

ARTICLE · CULTURA & SAÚDE

Background music increases viewer engagement: a systematic review and meta-analysis

A meta-analysis of 26 effects separates the attentional benefit from the one known cost

Alexandre Campos Moraes Amato

PhD, University of São Paulo (USP). Vascular and endovascular surgeon, Amato – Institute of Advanced Medicine. Brazilian Lipedema Association (ABL). São Paulo, SP, Brazil

Received for publication in August 2026 · Section: Article · Language: English

ABSTRACT

The author synthesises, in a systematic review and meta-analysis, the full range of outcomes tested during exposure to video with and without background music. A total of 2,574 records were screened across six bibliographic databases and web sources; thirteen reports contributed 15 independent samples and 26 effect sizes, pooled in multilevel random-effects models. Music raised attentional engagement largely and consistently (Hedges $g = 1.06$), a result replicated across autonomic, ocular, electrophysiological and self-report measures, whose failure modes are unrelated to one another. No outcome family showed a statistically reliable cost. The memory penalty proved confined to tests competing with concurrent narration ($g = -0.46$), approaching zero without such competition ($g = -0.13$). The author concludes that background music is a well-evidenced tool for increasing engagement, immersion and enjoyment, with a single boundary condition, which is a matter of placement rather than avoidance.

KEYWORDS: background music; attentional engagement; meta-analysis; instructional video; immersion; cognitive load

RESUMO

A música de fundo acompanha quase toda produção audiovisual profissional, mas a literatura acadêmica tem sido lida como advertência, dominada pelo achado de que ela prejudica a aprendizagem de material instrucional. Essa leitura repousa sobre fatia estreita da evidência: engajamento, imersão e fruição, que motivam a decisão de produção, nunca foram sintetizados quantitativamente ao lado da literatura de memória. O objetivo foi avaliar toda a gama de desfechos já testados durante a exposição audiovisual. Foram triados 2.574 registros de seis bases bibliográficas e de fontes web em busca de estudos que contrastassem presença e ausência de música durante a visualização de vídeo, em três famílias pré-especificadas: engaja-

mento atencional, memória do conteúdo e persistência comportamental. Os efeitos foram expressos em g de Hedges e agregados em modelos multiníveis de efeitos aleatórios. Treze relatos contribuíram com 15 amostras independentes e 26 tamanhos de efeito. A música aumentou o engajamento atencional de forma ampla e consistente ($g = 1,06$; IC 95% [0,76; 1,37]), replicado em quatro modalidades de medida com modos de falha não relacionados. Nenhuma família apresentou custo estatisticamente confiável. O prejuízo à memória restringiu-se a testes que competem com a narração simultânea ($g = -0,46$), sendo próximo de zero sem tal competição ($g = -0,13$). A música de fundo é ferramenta bem evidenciada para aumentar engajamento, imersão e fruição, com uma única condição de contorno: música sobreposta à fala

durante instrução prejudica a evocação verbal — questão de posicionamento, não de evitação.

Palavras-chave: Atenção. Imersão. Trilha sonora. Vídeo instrucional. Multimídia.

RESUMEN

La música de fondo acompaña casi toda producción audiovisual profesional, pero la literatura académica se ha leído como advertencia, dominada por el hallazgo de que perjudica el aprendizaje de material instruccional. Esa lectura se apoya en una porción estrecha de la evidencia: el compromiso, la inmersión y el disfrute, que motivan la decisión de producción, nunca se sintetizaron cuantitativamente junto a la literatura de memoria. El objetivo fue evaluar todo el rango de desenlaces ya probados. Se cribaron 2.574 registros de seis bases bibliográficas y fuentes web que contrastaran presencia y ausencia de música durante la visualización de vídeo, en tres familias preespecificadas: compromiso atencional, memoria del contenido y persistencia conductual. Los efectos se expresaron como g de Hedges y se agregaron en modelos multinivel de efectos aleatorios. Trece informes aportaron 15 muestras independientes y 26 tamaños de efecto. La música aumentó el compromiso atencional de forma amplia y consistente ($g = 1,06$; IC 95% [0,76; 1,37]), replicado en cuatro modalidades de medida con modos de fallo no relacionados. Ninguna familia mostró un costo estadísticamente confiable. El perjuicio a la memoria se restringió a pruebas que compiten con la narración simultánea ($g = -0,46$) y fue casi nulo sin tal competencia ($g = -0,13$). La música de fondo es herramienta bien evidenciada para aumentar el compromiso, la inmersión y el disfrute, con una única condición de contorno: la música superpuesta al habla durante la instrucción perjudica la evocación verbal — cuestión de ubicación, no de evitación.

