THE INSTITUTE FOR
RESEARCH & EVALUATION

6068 S Jordan Canal Rd
Salt Lake City, UT 84129
801-414-3443

A Brief Rebuttal to a Critique by the World Health Organization

October 17, 2023

In September 2018, at a side event of the UN Human Rights Council in Geneva, Switzerland, a research analyst from *The Institute for Research and Evaluation* presented the findings of their forthcoming review of the evidence for Comprehensive Sex Education (or CSE) titled, "Re-examining the Evidence for School-based Comprehensive Sex Education: A Global Research Review."[1] The Institute (IRE) had analyzed the international data cited by UNESCO[2] as evidence for its claims that CSE is effective at reducing adolescent sexual risk behavior, and the IRE findings contradicted these claims. Out of the 43 non-U.S. studies of school-based CSE in the UNESCO evidence base, IRE found only 3 that showed evidence of program effectiveness. The IRE definition of effectiveness was grounded in the scientific field of prevention research,[3] and was: that an effective program should cause a significant decrease in at least one key risk indicator for the targeted youth population (not just a subgroup), that the effect should last at least 12 months after the program's end (i.e., from one school year to the next), and that the program should not cause any negative effects on other measures of teen sexual health. In addition to finding little evidence of CSE success by this definition, the IRE reviewers found that 9 studies showed negative, harmful effects caused by school-based CSE. They concluded that when a credible scientific lens is used to evaluate school-based CSE, rather than the lenient standards employed in many favorable CSE reviews, there is little evidence of effectiveness and there appears to be more evidence of harm than real benefit.

Immediately following the presentation of these findings, an official from the World Health Organization (WHO) announced to the IRE presenter, "We disagree with your findings and will be actively working to refute them." In other words, without objectively examining the studies upon which the IRE findings were based, the WHO official decided, *a priori*, that the findings were false and should be rejected. Now, approximately five years later, the WHO has issued a critique of the IRE report, authored by personnel at WHO's *Department of Sexual and Reproductive Health and Research* (VanTreeck, et al., 2023).[4] The critique was published in a journal that is the publication arm of an advocacy organization which describes itself as a "community of researchers, activists and other experts" working "to shift ideology and power-driven politics… towards human rights and social justice …[with] explicit attention to sexual and reproductive justice." (See: https://www.srhm.org/about-us/) This lacks even the appearance of being a neutral, scientific publication.

The WHO critique declares the IRE report to be unscientific and full of errors, and labels it with biased terms like "misinformation research" and "a CSE opposition campaign." IRE has examined this WHO critique thoroughly, and has found, to the contrary, that *it* is full of errors and misinformation. A detailed rebuttal of this critique is being prepared for publication. Here are three key points:

1. **The WHO critique misrepresents the purpose and methods of the IRE review.**

a. A main purpose of the IRE review was to analyze a database previously identified by three authoritative scientific agencies in order to evaluate the evidence they claimed to show CSE effectiveness, rather than to conduct an original systematic review of CSE outcome studies. Yet the WHO reviewers criticized IRE for not conducting the latter type of review, even though IRE's purpose was stated in its report as "[an] examination of the best available sex education outcome research, as identified by three reputed scientific agencies…This allowed us to examine what other experts have independently identified as some of the best evidence for school-based CSE effectiveness."[1] The UNESCO[2] evidence base was the source for IRE's analysis of international

studies, which is the subject of the WHO critique. (For the U.S. analysis, IRE reviewed studies cited by the CDC and the HHS Teen Pregnancy Prevention Evidence Review.)

b. The WHO reviewers also criticized IRE for not specifying how it selected the individual studies for its review. However, as explained above, *IRE did not select the individual studies for its review*. The IRE review was an analysis of a pre-identified evidence base, a set of studies that had already been selected by three authoritative agencies as evidence for CSE effectiveness. These agencies and their documentation were specified in the IRE report. IRE's "re-examination" of the evidence was based on these agencies' selection of studies, not its own selection. A study's inclusion in the UNESCO database *was* the criterion for its inclusion in IRE's international database. Thus, this WHO criticism is not applicable.

