



## **Rebuttal to a Critique by the World Health Organization**

October 15, 2024 (Revised)<sup>1</sup>

*The Institute for Research & Evaluation*

### **Summary**

Advocates of comprehensive sexuality education (CSE) claim it has been proven effective at reducing teenage sexual risk behavior. Yet a published review by the Institute for Research and Evaluation (IRE) found only 3 out of the 43 non-USA school-based CSE studies in the United Nations database showed evidence of CSE effectiveness when using scientifically credible criteria to measure effectiveness. Moreover, 9 studies showed CSE *increased* teenage sexual risk behavior. There was a similar pattern for CSE evidence in the USA. A critique of the IRE review by the World Health Organization (WHO) asserts that it is “poor science.” IRE’s analysis of this critique finds *it* is full of scientific errors, misrepresentations of the IRE review, and double standards for research rigor (e.g., there was a 56% error rate in the discrepancies WHO reviewers claim to have found in the IRE report). These problems with the WHO critique nullify its validity and make the label of “poor science” appear self-referent. Notably, WHO reviewers still found poor CSE results in the 43 UN studies, similar to IRE’s review. IRE stands by its original findings and observes that 6 recent systematic reviews of research confirm the lack of evidence of effectiveness for school-based CSE.

### **Background**

In September 2018, at a side event of the United Nations Human Rights Council in Geneva, Switzerland, a research analyst from The Institute for Research and Evaluation (IRE) presented findings from their forthcoming report titled, “Re-Examining the Evidence for School-based Comprehensive Sex Education: A Global Research Review.”<sup>2</sup> In addition to USA studies, IRE had analyzed the international (non-USA) studies cited by UNESCO<sup>3</sup> as evidence for its claim that comprehensive sexuality education (or CSE) is effective at reducing teenage sexual risk behavior. And the IRE findings did not support this claim. Of the 43 non-USA studies of school-based CSE in the UNESCO evidence base (studies designed to test program impact), IRE found only 3 that showed evidence of effectiveness. IRE used a definition of program effectiveness grounded in the scientific field of prevention research: a) an effective CSE program should make a significant improvement on at least one key protective risk indicator (sexual abstinence, condom use, pregnancy, or STIs); b) the effect should occur for the targeted youth population (not just a subgroup); c) the effect should last at least 12 months after the program’s end; and d) the program should not have negative impact on other sexual risk indicators.<sup>4</sup> The IRE reviewers found little evidence of CSE success in non-USA schools by this definition, and also found that 9 studies showed negative CSE impact—increases in teen sexual risk behavior, pregnancy, or STIs. Similar poor results were found in the USA studies. The IRE researchers concluded that when a credible scientific lens is used to evaluate school-based CSE, rather than the low standards employed in many favorable CSE reviews, there is little evidence of effectiveness and appears to be more evidence of harm than real benefit.

Immediately following the IRE presentation in Geneva, an official from the World Health Organization (WHO) announced, “We disagree with your findings and will be actively working to refute them.” Without examining the studies upon which they were based, the WHO official had decided that the IRE findings were false and should be rejected. Five years later, the WHO has issued a critique of IRE’s international (non-USA) CSE findings, authored by personnel at WHO’s Department of Sexual and Reproductive Health and Research (VanTreeck, et al., 2023).<sup>5</sup> Instead of being published in a neutral journal, the article appears in the outlet of an advocacy organization, *Sexual and Reproductive Health Matters*, 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/>)

### **Five Key Points of the IRE Rebuttal**

The WHO critique declares the IRE report to be unscientific and full of errors, and labels it with strong terms like “misinformation research” and “a CSE opposition campaign.” IRE has examined this critique, and has found, to the contrary, that *it* is full of errors and misinformation. Five key points of the IRE rebuttal are:

