WHO says screens are bad? A critical review and meta-analysis refute the World Health Organization's guidelines on screen time

Yaakov Ophir1,2\*

Refael Tikochinski1

Hananel Rosenberg3

1The Hebrew University of Jerusalem, Jerusalem, Israel

2Technion, Israel Institute of Technology

3Ariel University

\*Corresponding author:

Yaakov Ophir, Seymour Fox School of Education, Hebrew University of Jerusalem, Jerusalem, Israel, 91905.

E-mail: yaakov.ophir@mail.huji.ac.il.

# Introduction

In April 2019, the World Health Organization (WHO) published comprehensive ‘guidelines on physical activity, sedentary behavior and sleep for children under 5 years of age’ (WHO, ‎2019). The guidelines were shaped with the full consensus of the Guideline Development Group (GDG) after a thorough scientific review of the available research. The rationale behind the guidelines was to provide policymakers, pediatricians and family nurses with practical recommendations on the amount of time that children should be: (1) physically active, (2) sleeping, and (3) spending on sedentaryactivities. Strikingly, this last activity included watching screen-based entertainment. As shown in Figure 1, the guidelines leave no room for doubt: “screen time is not recommended” for infants and for 1-year-old children. For older children aged 2–4 years, “screen time should be no more than 1 hour” and “less is better”. Inevitably, these official recommendations drew considerable public attention. Despite the richness of the complete document of guidelines, dozens of newspaper headlines turned the spotlight on the “[WHO guidelines on screen time](https://nypost.com/2019/04/24/who-releases-guidelines-on-screen-time-for-children/)”, warning against “[the big problem with the new screens](https://www.glamour.com/story/who-screen-time-guidelines-for-kids)”, and calling parents to [ban screen time for babies](https://www.bbc.com/news/health-48021224) and to [limit screen use for toddlers](https://www.euronews.com/2019/04/24/children-and-screens-who-tells-parents-to-ban-use-for-babies-and-limit-it-for-toddlers). Unfortunately, these headlines were essentially fake news.

Figure 1. A screen shot of the WHO recommendations for sedentary time



From a position of great respect toward the systematic work of the GDG, in the following article we argue that the scientific basis for the WHO guidelines on screen time is unstable. The structure of the current article follows the rules of the popular children’s game of Jenga[[1]](#footnote-1). In each step of the article, we will carefully remove one block of assumptions; by the end of the article, the entire tower of guidelines and headlines will seem extremely unsteady. Using a meta-analysis of the same empirical data that were available to the GDG, we will demonstrate that even if all (in our view, weak) assumptions exist, the overall effect of screens on the psychological development of children is negligible. Moreover, we will provide evidence that the field suffers from publication bias favoring the publication of studies that echo the public moral panic regarding screens.

# Block 1: The binding of ‘sedentary behaviors’ with ‘sedentary screen time’ is misleading

The overall framework of the WHO guidelines, as described in its opening statements, is to promote healthy sleeping and physical behaviors and to reduce physical inactivity – a noteworthy risk factor for mortality and obesity. The term that was chosen to describe this problematic physical inactivity is ‘sedentary activities’. According to the WHO glossary of terms (p. Ⅳ), sedentary activities include two types of activities: ‘sedentary screen time’, which is the “time spent passively watching screen-based entertainment (TV, computer, mobile devices)” and other ‘sedentary behaviors’ that are characterized by low levels of metabolic activity in a “sitting, reclining or lying posture”. The examples given for sedentary behaviors (without screens) include the time spent in a car seat or a stroller but also the time spent listening to a story. The equivalence between listening to a story and sitting restrained in a car seat is understood, assuming the overall framework of physical inactivity being linked to the problem of childhood obesity: Too much sitting without enough physical movement can lead to obesity, no matter the content of the sedentary activity. However, the above-mentioned news headlines did not focus on the simple act of sitting without movement or better yet, sitting listening to a story. The headlines were also not about obesity and physical health. Not surprisingly, the headlines gave voice to a narrow aspect of the guidelines, warning against the psychosocial negative consequences of sedentary screen time.