c. The WHO reviewers seemed not to take note that the UNESCO database reviewed by IRE included the UNESCO 2018 version of its *International Technical Guidance on Sexuality Education*.[2] Both the 2009 and 2018 editions were listed in IRE's Endnote 29,[1] yet the WHO critique only references UNESCO's 2009 publication in its citations (see "Citation 2").[4] Their questioning of the IRE list of included studies and their stated difficulty identifying the study sources appears to refer solely to the 2009 publication, an important oversight by the WHO reviewers. In fact, IRE relied on the 2018 reference list because it was the most recent.[5] The 43 included studies are from this reference list and include the individual studies included in the systematic reviews cited in this list. (Many of the 2009 sources were included by default.) This misunderstanding on the part of the WHO reviewers could have been easily rectified if the reviewers had contacted the IRE analysts for clarification. The reason they gave for not doing so was explained as follows: "Given the polarised environment of CSE research, we did not reach out to the authors for additional information on their search strategy."[4] It is unfortunate that such a bias on the part of the WHO reviewers prevented this important clarification from occurring.

d. The WHO reviewers criticized IRE for not screening the included studies for scientific quality (rejecting those of lower quality) or assessing risk of study bias. Here again, the WHO critique misrepresents the IRE study, and in addition, employs a double standard. IRE did not conduct an assessment of study quality because, as noted in its report, IRE accepted whatever quality screening UNESCO employed in its criteria for included studies. However, as did the WHO reviewers, the IRE report commented on the unfortunate variation in the quality of studies included by UNESCO. For example, VanTreeck et al., found that "several of the studies had serious flaws [and] some low-quality studies had smaller sample sizes or were purely descriptive without employing robust statistical tests."[4] This illustrates the lack of high quality research evidence upon which UNESCO's positive claims about CSE are based. It should also be mentioned that the WHO critique praised several other systematic reviews of CSE research as strong evidence for CSE effectiveness, reviews which they said "underwent a peer-review process that verified their validity and rigour."[4] Among them is the Goldfarb and Lieberman review of 2021.[6] Unfortunately, authors of this study conducted no screening of scientific quality whatsoever for their 80 included sources and there was no assessment of "risk of bias." They actually acknowledge the "substantial number of studies with less rigorous designs, smaller samples, and/or more qualitatively based [i.e., subjective or non-experimental] approaches" (p.4) found in their evidence base. In fact, their citations include many sources that could not even be called studies, such as subjective write-ups by teachers about classroom discussions held with 15 to 20 students, a workshop in which testimonials were shared, and a subjective response to a musical performance. Moreover, at least one-half of the sources cited by Goldfarb and Lieberman did not test to see if program effects endured even past the end of the intervention. It should be asked whether this documented inclusion of inferior studies by both UNESCO and Goldfarb and Lieberman causes the WHO reviewers to question UNESCO's positive assertions about CSE or the validity of Goldfarb and Lieberman's enthusiastic claims about CSE's

wide-ranging benefits. Or perhaps VanTreeck, et al., will continue to endorse the low bar/double standard to which CSE programs have historically been held.

e. The WHO reviewers criticized IRE for including studies in its review (included because they were in the UNESCO database) that did not measure a 12-month post-program effect—one of the IRE criteria for effectiveness—and thus, of unfairly labeling these programs as ineffective. In fact, IRE did not ever identify these programs as ineffective, but only as lacking evidence of effectiveness—a crucial distinction. Indeed, one of the purposes of the IRE review was to report on the significant number of included UNESCO studies that did *not* measure long-term effects and therefore could not provide evidence of program effectiveness, as UNESCO claimed that they did.

f. The WHO reviewers criticized IRE for unfairly including these short-term studies in the denominator of its calculation of a CSE failure rate, which it expressly did not do, as stated in its report (the IRE methodology is stated on p. 167, under U.S. Findings[1]).

Because of such major misrepresentations of the purpose and methods of the IRE review, much of the criticism by the WHO reviewers is not valid. In large part, they criticized the IRE review based on faulty premises.