- **The WHO critique is full of scientific errors that negate the credibility of its claims.** For example, there is a 56% error rate in the WHO reviewers’ scientific interpretation of the 43 reviewed studies.
- **The WHO critique conveys misinformation; it misrepresents major aspects of the IRE review.** The WHO reviewers misrepresented the type of review IRE conducted and made many erroneous assertions about its research methodology, further undermining WHO’s credibility.
- **WHO employed a double standard in its critique of the IRE review.** WHO held IRE to a standard of research scrupulosity that it did not require of other reviews it cites as favorable to CSE.
- **WHO’s analysis of the 43 UN studies appears confirmatory of IRE’s original results.** Both IRE and WHO found that only a small minority of the CSE studies showed evidence of real effectiveness and about one in five showed evidence of harmful CSE impact (increased sexual risk behavior).
- **Six recent systematic research reviews show a lack of evidence of effectiveness for school-based CSE.** The WHO reviewers did not identify any recent school-based CSE studies showing 12-month effects for the target population on risk behaviors. Instead, they cited older and/or flawed research reviews to show positive CSE evidence. But this is not supported by current research.

Below are supporting details for each of these five rebuttal points.

#### 1. The WHO critique contains many scientific errors, negating its claims of IRE error.

The WHO reviewers assert that IRE’s reporting of the results of the 43 reviewed studies contains errors regarding 74% of the studies. This itself is erroneous. The WHO critique claims there are 66 instances<sup>5,6</sup> of these “discrepancies” between the findings reported in the IRE data table (Table 7) and the findings of the 43 studies IRE reviewed. The IRE data table contains 430 data points (43 studies x 10 potential findings for each study). If there *were* 66 discrepancies, this would be an error rate of 66 out of 430 data points, or 15%, not 74%. But there *are not* 66 discrepancies. IRE analysts have carefully examined each of these claims compared to the text and data from each of the relevant original studies as well as the corresponding entries in IRE’s Table 7. This analysis found that the claimed IRE discrepancies were verified in just 9 of the 66 instances, none of which changed the overall results or conclusions of the IRE analysis. (Most were entries of “not measured” rather than “non-significant,” or vice versa.) Twelve of the 66 supposed discrepancies were instead a reasonable decision by IRE on the interpretation of the study findings and 8 of the cases were a misunderstanding by WHO reviewers of IRE’s Table 7. None of these 20 instances were errors on the part of IRE. The remaining 37 purported IRE discrepancies were actually errors by the WHO reviewers, based on their mistaken interpretations of study data.

For example, the WHO reviewers:

- Incorrectly labeled a reduction caused by a CSE program in the percent of teens who had never had sex—which is a negative/harmful program effect—as a positive/desired program outcome and therefore “evidence of program potential.”<sup>7</sup>
- Failed to acknowledge statistical analyses in several studies which showed that program effects were not overall effects but actually subgroup effects, as IRE had reported them (details in Endnote 8). 2

- Mistook a data table reporting pre-test measures for the study sample at follow-up as a report of program effects measured at the follow-up survey.<sup>9</sup>
  - Failed to correctly identify the appropriate statistical indicator of a program effect, in several studies.<sup>10</sup>
  - 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 discrepancy in the study findings reported by IRE.
  - Based their review on a prior self-published version of the IRE report (not peer-reviewed);<sup>11</sup> several of the errors they noted do not appear in the final peer-reviewed journal publication.<sup>2</sup>
- (Full details on the 37 "discrepancies" by WHO—differences between their findings and study results—are documented at *institute-research.com*, under *Rebuttal to the WHO, IRE Review of Table B1.*)

In summary, the WHO critique claims there were 66 discrepancies out of 430 IRE data points. Only 9 of these were confirmed, which is an IRE error rate of 2% (9÷430). However, for the WHO, 37 of its 66 discrepancy claims were found to be mistaken interpretations of research, a 56% error rate (37÷66). These findings contradict the WHO's claim of a high IRE error rate and its accusation of "poor science."