Palabras clave: Atención. Inmersión. Banda sonora. Vídeo instruccional. Multimedia.

INTRODUCTION

Background music accompanies most professionally produced video, and the practitioner rationale is straightforward: music makes the viewing experience more absorbing. The academic literature, however, has often been summarised as advising against it. The coherence principle in the cognitive theory of multimedia learning holds that non-essential audio consumes limited processing capacity, and the finding that adding an instrumental music loop to a narrated animation reduced both recall and problem-solving transfer (MORENO; MAYER, 2000) has been widely generalised into a recommendation to strip music from instructional video.

That generalisation outruns its evidence in two ways. First, it treats a result obtained under one specific

configuration — instrumental music playing concurrently with narration during a short animation — as a property of music in general. Second, it privileges a single outcome, verbal recall, over the outcomes that motivate the production choice in the first place: whether viewers are engaged, absorbed and enjoying what they watch. The gap, then, is not one of data but of synthesis: what is missing is a common metric on which the documented benefits can be weighed against the documented cost.

A parallel literature has been measuring exactly those outcomes, and its results are strikingly consistent. Music raises autonomic arousal across video genres (WÖLLNER; HAMMERSCHMIDT; ALBRECHT, 2018), increases pupil diameter (HAMMERSCHMIDT; WÖLLNER, 2018), produces greater alpha-rhythm suppression and larger P300 amplitudes than silence (KWON; LEE; LEE, 2022), and suppresses default-mode network activity associated with mind-wandering during free viewing (BEZDEK; WENZEL; SCHUMACHER, 2017). This is the state of the art on the engagement side, and it has never been confronted quantitatively with the memory literature.

This work provides that synthesis. The question is not whether music helps or harms in the abstract, but what it demonstrably does, how large those effects are, how consistently they replicate, and under precisely which conditions the one known cost appears.

MATERIAL AND METHODS

ELIGIBILITY

The review followed PRISMA reporting conventions. Eligible records involved human participants, an explicit contrast between a music-present and a music-absent condition, and a quantitative outcome in one of three prespecified families: attentional engagement (psychophysiological indices, eye movements, self-reported immersion, arousal or enjoyment); content memory (recall, recognition, comprehension, transfer); and behavioural persistence (automatically logged duration of voluntary exposure). Designs contrasting only different kinds of music, with no silence baseline, were recorded but excluded from pooling. No date, language or publication-status restrictions applied.

SEARCH AND SCREENING

Searches ran across OpenAlex, PubMed, arXiv, Crossref, bioRxiv/medRxiv and general web sources in three parallel tracks. Because the persistence literature proved sparse for video, that search was extended to adjacent domains where music's effect on time spent is the dependent variable, with domain recorded as a coded variable. After keyword searching, the three most relevant hits per track were expanded one step backward through their reference lists and one step forward through their citing works. Every DOI was resolved against Crossref or

OpenAlex, and all bibliographic metadata was taken from the resolved record.

EFFECT SIZES AND SYNTHESIS

All effects were expressed as Hedges g , signed so that positive values indicate music improved the outcome. Where per-condition means and standard deviations were reported, g was computed directly, pooling music-present cells across tempo or congruency levels. Otherwise t and F were converted to d and bias-corrected, and partial eta squared was converted via Cohen's f . Within-subject contrasts used the d_{av} metric assuming $r = 0.5$. Each conversion is documented, effect by effect, in the supplementary dataset.

Two extraction decisions warrant note. In Kwon, Lee and Lee (2022), the dispersions reported around the EEG and event-related potential means are implausibly small relative to the reported F statistics and appear to be standard errors; effect sizes were therefore derived from the reported partial eta squared. In Moreno and Mayer (2000), the marginal means presented in the text are inconsistent in