To some extent, the news reporters were right. The binding of sedentary behaviors with screen time in the recommendations (Figure 1) is not trivial, especially when viewed from the general perspective of the guidelines, which singles out inactivity and obesity. Consider a familiar scenario in which a young child cries in his/her car seat during a long drive. According to the guidelines, is it alright for parents to play him/her children’s songs on YouTube? Ostensibly, the child is already sitting in a state of a sedentary behavior and the additional screen time would not add more risk for obesity. In other words, assuming that children are physically active and that their sleeping habits are decent, can we expose them to screens? To answer this question, we should disregard the general framework of the guidelines and dive into the psychological outcomes of screen time.

The ‘summary of evidence’ section of the WHO guidelines (page 8) implies that screen time is a possible risk factor for both the physical and psychological development of children: the evidence on the negative physical effects of screens addresses adiposity and motor development, and the evidence on the psychological effects address children’s cognitive development and psychosocial health (page 8). Yet, in order to examine the unique consequences of screens beyond other forms of inactivity or sedentary behaviors, such as listening to a story, special attention should be given to the alleged cognitive and psychosocial implications. This is because there is no reason to assume that sedentary time with screens poses a greater risk for adiposity and motor development than sedentary time without screens (i.e., it is the physical inactivity that leads to adiposity, not the content of a TV show).

# Block 2: The quality of evidence for negative psychological outcomes of screens is very poor

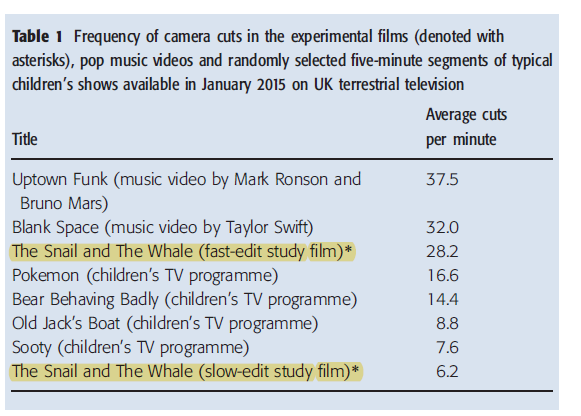
According to the authors of the WHO guidelines, the guidelines were formed based on a large number of empirical studies from high-quality systematic reviews published between 2017 and 2018 in six WHO languages. The scientific justification for the guidelines was provided in separate supplementary material named: [Web Annex Evidence Profiles](https://apps.who.int/iris/handle/10665/311663). The Web Annex includes 28 tables, which summarize findings from a total of 251 references. However, only two tables address the empirical studies examining the psychosocial or cognitive outcomes of screen time. The evidence for the psychosocial consequences of screens (Web Annex Table 1.2.3) is based on 20 studies: 2 randomized trials, 11 longitudinal studies, and 7 cross-sectional studies. The evidence for the cognitive consequences (Web Annex Table 1.2.4) is based on 29 studies: 12 longitudinal studies, 1 case-control, and 16 cross-sectional studies. In light of the fact that some studies appear in both tables (i.e., they examined both cognitive and psychosocial associations) and the fact that 4 studies in Web Annex Table 1.2.4 (studies #87, #160, #173, and #174) did not examine screen time but other sedentary behaviors, the final number of relevant studies is 33.

The quality of all the studies in the Web Annex was rated by the GDG using a common approach for the Grading of Recommendations, Assessment, Development and Evaluations (GRADE; Guyatt et al., 2011). Remarkably, the first severe criticism of the validity of the guidelines has been generated by the GDG themselves, who summarized the recommendation section on sedentary time with the phrase: “strong recommendations, very low-quality evidence” (Figure 1). Overall, 31 of 33 studies received a GRADE score of a “very low quality”. This troubling information removes a first fundamental ‘Jenga block’ from the scientific construction of the warnings against negative psychological outcomes of screens. Moreover, even the other two “moderate quality” studies are troublesome, as will be shown in the following two sections.

Block 3: One of the two “moderate quality” studies is irrelevant to the recommendations

The GDG judged 2 out of the 33 studies to have moderate scientific quality; however, one of these studies, Study #157, is irrelevant to the final recommendations. In Study #157, Kostyrka-Allchorne and colleagues (2017) investigated the effects of fast versus slow videos (measured in the frequency of camera cuts per minute) on preschool children. They created two versions of the ‘Snail and the Whale’ children’s story, one that simulated fast videos (28.2 cuts per minute) and one that simulated slower videos (6.2 cuts). The results of this experiment suggested that children who are exposed to fast‐paced videos may switch toys more frequently than children who watch slower paced videos.

