time with regard to the relative ranking between states (i.e., states with higher sexism rank consistently higher than low-sexism states, regardless of overall downward trends; Charles et al., 2018).

The factor scores for all 50 states and D.C. ranged from a low of -2.61 (D.C.) to a maximum score of 1.90 (Alabama). The factor scores for each state are presented in Table A1 below. The final variables, along with their original response format, sources, and years represented, are included in Table A2 below.

*County-level cultural sexism*

To measure county-level cultural sexism, we utilized the same items as were included in the state measure, described above. There were 3 items that we included in the state measure that could not be included in the county measure: these were the 2 GSS attitudinal items, as GSS measures were not available to us at the county level. The only other item excluded from the county-level sexism measure was the Project Implicit Gender-Career IAT, as it was asked on a different survey than the other Project Implicit items and therefore did not have the same county-level representation; including it in our models dramatically reduced the number of counties that could receive a sexism score. Therefore, the final county-level sexism measure was composed of 8 of the 11 items from the state-level measure.

For each study, state and county was determined based on where the treatment was administered (e.g., the clinic identified in the methods section). For studies in which the treatment location was not explicitly stated, the location was determined based on the authors’ affiliation if all affiliations matched. For the 20 studies in which affiliations did not match, we contacted the authors via email (usually the corresponding or first author) and received responses from each one.

While it is possible that some participants did not live in the county and state where they received treatment, potential misclassification of the county or state would introduce random error and thus bias our results toward the null.

**References**

Charles KK, Guryan J, Pan J. *The Effects of Sexism on American Women: The Role of Norms vs.*

*Discrimination*. National Bureau of Economic Research; 2018. doi:10.3386/w24904

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., ... & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*(1), 36-88.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table S1. Factor Analysis of Cultural Sexism: Item Selection, Response Options, Sources, and Years** | | | | |
| *Item* | *Original scale prior to standardization* | *Source* | *Years available* | *Geography* |
| Gender-Career Implicit Association Test: Implicit Sexism | -2 to +2 (High scores = more implicit sexism) | Project Implicit: gender-career | 2005-2018 | State only |
| Gender-Science Implicit Association Test: Implicit Sexism | -2 to +2 (High scores = more implicit sexism) | Project Implicit: gender-science | 2003-2018 | State and county |
| Prompt: *Women hold a smaller portion of the science and engineering faculty positions at top research universities than do men. The following factors are sometimes offered as reasons for this difference. Please rate how important you think each factor is for explaining this difference.*  Item 3) On average, whether consciously or unconsciously, men are favored in hiring and promotion. (r) | Likert-style, ranging from “not at all important” to “extremely important” | Project Implicit: gender-science | 2003-2018 | State and county |
| Prompt: *Women hold a smaller portion of the science and engineering faculty positions at top research universities than do men. The following factors are sometimes offered as reasons for this difference. Please rate how important you think each factor is for explaining this difference.*  Item 4) Directly or indirectly, boys and girls tend to receive different levels of encouragement for developing scientific interest. (r) | Likert-style, ranging from “not at all important” to “extremely important” | Project Implicit: gender-science | 2003-2018 | State and county |
| Prompt: *Women hold a smaller portion of the science and engineering faculty positions at top research universities than do men. The following factors are sometimes offered as reasons for this difference. Please rate how important you think each factor is for explaining this difference.*  Item 5) Different proportions of men and women are found among people with the very highest levels of math ability. | Likert-style, ranging from “not at all important” to “extremely important” | Project Implicit: gender-science | 2003-2018 | State and county State and county |
| Prompt: *Women hold a smaller portion of the science and engineering faculty positions at top research universities than do men. The following factors are sometimes offered as reasons for this difference. Please rate how important you think each factor is for explaining this difference.*  Item 6) On average, men and women differ in their willingness to spend time away from their families. | Likert-style, ranging from “not at all important” to “extremely important” | Project Implicit: gender-science | 2003-2018 | State and county |
| Prompt: *Women hold a smaller portion of the science and engineering faculty positions at top research universities than do men. The following factors are sometimes offered as reasons for this difference. Please rate how important you think each factor is for explaining this difference.*  Item 7) On average, men and women differ in their willingness to devote the time required by such ‘high-powered’ positions. | Likert-style, ranging from “not at all important” to “extremely important” | Project Implicit: gender-science | 2003-2018 | State and county |
| Prompt: *Women hold a smaller portion of the science and engineering faculty positions at top research universities than do men. The following factors are sometimes offered as reasons for this difference. Please rate how important you think each factor is for explaining this difference*  Item 8) On average, men and women differ naturally in their scientific interest. | Likert-style, ranging from “not at all important” to “extremely important” | Project Implicit: gender-science | 2003-2018 | State and county |
| Item 9) Suppose that ten men at a typical U.S. university were picked at random. How many would you predict will graduate with a scientific major (science, technology, engineering, or mathematics)? | Numerical, ranging from a possible 0 to 10 | Project Implicit: gender-science | 2003-2018 | State and county |
| Item 10) It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family | Likert-style, ranging from strongly disagree to strongly agree | General Social Survey | 1977, 1985, 1986, 1988, 1989, 1990, 1991, 1993, 1994, 1996, 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012, 2014 | State only |
| Item 11) Tell me if you agree or disagree with this statement: Most men are better suited emotionally for politics than are most women | Likert-style, ranging from strongly disagree to strongly agree | General Social Survey | 1974, 1975, 1977, 1978, 1982, 1983, 1985, 1986, 1988, 1989, 1990, 1991, 1993, 1994, 1996, 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012, 2014 | State only |