2. The WHO misrepresents major aspects of the IRE review, conveying outright misinformation.

- a. The express intent of the IRE review was to analyze a database previously identified by three authoritative scientific agencies in order to evaluate the evidence those agencies claimed shows CSE effectiveness, *rather than conducting an original systematic review of CSE research*. IRE described the intent of its review 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."<sup>11</sup> (The UNESCO<sup>3</sup> evidence base was the source for the analysis of non-USA studies. For the USA analysis, the studies cited by the CDC and U.S. Teen Pregnancy Prevention Evidence Review were examined.<sup>2</sup>) *Ignoring this, the WHO reviewers mistakenly criticized IRE for not employing the methodology of a systematic research review*, for example, for not specifying IRE's criteria for selecting the reviewed studies. (As explained in its report, IRE *did not select* the reviewed studies. A study's inclusion in UNESCO's database *was* the criterion for inclusion in IRE's review of non-USA school-based CSE.)
- b. The WHO critique claims that IRE's "designation of indicators as key protective indicators vs. less protective indicators" was not explained, a clear misstatement of fact. The IRE report gave a detailed review (p. 6) of the scientific rationale for *consistent condom use* and *abstinence* as "key protective indicators" (*pregnancy* and *STIs* are self-evident). And Endnote 20 in the IRE report (cited on p. 5, where key indicators were first named in the text) gave a detailed rationale for the designation of "less-protective indicators," including *unprotected sex* as a "less-protective" outcome.<sup>11</sup>
- c. The WHO reviewers seemed not to perceive that the UNESCO database reviewed by IRE was based on UNESCO's 2018 version of its *International Technical Guidance on Sexuality Education*.<sup>3</sup> Both the 2009 and 2018 editions were listed as sources in IRE's Endnote 29,<sup>11</sup> yet the WHO critique only references UNESCO's 2009 publication in its citations (see "Citation 2").<sup>5</sup> WHO's questioning of the IRE list of included studies and their stated difficulty identifying the study sources appears to be because they were looking solely at the 2009 publication, an important oversight. In fact, IRE relied on the 2018 reference list because it was the most recent.<sup>2</sup> (Details in Endnote 12, this document.)
- d. The WHO reviewers criticized IRE for not screening the included studies for scientific quality (i.e., rejecting those of lower quality). Here again, the WHO critique misrepresents the IRE study. IRE did not conduct its own assessment of study quality because, as noted in its report, IRE accepted the quality screening that UNESCO 2018 employed as *its* criteria for included studies, in order to review the evidence endorsed by UNESCO.<sup>3</sup> However, as did the WHO reviewers, IRE's report mentions

the lower quality of some of the included studies.<sup>11</sup> For example, as stated by the WHO critique, “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.”<sup>5</sup> If this shows anything, *it is the lack of good quality research evidence upon which UNESCO’s positive claims about CSE are based.*

- 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 describe these programs as *ineffective*, but only as *lacking* “sufficient evidence of effectiveness”<sup>11</sup>—a crucial distinction. Including such studies enabled the IRE review to report on the substantial number of studies included by UNESCO that did *not* measure sustained effects and therefore could not provide evidence of program effectiveness, as UNESCO claimed that they did. 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 (this IRE methodology is stated on p. 167 of the journal article<sup>2</sup> and pp. 7 and 9 of the self-published report<sup>11</sup>).
- f. The WHO reviewers claim the IRE report “minimised positive intervention effects and privileged negative intervention effects.”<sup>5</sup> An express purpose of IRE’s review was to identify evidence of CSE effectiveness measured by credible scientific criteria, instead of the low standards used by most favorable CSE reviews. In these lax reviews, the existence of one minor positive outcome—a short-term or subgroup effect—is claimed as evidence of an effective program, even though there was no overall effect, no sustained effect, or no effect on the truly protective outcomes, and even if there was negative impact on other sexual risk outcomes. There is no scientific support for this definition of program “effectiveness” and it is meaningless in the real world. To give credit to such minimal positive effects and downplay the occurrence of negative effects would be to perpetuate the misrepresentation of the lack of evidence for CSE effectiveness. Furthermore, the attention IRE gave to negative program effects complies with the standards of The Society for Prevention Research.<sup>4</sup>

These major misrepresentations of the nature and methods of the IRE review compromise the validity of the WHO critique. In large part, it criticized the IRE review based on its own faulty premises.