This is an intriguing experiment; however, two characteristics of this study prevent it from being relevant to the WHO recommendations on screen time. First, this study does not address or examine sedentary screen time per se, because bothgroups participated in sedentary screen time behaviors. Second, the artificially enhanced video in the experimental group does not simulate typical children’s movies. As Kostyrka-Allchorne and collegues illustrated in their article (Figure 2), the frequency of the cuts in the fast video is comparable to extremely energetic videos that do not address, and are not appropriate for, preschool children. The samples provided by the authors are the music videos by Taylor Swift (‘Blank Space,’ 32 cuts per minute) and Mark Ronson (‘Uptown Funk,’ 37.5 cuts per minute). On the contrary, the slow video actually resembles a popular children’s TV program (‘Sooty,’ 7.6 cuts per minute). Therefore, an alternative and more accurate recommendation for parents from this study may be to control the content of the screens. In other words, the study suggests a problem with fast-moving content that is inappropriate for young children; it does not suggest a problem with sedentary screen time.

Figure 2. Illustration of fast and slow videos (Kostyrka-Allchorne et al., 2017) 

# Block 4: The other “moderate quality” study has significant methodological flaws

Allegedly, Study #101 (Yilmaz et al., 2015) that was also judged to be of moderate quality provides the holy grail. In Study #101, Yilmaz and colleagues conducted a randomized controlled trial (RCT) to assess the outcome of a designated intervention for reducing screen time among preschool children. Unfortunately, even this RCT suffers from significant methodological weaknesses that challenge its fit to the WHO guidelines. We outline four comments below in regards to this research. First, the main finding of the study, in which the authors show a significant reduction in screen time among children in the intervention group, is irrelevant to the WHO guidelines. Children in the intervention group started the research with a screen time average of 86.03 (SD = 20.46) minutes per day and completed it after 9 months with only 21.15 (SD = 6.12) minutes, but these numbers were essentially a straightforward manipulation check that ensured that the designated intervention actually worked. The only implication of this finding is that reducing screen time is possible. Although some parents may be encouraged by this finding, it has nothing to do with the outcome of the screen time.

Second, the body mass index (BMI) scores did not differ between children in the intervention and control groups. This non-finding is especially interesting in light of the scientific strength of the examined variable; BMI is an objective assessment that is not susceptible to subjective impressions or to research biases. Considering the experimental design of the study (RCT) and the validity of the examined variable, this non-finding actually challenges the general framework of the guidelines that focus on physical outcomes of sedentary activities.

Third, the analysis of psychosocial outcomes was not conducted correctly. Yilmaz and colleagues claimed that children in the intervention group displayed less aggressive and delinquent behavior compared with children in the control group. They support this claim on the fact that the two negative behaviors were not statistically different between the groups before the intervention and that fact that they were statistically different after the intervention. After 9 months from the intervention, the average aggressive score in the control group was 3.85 (SD = 1.38) compared with 3.35 (SD = 1.46) in the intervention group. The average delinquent behaviors score in the control group was 3.83 (SD = 0.95) compared with 3.45 (SD = 1.56) in the intervention group; however, these analyses were based on between-group, Student’s *t*-tests for independent samples, which are not appropriate for this type of experimental design and therefore should be cautiously interpreted.

Well-designed RCTs that include two groups (experimental and control) as well as repeated measures (before and after the manipulation/intervention) should incorporate mixed-design analyses (a split-plot analysis of variance) that address the differences between the independent groups, while considering the changes that occur within each group. This is because our main interest in RCTs was in the difference between the changes that occurred in the intervention group and the changes that occurred in the control group. This mixed analysis was especially required in the research at hand in light of its peculiar findings. On the one hand, the average scores of delinquent behaviors increased, both in the control group (3.02 → 3.83) and in the intervention group (3.02 → 3.45), despite the reduction in screen time, which should have led to less delinquent behavior. On the other hand, there was a sizable decrease in aggressive behaviorin both groups. As expected, the scores in the intervention group decreased from an average of 6.94 (SD = 1.66) to 3.35 (SD = 1.46) but unexpectedly, a similar trend also occurred in the control group. The average aggressive behavior scores in the control group decreased from 7.17 (SD = 1.52) to 3.85 (SD = 1.38). This reduction in both groups (3.32 and 3.59) seems significantly larger that the difference between the changes (0.27) that occurred in each group, 9 months after the intervention. The large reduction in both groups may actually imply that the principle cause for the reduction in aggressive behavior had nothing to do with screen time. In order to conduct the essential mixed analyses, we requested the original data file from Prof. Yilmaz; at the time of this writing, we have not receive her response.