2. **The WHO critique has many factual errors.**

The WHO reviewers claimed that IRE's reporting of the results of the 43 studies contained errors regarding 74% of the studies. If true, this would seriously undermine the validity of the IRE findings. The WHO critique listed 66 instances[7] in which it found "discrepancies" between the IRE data table and the findings reported in the 43 cited studies. IRE analysts have examined each of these purported discrepancies, comparing them against the text and data tables published in each of the cited studies and the entries in the IRE data table. They found that the claims of the WHO were verified in only 9 of the 66 instances, while 11 of the cases were debatable disagreements on the interpretation of research findings. The large majority of these instances were minor issues that did not change the overall results or conclusions of the IRE analysis. (Most were entries of "not measured" rather than "non-significant," or vice versa.) The remaining 46 purported discrepancies were identified mistakenly by the WHO reviewers, based on their misinterpretation of study data or of the IRE data table. A number of these mistakes were such as would not be expected from someone in a research position at the WHO.

For example, the WHO reviewers:

- Labeled a significant increase in the number of teens becoming sexually active—which is a negative program outcome, and was labeled so by the study in question—as "a positive impact on sexual initiation" and as evidence of program effectiveness.[8]
- Failed to acknowledge statistical analyses in several studies which found that program effects were subgroup effects, consistent with how IRE reported them.[9]
- Mistook a data table reporting pre-test numbers for the study sample as a report of program effects measured at the follow-up survey.[10]
- In a misinterpretation of the IRE data table, repeatedly claimed that single program outcomes on the IRE data table were labeled as both positive and negative results, which was not ever done. Positive and negative outcomes were all labeled separately and clearly in the data table.
- Committed a number of other technical errors that will be detailed in the full IRE rebuttal.
- Counted the two times that a study author's name was slightly misspelled on the IRE data table (in other words, a "typo") as a misrepresentation of study findings by IRE.
- Based their review on an earlier version of the IRE report; several of the supposed errors they noted were not contained in the final published journal article (in *Issues in Law and Medicine*, 2019).[1]

The WHO critique claims there are 66 discrepancies out of the 430 data points in the IRE data table, which would be an error rate of 15%. However, only 9 of these were confirmed, leaving 9 out of 430, an actual error rate of 2%. This is less than the 5% that would statistically be expected to occur by chance. On the other hand, for the WHO analysis, there were 46 mistakes out of the 66 data points in their discrepancy table, which is an error rate of 70%.

3. <u>The results of the WHO data re-analysis report findings similar to the original IRE findings</u>.

Using the scientifically derived definition of program effectiveness employed by IRE, the WHO analysis of study findings still reported just 6 out of 43[11] international studies showing evidence of effectiveness for school-based CSE, only 3 more than IRE reported. (The WHO critique did not identify what specific studies these were, so we can only assume them to be the 6 studies listed in their Table B1 as showing "a positive effect,"[12] which is a completely different set of studies than the 3 identified by IRE.[13]) They also reported that 7 studies showed evidence of harmful impact, only two less that IRE reported. Correcting for the unarguable error by WHO reviewers in which they reported a negative result but mislabeled it as positive (Merakou, 2006),[8] their count of studies showing negative impact is 8, which compares to the 9 reported by IRE. These WHO numbers show little evidence of CSE effectiveness and an inverse ratio of program effectiveness to harmful impact, similar to the IRE results.

Although the WHO critique only claims 6 of these CSE programs have shown effectiveness (labeled "a positive effect" in their Table B1), the IRE analysts disagree with that designation. In each case, that claim is based on misinterpretation by the WHO reviewers of the study findings. For example, they called a subgroup effect an overall effect,[14] gave credit for a 12-month post-program effect where none was indicated,[15] and counted a program as effective that had produced multiple negative effects on program recipients.[16]

One thing the WHO reviewers have not made clear is that for all 6 of the CSE programs they claimed show effectiveness, the evidence did not come from independent studies. In each case, the evaluation study was conducted by the either the program's developer or by a researcher at the institution that developed or implemented the program. In other words, *the evaluation studies were not conducted by independent evaluators*.