### 3. The WHO employed a double standard in its critique of the IRE report.

The WHO reviewers cite 5 research reviews as counterpoints to the IRE review—studies whose “validity and rigour” they say has been “verified”<sup>5</sup> and that they claim show favorable CSE results. While the school-based CSE results in these studies are actually quite minimal, the studies themselves do not meet standards of research rigor that WHO required of the IRE report (although misapplied). While this is true for each of the 5 reviews, one example will illustrate. As noted above, IRE was criticized for relying on UNESCO’s screening for study quality and not conducting its own assessment. At the same time, the WHO reviewers gave high praise to a review of research by Goldfarb and Lieberman (2021)<sup>13</sup> for its “validity and rigour,” even though its authors conducted no screening whatsoever for the scientific quality of the 80 studies included in their review. Goldfarb and Lieberman actually acknowledged the “substantial number of studies with less rigorous designs, smaller samples, and/or more qualitatively based [i.e., subjective] approaches” (p.4) included in their evidence base. In fact, their citations include many sources that could not even be called studies, such as subjective write-ups by school teachers about classroom discussions with 15-20 students, a workshop in which testimonials were shared, and a subjective response to a musical performance. It should be asked why this documented problem of low-quality evidence did not cause the WHO reviewers to question the “validity” of Goldfarb and Lieberman’s review and its claims about CSE’s wide-ranging benefits. It appears the WHO holds a double standard for research rigor, giving a “pass” to favorable CSE reviews while demanding that the IRE review meet a different set of requirements.

#### 4. WHO's re-analysis of the CSE data actually reports results similar to IRE's original findings.

Using the scientifically derived definition of program effectiveness employed by IRE, the WHO analysis of study findings reported that just 6 out of the 43 international studies showed evidence of effectiveness for school-based CSE, only 3 more than IRE reported.<sup>14</sup> (The WHO critique did not identify which studies these were, so we can only assume them to be the 6 studies listed in their Table B1 as showing “a positive effect,”<sup>15</sup> which is a completely different set of studies than the 3 identified by IRE.<sup>16</sup>) They also stated that 7 studies showed harmful program impact (increases in teen risk behavior), only two less than IRE reported. Correcting for the inarguable error in which WHO reviewers correctly documented a negative program effect, but then mislabeled it as a positive/desired outcome,<sup>7</sup> they show 8 studies (19%) with negative effects, which compares to the 9 (21%) reported by IRE. Similar to IRE's results, WHO found little CSE effectiveness and the appearance of more evidence of CSE harm than effectiveness.

Although the WHO critique finds that only 6 of the 43 international school-based CSE studies have shown effectiveness, the IRE analysts disagree with that designation for those 6 programs. In each case, the designation of effectiveness is based on a misinterpretation by the WHO reviewers of the respective study's findings. For example, in one case, they called a subgroup effect an overall effect,<sup>17</sup> in another they gave credit for a 12-month post-program effect where none was indicated,<sup>18</sup> and in another case, counted a program as effective that had produced multiple negative effects on program participants.<sup>19</sup>

In addition, the WHO reviewers did not make clear that for all 6 of these CSE programs, the evidence they claim shows “a positive effect” did not come from an independent study. In each case, the study was conducted by either the program's author or by a researcher at the institution that developed or implemented the program. In other words, *the evaluation studies were not by independent evaluators.*

#### 5. Six recent reviews of research show a lack of evidence of effectiveness for CSE in schools.

The WHO critique praised several systematic reviews of CSE research as strong evidence for CSE effectiveness.<sup>5</sup> Upon close examination, however, these reviews are found to either be outdated,<sup>20</sup> of poor scientific quality,<sup>13,21</sup> contain few positive results for school-based CSE,<sup>13</sup> and/or suffer from the dual problem of using inadequate criteria for effectiveness (e.g., the finding of only short-term effects or subgroups effects) and discounting harmful program effects,<sup>22</sup> or all of the above. Moreover, roughly one-half of the studies included in these reviews were not independent evaluations; they were conducted by the intervention's developer or implementer, who has an investment in the findings.