*Note:* (r) indicates items that were reverse scored so that higher ratings represented higher levels of cultural sexism.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table S2. Factor Scores for All 50 States plus Washington, D.C.** | | | |
| **State name** | **Cultural Sexism Factor Score** | **State name** | **Cultural Sexism Factor Score** |
| 1. Alaska\* | -0.70 | 1. Montana | -0.32 |
| 1. Alabama | 1.90 | 1. North Carolina\* | 0.32 |
| 1. Arkansas | 1.39 | 1. North Dakota | 0.76 |
| 1. Arizona\* | 0.91 | 1. Nebraska\* | 0.39 |
| 1. California\* | -0.87 | 1. New Hampshire | -0.29 |
| 1. Colorado | -1.03 | 1. New Jersey\* | 0.19 |
| 1. Connecticut\* | -0.40 | 1. New Mexico | -0.57 |
| 1. District of Columbia | -2.61 | 1. Nevada | 0.53 |
| 1. Delaware | 0.10 | 1. New York\* | -1.12 |
| 1. Florida\* | 0.79 | 1. Ohio\* | 0.26 |
| 1. Georgia\* | 1.08 | 1. Oklahoma | 1.03 |
| 1. Hawaii\* | 0.07 | 1. Oregon\* | -1.81 |
| 1. Iowa | 0.46 | 1. Pennsylvania\* | -0.11 |
| 1. Idaho | 0.10 | 1. Rhode Island\* | -1.26 |
| 1. Illinois\* | -0.02 | 1. South Carolina\* | 1.06 |
| 1. Indiana\* | 0.33 | 1. South Dakota | 0.41 |
| 1. Kansas\* | 0.38 | 1. Tennessee\* | 0.55 |
| 1. Kentucky\* | 1.13 | 1. Texas\* | 0.71 |
| 1. Louisiana | 0.43 | 1. Utah\* | 0.65 |
| 1. Massachusetts\* | -1.66 | 1. Virginia\* | -0.49 |
| 1. Maryland\* | -0.92 | 1. Vermont\* | -1.91 |
| 1. Maine | -1.40 | 1. Washington\* | -1.66 |
| 1. Michigan\* | -0.22 | 1. Wisconsin\* | 0.23 |
| 1. Minnesota\* | -0.93 | 1. West Virginia | 1.35 |
| 1. Missouri\* | 0.32 | 1. Wyoming | 0.88 |
| 1. Mississippi | 1.61 |  |  |

\**Note:* States represented in the current study are denoted with an asterisk.