Notwithstanding WHO's unsubstantiated claims to have discredited it, the lack of evidence for CSE effectiveness identified in the IRE review is confirmed by multiple recent reviews of CSE outcome research. A landmark meta-analysis of sex education effectiveness, sponsored by the U.S. Centers for Disease Control and Prevention (2012), found no evidence that school-based CSE programs significantly increased teen condom use or reduced teen pregnancy or STIs.[17] A more recent meta-analysis of 19 U.S. school-based CSE programs (2018) found "no consistent evidence" that school-based CSE programs significantly increased teen condom use or abstinence or reduced teen pregnancy.[18] A recent meta-analysis of 44 programs on the U.S. Teen Pregnancy Prevention list (2019) found no evidence that school-based CSE increased teen abstinence or condom use or reduced teen pregnancy or STIs.[19] And a recent research review claiming to show evidence of wide-ranging CSE benefits (Goldfarb & Lieberman, 2021)[6], did not hold up to an objective analysis that documented the lack of scientific rigor behind its purported evidence, finding that very few of the cited sources were studies of CSE programs, and most of those did not meet scientific standards for study quality.[20] Most recently, the U.S. Department of Health and Human Services 2023 update of the Teen Pregnancy Prevention Evidence Review did not identify any new school-based CSE programs that show long-term (12-month) effects for the target population, on any protective outcomes.[21]

<u>Summary</u>

The WHO critique of IRE's report on international CSE programs is full of errors and misinformation, of which the above three points are emblematic. It is also tainted by the appearance of bias on the part of its

authors and publisher and a lack of independent research studies supporting its conclusions. The IRE analysis of the WHO critique, including a re-analysis of the IRE data and the 43 cited studies, found a far higher rate of error in the WHO critique's reporting of data (70%) than WHO claimed to have found in IRE's reporting of data (15%). And none of the 9 IRE discrepancies that were verifiable (an actual error rate of 2%) had any effect on IRE's original findings. Ironically, the WHO analysis, despite its inaccuracies, reports findings similar to those of IRE: that in an international database selected and screened by UNESCO there were few school-based CSE programs showing real effectiveness and somewhat more showing negative effects. Thus, the WHO review was confirmatory; it underscores the shaky foundation upon which school-based CSE stands—the lack of evidence of real program effectiveness and the unacceptable number of negative CSE effects (about 1 in 5 programs)—even when calculated by reviewers with a favorable bias.

IRE stands by its original finding that when a credible scientific lens was used to evaluate international school-based CSE studies in the database identified by UNESCO as evidence for CSE success, only 3 out of 43 studies showed evidence of real effectiveness while 9 studies showed evidence of harmful impact. IRE concludes that there appears to be too little evidence of benefit and too much evidence of harm by school-based CSE programs in international settings.

A more detailed IRE rebuttal is forthcoming that answers the remaining criticisms of the WHO review and provides detailed documentation of the findings reported above.