Fourth, even if the appropriate analyses were applied, the specific methodology of the experiment did not provide conclusions regarding the harmful effects of screens. Whereas families allocated to the intervention group were exposed to four sets of materials and counseling sessions in which they were taught the “harmful effects of TV, video, and computer games”, families in the control group “were not aware of (the) counselling interventions” (p. 444). As the authors themselves acknowledge, unlike parents in the control group, “the parents’ reports of aggressive behaviors (in the intervention group) are likely to be biased” (p. 448) as they were fully aware of the research goals. These parents were exposed to the demand characteristics of the study; they understood what was expected from them and knew what the “right” answers were. In contrast with the objective BMI scores, parents could provide biased responses to the subjective psychosocial questionnaires. This bias might have even occurred without the parents’ awareness. Moreover, the participants in the intervention group were subjected to a relatively strong placebo effect; they went through a persuasion campaign that emphasized the harmful effects of screens and the beneficial effects of the active reduction of screen time. Therefore, even if the parents’ reports are genuine, the design of the research does not allow the differentiation between the placebo effect and the active ingredient of reducing screen time.

# Block 5: The 33 studies on psychological outcomes do not converge into a consistent pattern

Having established that the WHO guidelines are based on a very poor-quality research, one might claim that, together, the 33 studies form a critical mass that tells a consistent troubling story. Regrettably, this is also not true. In order to extract an overall impression from the entire set of studies, we revisited each one of the 33 articles assembled by the GDG (Tables 1.2.3 and 1.2.4 in the Web Annex) and mapped the various findings to ‘good’, ‘bad’, or ‘null’ effects of screens. Altogether, including the two experimental yet problematic studies from above (Blocks 3 and 4), we counted a total of 66 psychological effects of screens.

The first impression that arises from this aggregation is that the 33 studies examine a very wide range of psychosocial and cognitive variables. Speech disorders, classroom behaviors, attention difficulties, and victimization of bullying are only few examples. Therefore, the claim that a consistent pattern can be seen from the 33 studies relies on a problematic ‘logical jump’ – that all of the studies address similar psychological effects. The assumption that the studies can be aggregated into one unified effect of screen time is, in our opinion, a far-reaching logical step. The second impression from this aggregation is that most of the studies (*n* = 21) also reported null results in some of the variables that were examined. Taking into account the valid suspicion of publication bias, i.e. the scientific tendency not to publish non-results (see Block 7), it is possible that additional studies with null results ended up not being published. The third and somewhat surprising impression is that a few studies (*n* = 5) have actually documented favorable associations with screen use. These favorable associations mean that screen time has also been documented to be associated with positive psychological outcomes, such as improved social skills and cognitive functions. Taking together these three impressions, we conclude that the 33 studies do not share a common and clear (negative) scientific ground.