On the other hand, the lack of evidence for school-based CSE effectiveness identified in the original IRE review is confirmed by the results of six recent systematic reviews of CSE research. A landmark meta-analysis sponsored by the U.S. Centers for Disease Control and Prevention found school-based CSE did not significantly increase teen condom use or “protection” or reduce teen pregnancy or STIs.<sup>23</sup> A 2018 meta-analysis of 19 USA school-based CSE programs found “no consistent evidence” that they significantly increased teen condom use or abstinence or reduced teen pregnancy.<sup>24</sup> A 2019 meta-analysis of 44 sex education programs on the U.S. Teen Pregnancy Prevention approved list found no evidence that school-based CSE had a significant effect on sexual risk indicators, including teen abstinence, condom use, pregnancy, or STIs.<sup>25</sup> The Goldfarb & Lieberman review (2021),<sup>13</sup> which claimed to show evidence of wide-ranging CSE benefits, did not hold up under an objective analysis which revealed that very few of its cited sources were studies of CSE, and those that were did not produce scientific evidence of CSE effectiveness.<sup>26</sup> And the 2023 update of the U.S. Teen Pregnancy Prevention Evidence Review could not identify any new studies (since 2016) of school-based CSE that showed 12-month effects for the targeted teenage population on any sexual risk indicators.<sup>27</sup>

Finally, a 2023 systematic review of the effects of CSE on youth, by Kim, et al., also failed to provide credible evidence for school-based CSE effectiveness at reducing sexual risk behavior.<sup>28</sup> The

review called itself a meta-analysis but it violated basic principles for valid meta-analysis. For example, the results of wildly disparate types of interventions—a text message program for gay teens on social media, a clinic program for adult sex workers, a social network program for adults who inject illegal drugs—were combined with the results of CSE programs for teens in school settings. Disparate *outcomes* were also combined—*cognitive* measures (knowledge, attitudes) were averaged together with measures of sexual *behavior*. And the internal inconsistency of the findings was extremely high (the  $I^2$  was 99%, well above the 56% to 75% range indicating “high” or “severe” heterogeneity).<sup>28,29</sup> Scholars of meta-analysis contend that such high heterogeneity in the interventions, outcome measures, and findings of a meta-analysis calls into question its validity.<sup>30</sup> Moreover, 88% of the included studies were rated low in quality, according to Kim, et al. For these reasons, we ignored this meta-analysis’ questionable results and instead examined each included study. Only 10 of the 34 studies tested the impact of school-based CSE programs on sexual risk behavior.<sup>31</sup> Of the 10 studies, one was of inadequate quality<sup>32</sup> and none of the rest demonstrated 12-month post-program effects on the targeted youth population for any of the protective behavioral outcomes (reduced sexual risk behavior, pregnancy, or STIs).<sup>33</sup>

It should also be mentioned that the WHO critique accused IRE of “actively working with UN member states to influence policymaking and decisions around funding allocation to CSE,” another misrepresentation of fact. While IRE *has* presented its research at UN Side Events and at conferences attended by UN member states (events sponsored by other entities, not IRE), IRE personnel *have neither had, nor sought*, any other communication with representatives from *any* UN member states about these topics. Again, a double standard is evident. No doubt, VanTreeck, et al., or other researchers from WHO’s Department of Sexual and Reproductive Health and Research, have actively engaged in many communications with UN member states to influence them in support of CSE, communication which the WHO would consider as totally appropriate.