Endnotes and Citations

1. Ericksen, I.H. and Weed, S.E. (2019). "Re-Examining the Evidence for School-based Comprehensive Sex Education: A Global Research Review." *Issues in Law and Medicine, 34(2):161-182.*
2. United Nations Educational, Scientific and Cultural Organization. (2018). *International Technical Guidance on Sexuality Education: An Evidence-Informed Approach*, Revised Edition. http://www.unaids.org/sites/default/files/media_asset/ITGSE_en.pdf; UNESCO. International Technical Guidance on Sexuality Education, Volume 1; 2009.
3. These standards or criteria for effectiveness are grounded in the work of the scientific field of prevention research. The criteria are: 1) the use of a reliable study designed to test cause and effect: an experimental or quasi-experimental design study with adequate sample size and reliable measures; 2) program results that show evidence of effectiveness: significant ($p<.05$) long-term protective effects (for school-based programs, lasting 12 months post-program), for the intended or target population of program recipients (not just a subgroup or subsample), on an important protective outcome (indicating a reduction in sexual risk behavior), without other negative program effects occurring on important outcomes. See the work of: Flay BR, Biglan A, Boruch RF, Castro FG, Gottfredson D. (2005). Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination. *Prev Sci*, 6(3):151–175; Gottredson DC, Cook TD, Gardner FEM, Gorman-Smith D, Howe GW, Sandler IN, Zafft KM. (2015). Standards of Evidence for Efficacy, Effectiveness, and Scale-up Research in Prevention Science: Next Generation. *Prev Sci*, 16(7):893-926. doi: 10.1007/s11121-015-0555-x; Blueprints for Healthy Youth Development: Blueprints Standards. Available at: https://www.blueprintsprograms.org/blueprints-standards/
4. VanTreeck K, Elnakib S, & Chandra-Mouli V. (2023) A reanalysis of the Institute for Research and Evaluation report that challenges non-US, school-based comprehensive sexuality education evidence base. *Sexual and Reproductive Health Matters*, 31:1, 2237791, DOI: 10.1080/26410397.2023.2237791
5. United Nations Educational, Scientific and Cultural Organization. (2018). *International Technical Guidance on Sexuality Education: An Evidence-Informed Approach*, Revised Edition. See p. 129, Appendix V. Studies referenced as part of the evidence review 2016. It states on this page that "Those [citations] marked with * were included in the analysis of systematic reviews and high-quality valuations." IRE included the studies marked with * and, where that study was a systematic review, also included the individual studies cited in that review, since they formed the basis for the systematic review's findings.
6. Goldfarb E and Lieberman L. (2021). Three Decades of Research: The Case for Comprehensive Sex Education. *J Adolesc Health*, 68(1):13-27. doi: 10.1016/j.jadohealth.2020.07.036
7. In Table B1, the WHO critique lists 59 entries of claiming discrepancies in the IRE data table (Table 7 in Endnote entry 1, above). Some of these entries note multiple discrepancies, giving a total of 66 itemized discrepancies or datapoints in the WHO critique.
8. Merakou K, Kourea-Kremastinou J. (2006). Peer education in HIV prevention: an evaluation in schools. *European Journal of Public Health*, Vol. 16, No. 2, 128–132.
9. James S, Reddy P, Ruiter R, McCauley A, van den Borne B. (2006). The impact of an HIV and AIDS life skills program on secondary school students in Kwazulu-Natal, South Africa. *AIDS Education and Prevention*, 18(4), 281–294; Mathews C,

Aarø LE, Grimsrud A, Flisher AJ, et al. (2012—listed as 2010 in IRE report). Effects of the SATZ teacher-led school HIV prevention programmes on adolescent sexual behaviour: cluster randomised controlled trials in three sub-Saharan African sites. *International Health*, (4) 111– 122, Site 3; Okonofua FE, Coplan P, Collins S, Oronsaye F, et al. (2003). Impact of an intervention to improve treatment-seeking behavior and prevent sexually transmitted diseases among Nigerian youths. *Int J Infect Dis*; 7: 61-73; Stanton BF, Li X, Kahihuata J, Fitzgerald A, et al. (1998). Increased protected sex and abstinence among Namibian youth following a HIV risk-reduction intervention: a randomized, longitudinal study. *AIDS* 1998, 12:2473–2480.