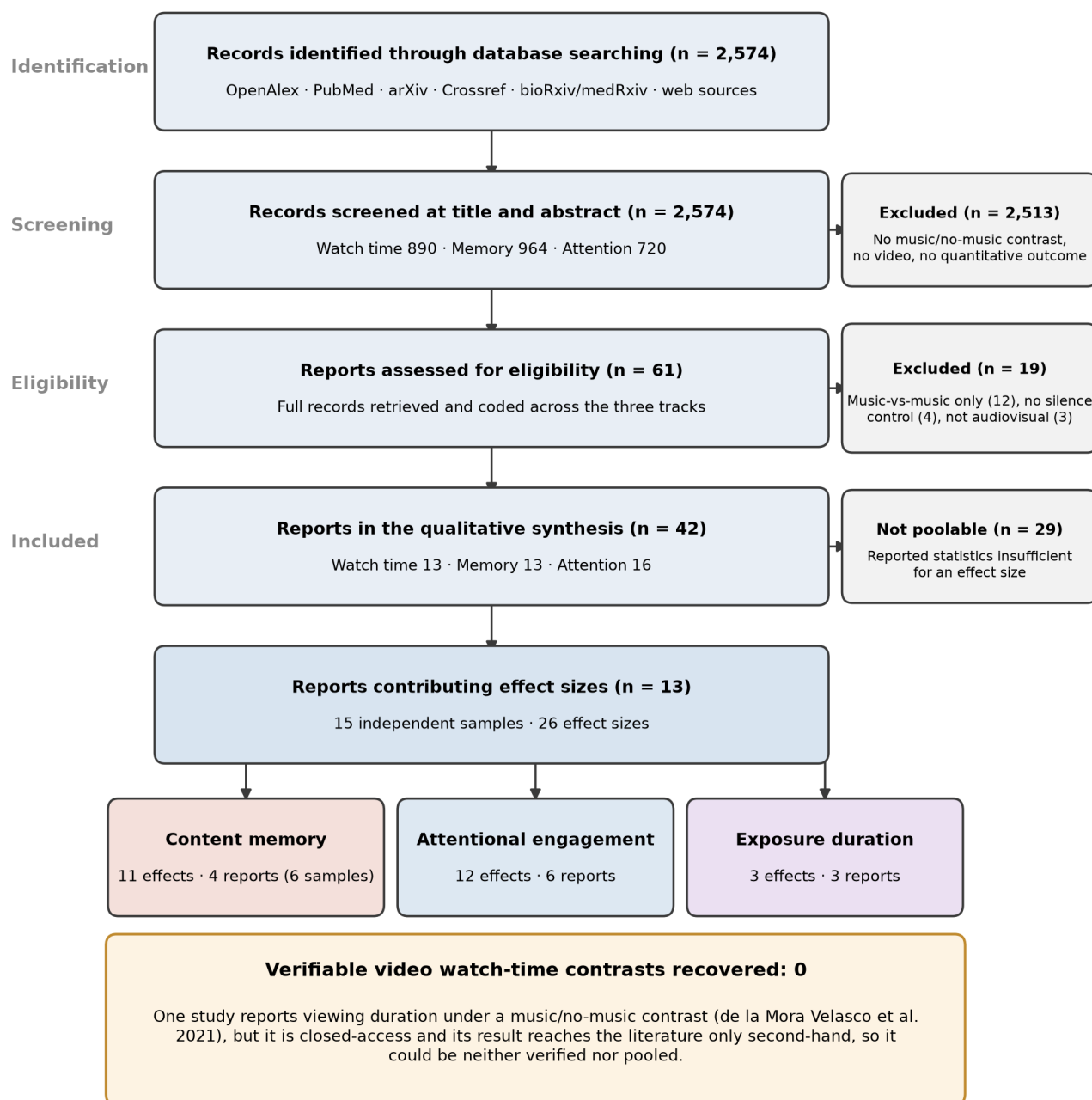
sign with the cell means in Table 2 of the original article; the cell means were used, as they match the reported direction.

Each family was pooled in a multilevel random-effects model (REML) with random intercepts at study and effect level. Pooled estimates are reported with 95% confidence and prediction intervals, τ^2 at both levels, Cochran's Q , and a multilevel I^2 . Prespecified subgroup analyses split the memory family by test depth and the attention family by measurement type. Small-study effects were assessed by regressing effect size on standard error and by trim-and-fill on study-aggregated estimates. Analyses used the metafor 5.0-1 package in R.

RESULTS AND DISCUSSION

Of 2,574 screened records, 61 were assessed in full, 42 entered the qualitative synthesis, and 13 reports, contributing 15 independent samples and 26 effect sizes, were poolable (figure 1). The 26 individual effects, organised by outcome family, are presented in figure 4.