# Block 6: Meta-analyses reveal negligible summary effects of screens

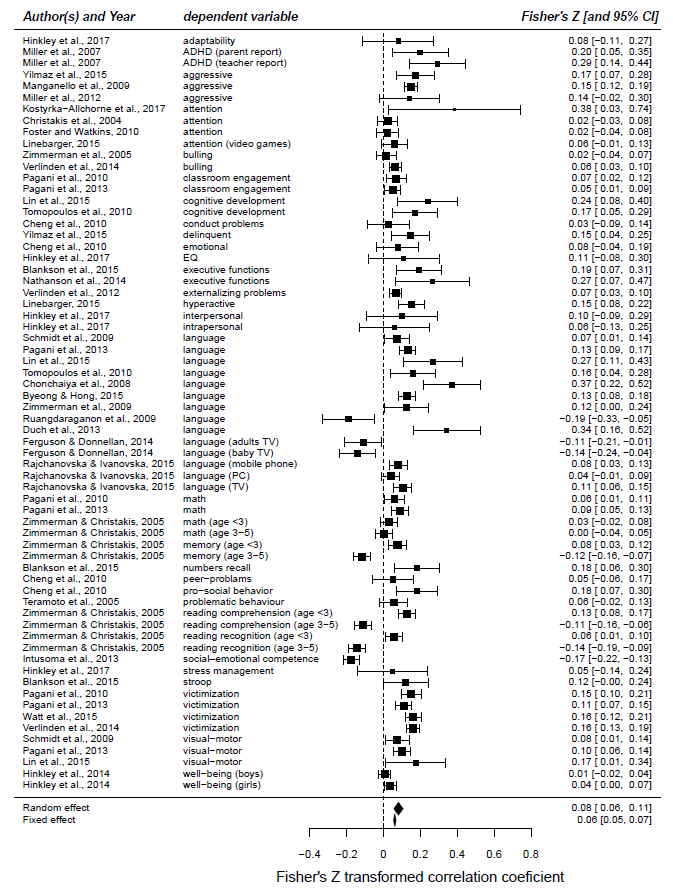
Even if we take the far-reaching logical step of aggregating the 66 effects from all 33 studies into one unified conclusion regarding the psychosocial effects of screens, an objective–mathematical method can be applied to estimate the overall effect. To examine the overall effect of screens, we conducted a meta-analysis that combines the 66 effects from all 33 studies and extracts an overall estimate for the actual effect of screens (Littell et al., 2008). For the purpose of this meta-analysis, we accepted the inclusion of all 33 studies’ data that were available to the GDG, despite our reservations about some of these studies and about the general aggregation assumption. For the application of the meta-analysis, we used R version 3.5.0 ([www.r-project.org](http://www.r-project.org)).

Prior to the meta-analysis, we collected all of the relevant effects on psychosocial and cognitive variables from all 33 studies and transformed them into a unified parameter of Pearson’s correlation coefficient (*r*). We used the R package ‘Psych’ (version 1.8.12; <https://personality-project.org/r/psych>) to transform the effect sizes from their original modes (e.g., Cohen’s D, odds ratio) into Pearson’s *r*. In cases where the authors did not provide a value for the effect size, we computed the Pearson’s *r* from the reported means and standard deviations and from the given test statistics (i.e., *T*, *F* or *χ2*). In cases where multiple regression was conducted, we transformed the standardized beta coefficients to Pearson’s *r* using a conventional formula: *r* = *β* + 0.5γ, whereby γ equals 1 when *β* is positive and zero when *β* is negative (Peterson and Brown, 2005). Only one study lacked standardized beta coefficients or other interpretable statistics (Mistry et al., 2007) and therefore could not be included in the meta-analysis.

After the collection and transformation of all the effect sizes, we standardized the total of 66 Pearson’s coefficients into a set of Fischer *Z* scores and conducted the meta-analysis, using the statistical package, Metaphor (version 2.1-0; [www.metafor-project.org](http://www.metafor-project.org)). To ensure the validity of the results, we applied both conventional statistical models: the random effects model and the fixed model. The random effects model is considered more lenient because it allows the assumption that studies are not functionally identical (Borenstein et al., 2011).

The image that arises from the entire meta-analysis is illustrated in a forest plot (Figure 3). Allegedly, the results indicate that screen time is positively related with psychosocial and cognitive variables (*r* = 0.08, *p* < 0.01; 95% CI [0.06, 0 .11]). The two additional meta-analyses that were performed separately for each cluster of variables provided similar results, for the psychosocial (*r* = 0.08, *p* < 0.01; 95% CI [0.05, 0 .11]) and cognitive variables (*r* = 0.08, *p* < 0.01; 95% CI [0.05, 0 .12]); however, although significant, the summary effect value (*r* = 0.08) and the confidence intervals that approached zero (0.06) imply that screen time has a very low effect on a child’s psychological development (Cohen, 2013). De facto, the summary effect value (0.08) explains only 1.6% of the variance in the child’s psychosocial development. Thus, even if we accept all of the assumptions made in the WHO guidelines, the actual effect of screens is negligible.