10. Walker D, Gutierrez JP, Torres P, Bertozzi SM. (2006). HIV prevention in Mexican schools: prospective randomized evaluation of intervention, BMJ, doi:10.1136/bmj.38796.457407.80
11. While the WHO critique states that it could not locate one of the 43 studies in the IRE database, it reported findings on this same study, which makes its database 43 studies as well.
12. Dente, M, Fabiani, M, Okwey, R, Conestà, N, et al. (2005). Impact of Voluntary Counselling and Testing and Health Education on HIV Prevention among Secondary School Students in Northern Uganda. *VCT AND HEALTH EDUCATION FOR HIV PREVENTION*; 3 (1) 1 – 11; Jemmott, J, Jemmott, L, Oleary, A, Ngwane, Z, et al. (2015). HIV/STI Risk-Reduction- Intervention Efficacy with South African Adolescents Over 54 Months. *Health Psychology*: 34(6): 610–621; Jewkes R, Nduna M, Levin J, Jama N, et al. (2008). Impact of Stepping Stones on incidence of HIV and HSV-2 and sexual behaviour in rural South Africa: cluster randomised controlled trial. *BMJ*; 337:a506 doi:10.1136/bmj.a506; Mathews C, Aarø LE, Grimsrud A, Flisher AJ, et al. (2012—listed as 2010 in IRE report). Effects of the SATZ teacher-led school HIV prevention programmes on adolescent sexual behaviour: cluster randomised controlled trials in three sub-Saharan African sites. *International Health*, (4) 111– 122, Site 3; Stanton BF, Li X, Kahihuata J, Fitzgerald A, et al. (1998). Increased protected sex and abstinence among Namibian youth following a HIV risk-reduction intervention: a randomized, longitudinal study. *AIDS* 1998, 12:2473–2480; Visser M. (2007). HIV/AIDS prevention through peer education and support in secondary schools in South Africa, SAHARA-J: *Journal of Social Aspects of HIV/AIDS*,4:3, 678-694, DOI: 10.1080/17290376.2007.9724891
13. Okonofua FE, Coplan P, Collins S, Oronsaye F, et al. (2003). Impact of an intervention to improve treatment-seeking behavior and prevent sexually transmitted diseases among Nigerian youths. *Int J Infect Dis*; 7: 61-73; Shuey DA, Babishangire BB, Omiat S, and Bagarukayo H. (1999). Increased sexual abstinence among in-school adolescents as a result of school health education in Soroti district, Uganda. *HEALTH EDUCATION RESEARCH*, 14 (3), 411–419; Stephenson J, Strange V, Allen E, Copas A, Johnson A, et al. (2008). The long-term effects of a peer-led sex education programme (RIPPLE): A cluster randomised trial in schools in England. *PLoS Med*, 5(11): e224. doi:10.1371/journal.
14. Mathews C, Aarø LE, Grimsrud A, Flisher AJ, et al. (2012—listed as 2010 in IRE report). Effects of the SATZ teacher-led school HIV prevention programmes on adolescent sexual behaviour: cluster randomised controlled trials in three sub-Saharan African sites. *International Health*, (4) 111– 122, Site 3.
15. Dente, M, Fabiani, M, Okwey, R, Conestà, N, et al. (2005). Impact of Voluntary Counselling and Testing and Health Education on HIV Prevention among Secondary School Students in Northern Uganda. *VCT AND HEALTH EDUCATION FOR HIV PREVENTION*; 3 (1) 1 – 11
16. Visser, 2007 Visser M. (2007). HIV/AIDS prevention through peer education andsupport in secondary schools in South Africa, SAHARA-J: *Journal of Social Aspects of HIV/AIDS*,4:3, 678-694, DOI: 10.1080/17290376.2007.9724891
17. Weed SE. Sex Education Programs for Schools Still in Question: A Commentary on Meta-Analysis. *Am J Prev Med.* 2012;42(3):313-315, doi: 10.1016/j.amepre.2011.11.004
18. Marseille E, et al. (2018) Effectiveness of school-based teen pregnancy prevention programs in the USA: a systematic review and meta-analysis, *Prevention Science*, 19(4):468–489.
19. Juras R, Tanner-Smith E, Kelsey M, Lipsey M, Layzer J. Adolescent Pregnancy Prevention: Meta-Analysis of Federally Funded Program Evaluations, *American Journal of Public Health*. 2019;*09*(4), e1-e8.
20. Ericksen IH and Weed SE. (2023). "Three Decades of Research:" A New Sex Ed Agenda and the Veneer of Science. *Issues in Law and Medicine*, 38(1):27-46.
21. Forrester e, Manzer J, Chesnut K, Knab J, et al. (2023). Updated Findings from the HHS Teen Pregnancy Prevention Evidence Review: October 2016-May 2022. U.S. Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation, April 2023. https://tppevidencereview.youth.gov/