### Conclusions

The WHO critique of IRE’s report on international CSE programs is full of misinformation and scientific errors, and employs a double standard regarding the assessment of scientific rigor. It also suffers from the appearance of bias on the part of its authors and publisher and the lack of independent research studies supporting its assertions. These problems with the WHO critique nullify its credibility. The IRE analysis of the WHO critique entailed a meticulous re-analysis of the IRE data table and the 43 cited studies. It found and documented a far higher rate of error in the WHO’s reporting of data (56%) than could be confirmed in IRE’s original reporting of data (2%). And none of the 9 IRE discrepancies that were verified had any effect on IRE’s original findings. Ironically, the WHO analysis, despite its inaccuracies, reported results similar to those of IRE: that in an international database vetted by UNESCO there were very few school-based CSE programs that showed credible evidence of effectiveness but a concerning number (about one in 5) that had harmful effects on teenage sexual behavior. In this respect, the WHO critique appears confirmatory of IRE’s original report. It confirms the shaky foundation upon which school-based CSE stands: the inadequate evidence of actual effectiveness and the uncomfortable number of harmful impacts, even when calculated by those favorably biased towards CSE. The WHO reviewers failed to present any recent scientifically credible positive evidence for CSE. Instead, they relied on outdated reviews and/or reviews that employ lax criteria for program effectiveness and/or low standards for study quality. In contrast, the results of six recent systematic reviews of CSE research confirm the lack of evidence of effectiveness for school-based CSE.<sup>23-28</sup>

IRE stands by its original finding that when a credible scientific lens was used to evaluate 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 (one in 5) 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. And the same can be said for CSE in USA schools.<sup>2</sup> The claim that CSE has been proven effective and worthy of widespread dissemination<sup>34</sup> requires credible scientific evidence of program effectiveness, and school-based CSE programs have not produced that evidence.<sup>35</sup>

## Endnotes and Citations

1. Earlier versions of this rebuttal have been published previously; they should be replaced by this revised and updated version.
2. 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.
3. United Nations Educational, Scientific, and Cultural Organization. (UNESCO). *International Technical Guidance on Sexuality Education: An Evidence-Informed Approach*, Revised Edition, 2018. [http://www.unaids.org/sites/default/files/media\\_asset/ITGSE\\_en.pdf](http://www.unaids.org/sites/default/files/media_asset/ITGSE_en.pdf); UNESCO. International Technical Guidance on Sexuality Education, Volume 1; 2009.
4. These standards or criteria for effectiveness are grounded in the work of the scientific field of prevention research, especially *The Society for Prevention Research* and *Blueprints for Healthy Youth Development*. 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 of sustained effects; 2) program results that show evidence of effectiveness: significant ( $p < .05$ ) sustained protective effects (for school-based programs, effects lasting 12 months post-program), for the intended or target population of program recipients (not just a subgroup or subsample), on one of the key outcomes (that actually protect from sexual harm), without other negative program effects occurring that increase sexual risk. 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; Gottfredson 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/>
5. VanTrececk 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
6. In Table B1, the WHO critique lists 59 entries of purported discrepancies in the IRE data table (Table 7 in Endnote entry 2, above). Seven of these 59 entries note two discrepancies, giving a total of 66 itemized discrepancies or datapoints in the WHO critique.
7. 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. (On p.131 the study states, "more students from the intervention group initiated sex." The significant effect cited by the WHO from Table 4 is a negative impact on teens, not a positive one. The number of virgin students decreased significantly in the intervention group,  $p < .001$ , but not in the control group,  $p < .064$ . WHO's claim of IRE error on this outcome is itself a very basic error in interpreting the study results.)
8. 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. (The "full implementation" group was a subgroup, since it was a product of "exploratory analyses" (p.287) that divided the sample into those who received "full" or "partial" implementation of the program. We took the "intent to treat" approach and considered these as subgroups, as the study also appeared to do, since they reported: "No effects were found on safe sex practices (condom use, sexual intercourse)" for the full sample.); 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. (When males and females were analyzed separately (p.117, Supplementary Tables 3 & 4), the effect was found only for males and not females. Thus, it was a subgroup effect); 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. (The outcome of interest is the relative change in the intervention group compared to the control group. This is given on p.66, where it states: "this statistically significant effect of the intervention was due to the reported increase among females (OR=1.80, 95%CI=1.11-2.92) rather than among males (OR=1.13, 95% CI=0.84-1.51) when comparing the intervention to control groups." Thus, it was a subgroup effect, occurring for females only); 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. (On p.2475, the study states: "This effect appears to have been contributed primarily by changes among the females (e.g. there was no statistically significant effect on abstinence among males who were virgins at baseline)." Thus, it was a subgroup effect—occurring for females but not males.)
9. 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
10. See, for example: Ajuwon A, Brieger W. (2007). Evaluation of a school-based Reproductive Health Education Program in rural South Western, Nigeria. *African Journal of Reproductive Health*; 11 (2), 47-59; Borgia P, Marinacci C, Schifano P, Perucci C. (2005). Is peer education the best approach for HIV prevention in schools? Findings from a randomized controlled trial. *JAdolescHealth*; 36, 508-516.
11. Weed SE, Ericksen IH. (2019). Re-Examining the Evidence for Comprehensive Sex Education in Schools: A Global Research Review, p.5. See: [https://www.institute-research.com/CSEReport/Global\\_CSE\\_Report\\_12-17-19.pdf](https://www.institute-research.com/CSEReport/Global_CSE_Report_12-17-19.pdf)
12. 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. With one exception, the 43 included studies are from this reference list. The exception was an update of one of the reviews on the list (Mason-Jones, 2016). It should be noted that if the older studies from the 2009 reference list were included in the IRE review, it would not increase how many studies show evidence of CSE effectiveness.
13. 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
14. 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, not 42.
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; 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