Figure 3. Forest plot of the meta-analysis

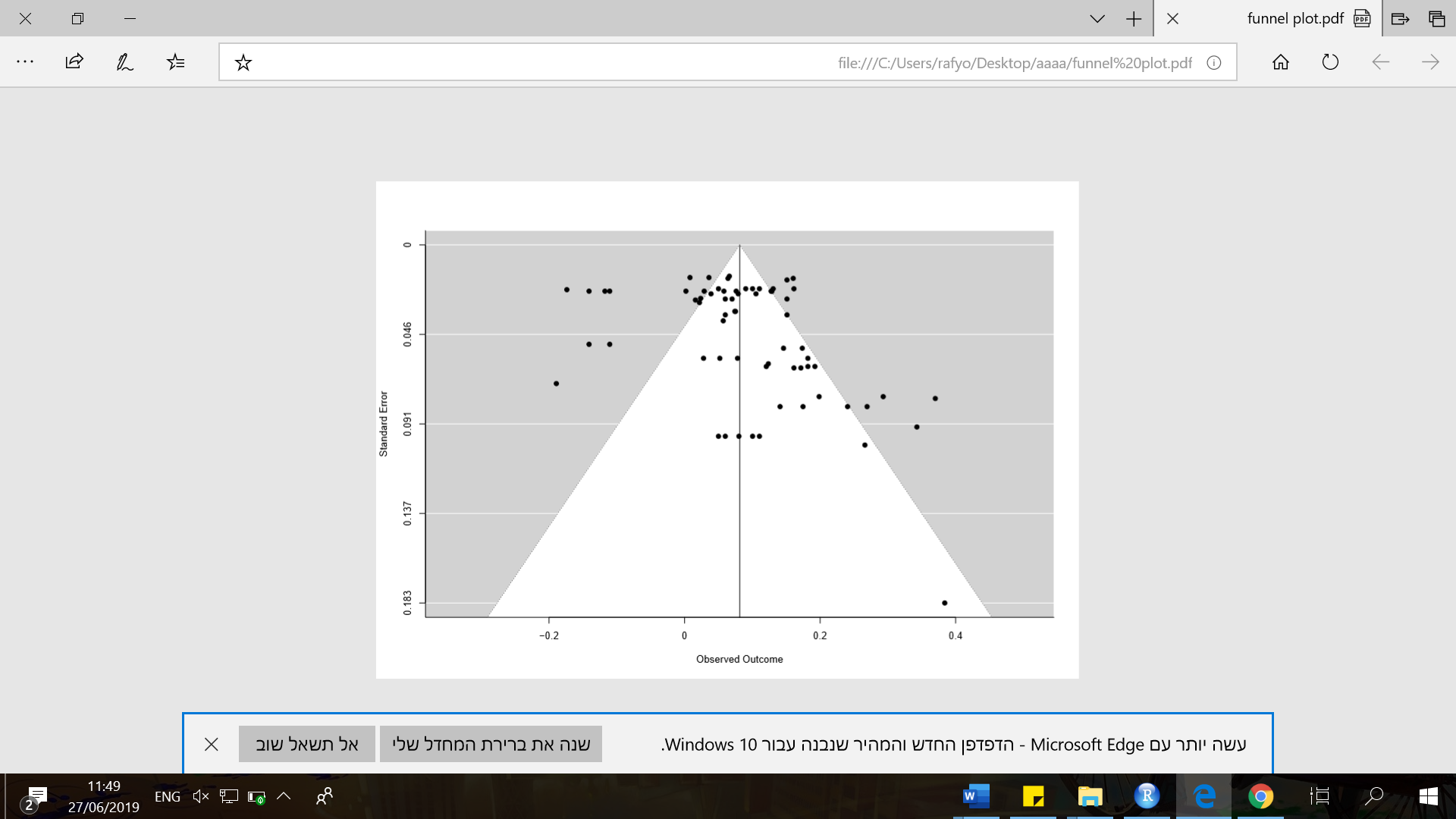


# Block 7: Meta-analysis shows a significant publication bias toward favoring negative effects

As implied in Block 5, one of the inherent problems in the publication process of scientific research is the tendency of both editors and authors to favor studies that yield significant findings (i.e., that reject the null hypothesis). Empirical studies that do not result in statistically significant effects are usually not even considered for publication, unless they are based on a particularly large data set or on a very strong experimental design. Poor-quality cross-sectional studies with small to medium samples, such as the ones used to form the guidelines, are therefore especially vulnerable to this bias. Those poor-quality studies that are ‘lucky’ enough to find an effect are likely to be published, whereas those that do not find significant effects would not.