**Section S3: Addressing Contextual Confounders and Study Characteristics**

**Contextual Confounders**

Because cultural sexism is related to other state-level contextual factors that may also influence treatment efficacy, we investigated 6 contextual covariates to rule out these alternative explanations: population density is defined as the average population per square mile; percent foreign-born, defined as the percentage of the state population not born a U.S. citizen; percent non-Hispanic White; poverty rate, defined as the number of state residents living at or below the federal poverty line; median household income; and Gini coefficient, which is a summary index of income inequality ranging from 0 (perfect equality) to 100 (perfect inequality).

State-level estimates for population density, percent foreign born, and percent non-Hispanic White were available from the US Census in 10-year intervals; these measures were pooled and averaged across the 2000-2010 censuses. Poverty rate and median household income were available annually from the US Census Current Population Survey, and were pooled and averaged across all years 2000-2010. Gini coefficient was available from the US Census American Community Survey 5-year estimates, pooled and averaged for the years 2006-2010.

**Study- and Effect-Size Level Characteristics**

We examined one study-level variable and two effect size (ES)-level variables that predicted intervention efficacy in the larger meta-analytic database (Weisz et al., 2019): 1) targeted problem (e.g., internalizing problems); 2) informant (e.g., youth-reported symptoms); and 3) control condition (e.g., waitlist, usual care). The analyses described below were conducted using the state-level measure of cultural sexism, as the county-level sexism measure was not significantly associated with psychotherapy efficacy.

1. **Targeted problem:** To examine whether the mental health problem targeted by the treatment impacted our results, we added a binary variable to our model measuring whether a study focused on internalizing (i.e., depression, anxiety, or both) or externalizing (i.e., ADHD, conduct) problems. Results indicated that this variable was not significantly associated with effect size (*b* = -.002, *p* = 0.98), and its inclusion in the model did not affect the association between cultural sexism and treatment efficacy (*b* = -.07, *p* <.05).Further, we tested an interaction term between cultural sexism and targeted problem (i.e., internalizing vs. externalizing) which was not significant (*b* = .099, *p* = 0.41), indicating that results were similar across both outcomes.
2. **Informant:** Individual effect sizes were calculated from individual measures completed by different informants (e.g., caregiver-report, child self-report, teacher-report). We added this categorical variable as a moderator to our main model and found that it was not associated with treatment efficacy (overall *b*=-.003, *p*=.94). Further, its inclusion in the model did not affect the relationship between sexism and treatment efficacy (*b*=-.07, *p*<.05). We additionally tested an interaction term (cultural sexism by informant) and found that it was not significant (*bs:* -.10 to -.01, *ps*>.05), suggesting that results were similar across informants.
3. **Control condition:** Control condition (e.g., waitlist, placebo treatment) was added to the main model and results indicated that ESs were significantly higher when the control condition was case management (*b*=0.60, *p* <.01) or no treatment (*b*=0.36, *p*<.05). Including this variable minimally changed the association between cultural sexism and psychotherapy efficacy (*b*=-.085, *p*<.01), and thus was not included in the main model.

While treatment type did not predict treatment efficacy in the larger database (Weisz et al., 2017), some studies have found larger effects for youth behavioral therapies as compared to non-behavioral treatments (e.g., Weiss & Weisz, 1995; Weisz et al., 1995). To examine whether treatment type affected the association between cultural sexism and treatment efficacy, we added an interaction term representing cultural sexism by treatment type (a binary indicator measuring whether a study examined one or more nonbehavioral treatments or one or more behavioral treatments) to our main model. The interaction was non-significant (*b*=-.05, *p*=0.60), indicating that results were similar across these two treatment types.

Gender-client therapist match was associated with stronger therapeutic alliance and treatment completion in at least one study of adolescents (Wintersteen et al., 2005), though findings from adult studies are mixed (Bhati, 2014; Behn et al., 2018; Fujino, Okazaki, & Young, 1994; Shiner et al., 2017). We were unable to include therapist-gender in our analyses, as most RCTs in our study did not provide therapist-gender data. Accordingly, exploring this as a potential predictor of treatment effectiveness is suggested for future research.