16. 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.
17. 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. (When males and females were analyzed separately (p.117, Supplementary Tables 3 & 4), the effect was found only for males and not females. Thus, it became clear that it was a subgroup effect.)
18. 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. (p.2, says it was “a post-test only control group study.” There was no indication that a long-term (12-month) post-program effect was measured.)
19. Visser, 2007 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
20. Grunseit A, Kippax S, et al. (1997). Sexuality Education and Young People's Sexual Behavior: A Review of Studies. *Journal of Adolescent Research*, Volume 12, Issue 4. <https://doi.org/10.1177/0743554897124002>; Kirby DB, Laris BA, Rollieri LA.(2007). Sex and HIV education programs: their impact on sexual behaviors of young people throughout the world. *J Adolesc Health Off Publ Soc Adolesc Med*, 40(3):206–217. doi:10.1016/j.jadohealth.2006.11.143; Underhill K, Operario D, Montgomery P. (2007). Systematic review of abstinence-plus HIV prevention programs in high income countries. *PLoS Med*, 4(9):e275. doi:10.1371/journal.pmed.0040275
21. Grunseit A, Kippax S, et al., (1997). Sexuality Education and Young People's Sexual Behavior: A Review of Studies *Journal of Adolescent Research*, Volume 12, Issue 4. <https://doi.org/10.1177/0743554897124002>. Note: This study appeared to conduct no screening for study quality and stated that its own results “often [were] compromised...because of inadequacies in study design, analytic techniques, outcome indicators, and reporting of statistics.”
22. Fonner VA, Armstrong KS, Kennedy CE, et al.(2014). School based sex education and HIV prevention in low- and middle income countries: a systematic review and meta-analysis. Vermund SH, ed. *PLoS ONE*, 9(3):e89692. doi:10.1371/journal.pone.0089692; Kirby DB, Laris BA, Rollieri LA.(2007). Sex and HIV education programs: their impact on sexual behaviors of young people throughout the world. *J Adolesc Health Off Publ Soc Adolesc Med*, 40(3):206–217. doi:10.1016/j.jadohealth.2006.11.143; Underhill K, Operario D, Montgomery P. (2007). Systematic review of abstinence-plus HIV prevention programs in high income countries. *PLoS Med*, 4(9):e275.
23. Weed SE. (2012). Sex Education Programs for Schools Still in Question: A Commentary on Meta-Analysis. *Am J Prev Med*;42(3):313-315, doi: 10.1016/j.amepre.2011.11.004
24. 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.
25. Juras R, Tanner-Smith E, Kelsey M, Lipsey M, Layzer J. (2019). Adolescent Pregnancy Prevention: Meta-Analysis of Federally Funded Program Evaluations, *American Journal of Public Health*;09(4), e1-e8.
26. 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.
27. 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/>
28. Kim, E.J.; Park, B.; Kim, S.K.; Park, M.J.; Lee, J.Y.; Jo, A.R.; Kim, M.J.; Shin, H.N. (2023). A Meta-Analysis of the Effects of Comprehensive Sexuality Education Programs on Children and Adolescents. *Healthcare*, 11, 2511. <https://doi.org/10.3390/healthcare11182511>
29. Higgins JPT, Thompson SG. (2002). Quantifying heterogeneity in a meta-analysis. *Stat Med*;21:1539–58.
30. Sharpe D. (1997). Of apples and oranges, file drawers and garbage: why validity issues in meta-analysis will not go away. *Clin Psychol Rev*;17(8):881–901.