If a publication bias exists in a given field, then it would skew the entire meta-analysis (Borenstein et al., 2011). To estimate whether the collected studies underlying the meta-analysis suffer from a publication bias, we conducted a funnel plot. The funnel plot method relies on the simple notion that if a phenomenon exists in reality (e.g., that screen time has a negative effect on psychological development) then the studies that try to measure this phenomenon should show a relatively symmetric pattern. If, for example, the ‘real’ correlation between screen time and psychological variables is around 0.3, then some studies would have lower values (e.g., 0.1, 0.2) and others would have higher values (e.g., 0.4, 0.5), but they all should form a somewhat symmetric pattern that revolves around the center point. Worryingly, the funnel plot of all the 66 effects included in the meta-analysis showed a severely asymmetric pattern (Figure 4). Moreover, the Egger’s regression test for funnel plot asymmetry showed a significant effect of publication bias (*Z* = 3.656, *p* < 0.001). The clustering of effect sizes plotted in the lower part of the funnel, especially on the right side of the line, serves as a strong indication that the literature on screen time and its negative effects on the psychological development of the child suffers from significant publication bias.

Figure 4. Funnel plot indicating a significant publication bias



Unfortunately, there is no way to offer an accurate estimation of the number of null results that ended up not being published. Moreover, in light of the significant publication bias and in light of the moral panic of the media, we speculate that the 21 null effects and the 5 favorable effects that did infiltrate the current meta-analysis were only possible because many of the studies examined the multiple negative effects of screens, of which at least 1 negative effect was found to be significant. The current meta-analysis suggests that there are additional empirical studies that did not find any effects of screens but due to this publication bias, they were not brought to the scientific community and therefore not brought to the public discourse.

# Summary

Despite the thorough scientific work completed by the GDG, the current article shakes the Jenga tower of the WHO guidelines for (non-) screen use among children. In Block 1, we made a clear distinction between the overall framework of the guidelines (i.e., inactivity and obesity) and the topic at hand (i.e., sedentary screen time). Then, in Block 2, we illustrated the GDG’s major reservation that summarized the section on screens with the phrase: “strong recommendations, very low-quality evidence”. In Blocks 3 and 4, we looked at two “moderate quality” studies and showed that they suffered from troubling methodological issues and that they were not relevant to the WHO guidelines. In light of the fact that the 33 studies examined a wide range of outcomes, of which many resulted in null or even favorable outcomes of screens, Block 5 challenged the assumption that a consistent pattern can be extracted to begin with. Finally, the meta-analysis conducted in Block 6 showed that the overall psychological effect of screens was negligible, and Block 7 lends evidence that the entire field has a publication bias.

The WHO document states that: “The recommendations will be updated within 10 years, unless further research in the area provides additional evidence to warrant an earlier update” (page 16). It is our position that we should not wait 10 years before we correct the current version of the guidelines. The current version is based on poor-quality science. The scientific evidence does not support the clear-cut recommendations, while the troubling headlines and the current version of the guidelines only escalate the heavy guilt that parents are already feeling.

**References**

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2011). *Introduction to meta-analysis*: John Wiley & Sons.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*: Routledge.

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., et al. (2011). GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology, 64*(4), 383–394.

Kostyrka-Allchorne, K., Cooper, N. R., Gossmann, A. M., Barber, K. J., & Simpson, A. (2017). Differential effects of film on preschool children's behaviour dependent on editing pace. *Acta Paediatrica, 106*(5), 831–836.

Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*: Oxford University Press.

Mistry, K. B., Minkovitz, C. S., Strobino, D. M., & Borzekowski, D. L. G. (2007). Children's television exposure and behavioral and social outcomes at 5.5 years: does timing of exposure matter? *Pediatrics, 120*(4), 762–769.

Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology, 90*(1), 175.

a16

Yilmaz, G., Demirli Caylan, N., & Karacan, C. D. (2015). An intervention to preschool children for reducing screen time: a randomized controlled trial. *Child: Care, Health and Development, 41*(3), 443–449.

1. Players of Jenga remove blocks and place them on the top of a wooden tower. This creates a tall and unstable structure.

   Source: Pexels - Free stock photos & videos (<https://www.pexels.com/photo/balance-game-jenga-school-72766/>)

    [↑](#footnote-ref-1)