References

Behn, A., Davanzo, A., & Errázuriz, P. (2018). Client and therapist match on gender, age, and income: Does match within the therapeutic dyad predict early growth in the therapeutic alliance? *Journal of Clinical Psychology*, *74*(9), 1403–1421. https://doi.org/10.1002/jclp.22616

Bhati, K. S. (2014). Effect of client-therapist gender match on the therapeutic relationship: an exploratory analysis. *Psychological Reports*, *115*(2), 565–583. https://doi.org/10.2466/21.02.PR0.115c23z1

Fujino, D. C., Okazaki, S., & Young, K. (1994). Asian-American women in the mental health system: An examination of ethnic and gender match between therapist and client. *Journal of Community Psychology*, *22*(2), 164–176. https://doi.org/10.1002/1520-6629(199404)22:2<164::AID-JCOP2290220211>3.0.CO;2-K

Shiner, B., Leonard Westgate, C., Harik, J. M., Watts, B. V., & Schnurr, P. P. (2017). Effect of patient-therapist gender match on psychotherapy retention among united states veterans with posttraumatic stress disorder. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*(5), 642–650. https://doi.org/10.1007/s10488-016-0761-2

Weiss, B., & Weisz, J. R. (1995). Relative effectiveness of behavioral versus nonbehavioral child psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 317–320. <http://dx.doi.org/10.1037/0022-006X> .63.2.317

Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin, 117*, 450–468. <http://dx.doi.org/10.1037/0033-2909.117.3.450>

Wintersteen, M., Mensinger, J., & Diamond, G. (2005). Do gender and racial differences between patient and therapist affect therapeutic alliance and treatment retention in adolescents? *Professional Psychology: Research and Practice*, *36*, 400–408. https://doi.org/10.1037/0735-7028.36.4.400

**Table S3. Correlation Matrix of Contextual Variables**

*Intercorrelations for contextual variables, effect size, and cultural sexism*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1. Effect size (g) |  |  |  |  |  |  |  |
| 2. Cultural sexism (state level) | -.13\*\* |  |  |  |  |  |  |
| 3. Gini coefficient | .08\* | .05 |  |  |  |  |  |
| 4. Median income | .09\*\* | -.28\*\* | -.16\*\* |  |  |  |  |
| 5. Population density | -.03 | .06 | .08\* | .55\*\* |  |  |  |
| 6. % foreign-born | -.01 | -.36\*\* | .54\*\* | .42\*\* | .18\*\* |  |  |
| 7. % Non-Hispanic White | .03 | .00 | -.51\*\* | -.23\*\* | -.06\*\* | -.83\*\* |  |
| 8. Poverty rate | -.08\* | .09\* | .55\*\* | -.68\*\* | -.48\*\* | .27\*\* | -.46\*\* |

\**p* < .05. \*\* *p* < .01

**Section S4. Examination of Clustering by Study**

As mentioned in the main text, our 2-level model included multiple outcomes from each study, thereby violating the assumption of independent ESs. Practically, this introduces two issues. First, a study could have multiple ESs, and thus these samples could not be considered truly independent. To address this issue, we ran a sensitivity analysis accounting for ES-dependency within studies. Specifically, we used robust variance estimation (RVE), which allows for the inclusion of dependent ESs by correcting the study standard errors to account for associations between ESs from the same sample (see R code in Section S5; Tanner-Smith & Tipton, 2014). RVE and other approaches that nest ESs within studies (e.g., 3-level meta-regression) are limited in computational power because the number of observations is reduced from total number of ESs (*k* = 702 in the present study) to the total number of studies (*N* = 93 in this study; Hedges, Tipton & Johnson, 2010; Tipton & Pustejovsky, 2015). Results from the nested models (unadjusted β=-0.074, *p*=0.15; adjusted β=-0.070, *p*=0.20) were identical in direction and magnitude to the 2-level models presented in the main text (unadjusted β=-0.074, *p*<0.01; adjusted β=-0.070, *p*<0.05), but did not reach statistical significance, likely due to the reduction in statistical power in this model (see changes in degrees of freedom in relevant code and results in Section S5 below).