31. An additional (11<sup>th</sup>) study used a cross-sectional correlation methodology to “estimate the association between [perceived] exposure to CSE” and sexual risk indicators. However, this type of study design is not able to test causal impact (see George, et al., p.274) and therefore, as a rule, is not included in reviews of evidence for CSE causal impact. (George, G.; Beckett, S.; Reddy, T.; Govender, K.; Cawood, C.; Khanyile, D.; Kharsany, A.B. (2022). Role of Schooling and Comprehensive Sexuality Education in Reducing HIV and Pregnancy Among Adolescents in South Africa. *J. Acquir. Immune Defic. Syndr.*, 90, 270.)
32. Mellanby, A.R.; Phelps, F.A.; Crichton, N.J.; Tripp, J.H. (1995). School sex education: An experimental programme with educational and medical benefit. *BMJ*, 311, 414–417. This study employed a questionable study design, using 3 different cohorts of program and matched comparison students to infer change over a 3-year time period, but each year was a different cohort of students, with the later cohorts having received more of the program content. There was no indication of the baseline differences between the program and comparison groups, for each cohort, on the outcome variables or the known confounding variables. This, along with the use of “local” and “distant” comparison groups, again, with no baseline measures of the outcomes for each cohort, makes the validity of the study’s results highly suspect.
33. Grossman, J.M.; Tracy, A.J.; Charamaraman, L.; Ceder, I.; Erkut, S. (2014). Protective effects of middle school comprehensive sex education with family involvement. *J. Sch. Health*, 84, 739–747, found a significant reduction by CSE in sexual initiation for the intervention group, but the follow-up period was less than 9 months post-program; Taylor, M.; Jinabhai, C.; Dlamini, S.; Sathiparsad, R.; Eggers, M.S.; De Vries, H. (2014). Effects of a teenage pregnancy prevention program in KwaZulu-Natal, South Africa. *Health Care Women Int.*, 35, 845–858, found a significant effect by CSE on condom use, but the follow-up period was only 8 months post-program and the intervention group’s condom use actually decreased. The other 7 school-based CSE programs either produced no significant behavioral effects (Kemigisha, 2019; Manaseri, 2019; Rotz, 2018) or only short-term behavioral effects, detected only at the end of the program or 3 months post-program (Coyle, 2021; Hegdahl, 2022; Menna, 2015; Yakubu, 2019). Short-term effects may indicate program potential, but an effective program needs to show impact that is sustained well beyond its end.
34. As claimed in: *Advocates for Youth*. (2009). Comprehensive Sex Education: Research and Results. *The Facts, September 2009*. Retrieved from <https://www.advocatesforyouth.org/wp-content/uploads/storage/advfy/documents/fscse.pdf>; United Nations Educational, Scientific, and Cultural Organization. (UNESCO). *International Technical Guidance on Sexuality Education: An Evidence-Informed Approach*, Revised Edition, 2018. [http://www.unaids.org/sites/default/files/media\\_asset/ITGSE\\_en.pdf](http://www.unaids.org/sites/default/files/media_asset/ITGSE_en.pdf), p.12.
35. It is worth noting that as of this date, the *Blueprints for Healthy Youth Development* website does not show any CSE programs (whether school-based or not) as qualifying for the label “Model Program,” a designation they require for a prevention program to be deemed “ready for widespread use” <https://www.blueprintsprograms.org/faq/>.