Second, without fully accounting for dependency, a study with more ESs could have greater influence on the results than a study with fewer ESs. However, for dependency to introduce bias related to our research question (i.e., be a confounder), studies with more ESs would have to 1) be associated with our moderator—that is, studies with more ESs would have to be more likely to occur in higher cultural sexism states—and 2) be more likely to influence overall study ES (i.e., mean ES; the average magnitude of difference on all outcomes between the intervention vs. control groups) compared to studies with fewer ESs. However, we do not find evidence for either. Results from bivariate correlations indicated that the total number of ESs per study was not associated with state-level cultural sexism (*r* = 0.15, *p* = 0.15) or with mean study ES (*r* = -0.02, *p* = 0.81). In other words, studies with larger numbers of ESs were not associated with larger or smaller mean ES, nor were they more likely to occur in states with higher cultural sexism.

Taken together, both sets of analyses minimize concerns that dependency serves as a source of bias in our study.

**Section S5. Section S5. R Analysis Code**

**#NULL MODEL IN MAJORITY GIRLS DATASET**

Library(metafor)

null\_model <- rma(yi = gALL, #effect size

vi = vgALL, #variance

data =majgirls\_Dec2020,

method ="REML", #estimate tau2 using REML

test="knha") #use Knapp-Hartung t-tests

summary(null\_model)

#Random-Effects Model (k = 702; tau^2 estimator: REML)

#logLik deviance AIC BIC AICc

#-749.1928 1498.3855 1502.3855 1511.4906 1502.4027

#tau^2 (estimated amount of total heterogeneity): 0.2036 (SE = 0.0176)

#tau (square root of estimated tau^2 value): 0.4512

#I^2 (total heterogeneity / total variability): 68.58%

#H^2 (total variability / sampling variability): 3.18

#Test for Heterogeneity:

# Q(df = 701) = 2099.0937, p-val < .0001

#Model Results:

# estimate se tval pval ci.lb ci.ub

#0.2604 0.0254 10.2499 <.0001 0.2105 0.3103 \*\*\*

# ---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**#MODERATION ANALYSIS IN MAJORITY GIRLS DATASET WITHOUT COVARIATES**

no\_covar\_model <- rma(yi = gALL, #effect size

vi = vgALL, #variance

mods = ~ state\_sexismR, #moderator

data =majgirls\_Dec2020,

method ="REML", #estimate tau2 using REML

test="knha") #use Knapp-Hartung t-tests

summary(no\_covar\_model)

#Mixed-Effects Model (k = 702; tau^2 estimator: REML)

#logLik deviance AIC BIC AICc

#-744.4181 1488.8361 1494.8361 1508.4894 1494.8706

#tau^2 (estimated amount of residual heterogeneity): 0.2027 (SE = 0.0175)

#tau (square root of estimated tau^2 value): 0.4503

#I^2 (residual heterogeneity / unaccounted variability): 68.46%

#H^2 (unaccounted variability / sampling variability): 3.17

#R^2 (amount of heterogeneity accounted for): 0.43%

#Test for Residual Heterogeneity:

# QE(df = 700) = 2094.2294, p-val < .0001

#Test of Moderators (coefficient 2):

# F(df1 = 1, df2 = 700) = 6.7084, p-val = 0.0098

#Model Results:

# estimate se tval pval ci.lb ci.ub

# intrcpt 0.2370 0.0268 8.8338 <.0001 0.1843 0.2897 \*\*\*

# state\_sexismR -0.0742 0.0287 -2.5901 0.0098 -0.1305 -0.0180 \*\*

**#MODERATION ANALYSIS IN MAJORITY GIRLS DATASET WITH COVARIATES**

covar\_model <- rma(yi = gALL, #effect size

vi = vgALL, #variance

mods = ~ state\_sexismR + median\_inc\_Rscale,

data =majgirls\_Dec2020,

method ="REML", #estimate tau2 using REML

test="knha") #use Knapp-Hartung t-tests

summary(covar\_model)

# Mixed-Effects Model (k = 702; tau^2 estimator: REML)

#

# logLik deviance AIC BIC AICc

# -743.8199 1487.6398 1495.6398 1513.8384 1495.6975

#

# tau^2 (estimated amount of residual heterogeneity): 0.2032 (SE = 0.0176)

# tau (square root of estimated tau^2 value): 0.4508

# I^2 (residual heterogeneity / unaccounted variability): 68.49%

# H^2 (unaccounted variability / sampling variability): 3.17

# R^2 (amount of heterogeneity accounted for): 0.20%

#

# Test for Residual Heterogeneity:

# QE(df = 699) = 2091.1027, p-val < .0001

#

# Test of Moderators (coefficients 2:3):

# F(df1 = 2, df2 = 699) = 3.5202, p-val = 0.0301

#

# Model Results:

#

# estimate se tval pval ci.lb ci.ub

# intrcpt 0.2399 0.0273 8.7932 <.0001 0.1863 0.2934 \*\*\*

# state\_sexismR -0.0702 0.0295 -2.3778 0.0177 -0.1282 -0.0122 \*

# median\_inc\_Rscale 0.0156 0.0270 0.5778 0.5636 -0.0374 0.0686

#

# ---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**#MODERATION WITH RVE AND NESTING BY STUDY IN MAJORITY GIRLS DATASET**

library(clubSandwich)

coef\_test(covar\_model, #estimation model above

cluster=majgirls\_Dec2020$studyid, #nesting by study

vcov = "CR2") #type of correction, standard errors corrected using

# Coef. Estimate SE t-stat d.f. p-val (Satt) Sig.

# 1 intrcpt 0.2399 0.0449 5.348 21.7 <0.001 \*\*\*

# 2 state\_sexismR -0.0702 0.0530 -1.324 17.4 0.203

# 3 median\_inc\_Rscale 0.0156 0.0388 0.402 11.1 0.695

#NOTE: degrees of freedom significantly reduced in this model, compared #to non-nested model directly above

**#SENSITIVITY ANALYSIS: MODERATION ANALYSIS IN STUDIES WITH >75% GIRLS**

**#NO COVARIATES**

seventy5\_model <- rma(yi = gALL, #effect size

vi = vgALL, #variance

mods = ~ state\_sexismR, #moderator

data = majgirls\_Dec2020 %>% subset(male<=25),

method ="REML", #estimate tau2 w/REML

test="knha") #use Knapp-Hartung t-tests

summary(seventy5\_model)

# Mixed-Effects Model (k = 226; tau^2 estimator: REML)

#

# logLik deviance AIC BIC AICc

# -205.6230 411.2461 417.2461 427.4810 417.3552

#

# tau^2 (estimated amount of residual heterogeneity): 0.1454 (SE = 0.0278)

# tau (square root of estimated tau^2 value): 0.3813

# I^2 (residual heterogeneity / unaccounted variability): 54.42%

# H^2 (unaccounted variability / sampling variability): 2.19

# R^2 (amount of heterogeneity accounted for): 7.28%

#

# Test for Residual Heterogeneity:

# QE(df = 224) = 520.3106, p-val < .0001

#

# Test of Moderators (coefficient 2):

# F(df1 = 1, df2 = 224) = 13.1535, p-val = 0.0004

#

# Model Results:

#

# estimate se tval pval ci.lb ci.ub

# intrcpt 0.1488 0.0448 3.3202 0.0010 0.0605 0.2371 \*\*

# state\_sexismR -0.2371 0.0654 -3.6268 0.0004 -0.3659 -0.1083 \*\*\*

#

# ---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**#SENSITIVITY ANALYSIS: MODERATION ANALYSIS IN STUDIES WITH >75% GIRLS**

**#WITH COVARIATES**

covar75\_model <- rma(yi = gALL, #effect size

vi = vgALL, #variance

mods = ~ state\_sexismR + median\_inc\_Rscale, #moderator

data = majgirls\_Dec2020 %>% subset(male<=25), #data name

method ="REML", #estimate tau2 w/REML

test="knha") #use Knapp-Hartung t-tests

summary(covar75\_model)

# Mixed-Effects Model (k = 226; tau^2 estimator: REML)

#

# logLik deviance AIC BIC AICc

# -201.1774 402.3549 410.3549 423.9836 410.5384

#

# tau^2 (estimated amount of residual heterogeneity): 0.1323 (SE = 0.0264)

# tau (square root of estimated tau^2 value): 0.3637

# I^2 (residual heterogeneity / unaccounted variability): 52.02%

# H^2 (unaccounted variability / sampling variability): 2.08

# R^2 (amount of heterogeneity accounted for): 15.65%

#

# Test for Residual Heterogeneity:

# QE(df = 223) = 484.9734, p-val < .0001

#

# Test of Moderators (coefficients 2:3):

# F(df1 = 2, df2 = 223) = 10.6742, p-val < .0001

#

# Model Results:

#

# estimate se tval pval ci.lb ci.ub

#intrcpt 0.1261 0.0445 2.8317 0.0051 0.0384 0.2139 \*\*

#state\_sexismR -0.2336 0.0641 -3.6464 0.0003 -0.3599 -0.1074 \*\*\*

#median\_inc\_Rscale 0.1170 0.0416 2.8102 0.0054 0.0350 0.1991 \*\*

#

# ---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**#SPECIFICITY ANALYSIS: MODERATION ANALYSIS IN MAJORITY BOYS DATASET**

majboys\_main\_covar\_model <- rma(yi = gALL, #effect size

vi = vgALL, #variance

mods = ~ state\_sexismR + median\_inc\_Rscale,

data =majboys\_post, #data name

method ="REML", #estimate tau2 using REML

test="knha") #use Knapp-Hartung t-tests

summary(majboys\_main\_covar\_model)

# Mixed-Effects Model (k = 1996; tau^2 estimator: REML)

#

# logLik deviance AIC BIC AICc

# -1817.1932 3634.3865 3642.3865 3664.7761 3642.4066

#

# tau^2 (estimated amount of residual heterogeneity): 0.1691 (SE = 0.0084)

# tau (square root of estimated tau^2 value): 0.4113

# I^2 (residual heterogeneity / unaccounted variability): 71.41%

# H^2 (unaccounted variability / sampling variability): 3.50

# R^2 (amount of heterogeneity accounted for): 0.27%

#

# Test for Residual Heterogeneity:

# QE(df = 1993) = 6260.8441, p-val < .0001

#

# Test of Moderators (coefficients 2:3):

# F(df1 = 2, df2 = 1993) = 4.1143, p-val = 0.0165

#

# Model Results:

#

# estimate se tval pval ci.lb ci.ub

# intrcpt 0.2654 0.0147 18.0041 <.0001 0.2365 0.2943 \*\*\*

# state\_sexismR 0.0030 0.0178 0.1657 0.8684 -0.0320 0.0379

# median\_inc\_Rscale-0.0470 0.0230 -2.0412 0.0414 -0.0921 -0.0018 \*

#

# ---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**#COUNTY-LEVEL ANALYSIS**

county\_model <- rma(yi = gALL, #effect size

vi = vgALL, #variance

mods = ~ county\_sexismR,

data =majgirls\_Dec2020,

method ="REML", #estimate tau2 using REML

test="knha") #use Knapp-Hartung t-tests

summary(county\_model)

#Mixed-Effects Model (k = 702; tau^2 estimator: REML)

#logLik deviance AIC BIC AICc

#-747.3573 1494.7146 1500.7146 1514.3678 1500.7491

#tau^2 (estimated amount of residual heterogeneity): 0.2049 (SE = 0.0177)

#tau (square root of estimated tau^2 value): 0.4526

#I^2 (residual heterogeneity / unaccounted variability): 68.68%

#H^2 (unaccounted variability / sampling variability): 3.19

#R^2 (amount of heterogeneity accounted for): 0.00%

#Test for Residual Heterogeneity:

# QE(df = 700) = 2099.0913, p-val < .0001

#Test of Moderators (coefficient 2):

# F(df1 = 1, df2 = 700) = 2.1679, p-val = 0.1414

#Model Results:

# estimate se tval pval ci.lb ci.ub

#intrcpt 0.1879 0.0555 3.3876 0.0007 0.0790 0.2969 \*\*\*

#county\_sexismR -0.1031 0.0700 -1.4724 0.1414 -0.2405 0.0344

#---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**####SPATIAL AUTOCORRELATION ANALYSIS###**

library("USAboundaries")

library(sp)

library(sf)

library(spdep)

#load state boundaries in simple features (sf) format

states <- us\_states()

#convert to sp format

states <- as(states, "Spatial")

#limit to 31 states found in dataset

ourstates <- subset(states, state\_abbr %in% majgirls\_factor\_norms$STATE)

#convert to neighborhoods

ourstates\_nb <- poly2nb(ourstates)

#convert to spatial weights

ourstates\_lw <- nb2listw(ourstates\_nb, zero.policy = T)

#First, examine spatial autocorrelation of residuals by state

#rename ‘majgirls’ dataframe for this analysis

df <- majgirls\_state\_county

#obtain model residuals from primary moderation model

df$modelresid <- residuals.rma(norms\_model)

#create dataframe with mean of residuals by state

ourstate\_resid <- df %>% group\_by(state\_abbr = STATE) %>% #changed #variable name here to match for joining

summarize(mean\_resid = mean(modelresid))

ourstates@data <- left\_join(ourstates@data, ourstate\_resid)

#Moran’s I test

moran.test(ourstates$mean\_resid, ourstates\_lw, zero.policy = T)

#Moran I test under randomisation

#data: ourstates$mean\_resid

#weights: ourstates\_lw n reduced by no-neighbour observations

#Moran I statistic standard deviate = 0.005349, p-value = 0.4979

#alternative hypothesis: greater

#sample estimates:

# Moran I statistic Expectation Variance

# -0.03494229 -0.03571429 0.02082988

#Second, examine spatial autocorrelation of mean ES by State

#create dataframe with mean ES by state

ourstate\_ES <- df %>% group\_by(state\_abbr = STATE) %>% #changed variable name here so #they'd match for joining

summarize(mean\_ES = mean(gALL))#calc mean

ourstates@data <- left\_join(ourstates@data, ourstate\_ES)

#Moran’s I test

moran.test(ourstates$mean\_ES, ourstates\_lw, zero.policy = T)

#Moran I test under randomisation

#data: ourstates$mean\_ES

#weights: ourstates\_lw n reduced by no-neighbour observations

#Moran I statistic standard deviate = 0.22977, p-value = 0.4091

#alternative hypothesis: greater

#sample estimates:

# Moran I statistic Expectation Variance

#-0.003217345 -0.035714286 0.020003710

**##MAJORITY GIRL STUDIES TARGETING INTERNALIZING PROBLEMS ONLY**

tp\_model <- rma(yi = gALL, #effect size

vi = vgALL, #variance

mods = ~ state\_sexismR + tpR + median\_inc\_Rscale,

data =majgirls\_Dec2020,

method ="REML", #estimate tau2 using REML

test="knha") #use Knapp-Hartung t-tests

summary(tp\_model)

# Mixed-Effects Model (k = 702; tau^2 estimator: REML)

#

# logLik deviance AIC BIC AICc

# -743.4312 1486.8624 1496.8624 1519.6035 1496.9491

#

# tau^2 (estimated amount of residual heterogeneity): 0.2039 (SE = 0.0176)

# tau (square root of estimated tau^2 value): 0.4515

# I^2 (residual heterogeneity / unaccounted variability): 68.56%

# H^2 (unaccounted variability / sampling variability): 3.18

# R^2 (amount of heterogeneity accounted for): 0.00%

#

# Test for Residual Heterogeneity:

# QE(df = 698) = 2089.7265, p-val < .0001

#

# Test of Moderators (coefficients 2:4):

# F(df1 = 3, df2 = 698) = 2.3457, p-val = 0.0717

#

# Model Results:

# estimate se tval pval ci.lb ci.ub

# intrcpt 0.2415 0.0850 2.8421 0.0046 0.0747 0.4083 \*\*

#state\_sexismR -0.0702 0.0298 -2.3529 0.0189 -0.1288 -0.0116 \*

#tpRinternalizing -0.0017 0.0875 -0.0190 0.9848 -0.1734 0.1701

#median\_inc\_Rscale 0.0156 0.0270 0.5772 0.5640 -0.0374 0.0686

# ---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1