Figure 1 – PRISMA flow diagram of study selection.



Caption: Records identified, screened, assessed for eligibility and included in the synthesis, with the distribution of the 26 effect sizes across the three outcome families. No verifiable video watch-time contrast was retrieved.

Source: prepared by the author.

THE ENGAGEMENT BENEFIT IS LARGE, CONSISTENT AND MODALITY-INDEPENDENT

Background music increased attentional engagement by $g = 1.06$ (95% CI [0.76, 1.37]; $p < 0.0001$; $k = 12$ effects from six reports; figure 2a). Every included effect was positive. Heterogeneity was high ($I^2 = 90\%$), but the prediction interval remained entirely above zero ([0.05,

2.08]), meaning a new study in this family would be expected to find a positive effect.

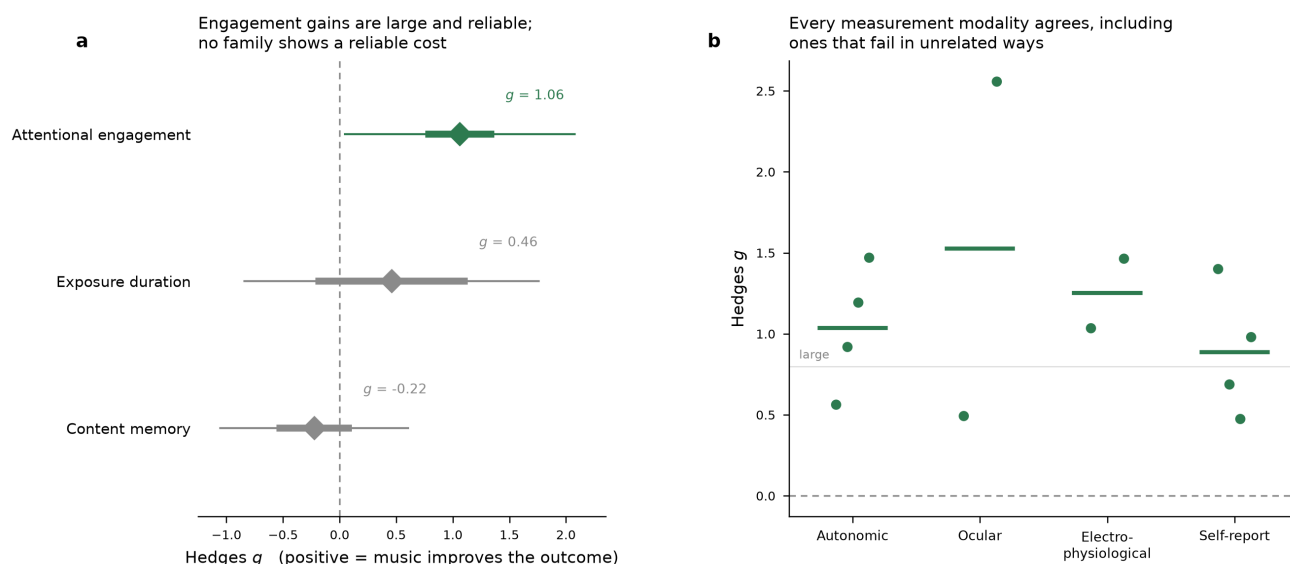
The strength of this result lies in its convergence (figure 2b). Autonomic measures ($g = 1.04$; for example, skin conductance during a suspense film, BENTE et al., 2022), ocular measures ($g = 1.53$), electrophysiological measures ($g = 1.25$) and self-report ($g = 0.89$) all detect the same benefit, and these methods have entirely unrelated failure modes: demand characteristics that could inflate a self-

report cannot inflate skin conductance, and electrode artefacts that could distort an EEG signal cannot distort a questionnaire. Agreement across them is the strongest form of evidence available for a construct of this kind.

Physiological and experiential estimates were also close to one another ($g = 1.18$; 95% CI [0.76, 1.61] against $g = 0.93$; 95% CI [0.46, 1.40]; figure 3b). This matters for

production: viewers do not merely register music sub-perceptually, they consciously experience the heightened engagement it produces. Muting the soundtrack of subtitled video reduced enjoyment ($g = 0.69$) and immersion ($g = 0.48$) in a within-subjects design with 161 participants (SZARKOWSKA et al., 2024), among the most precisely estimated effects in this review.

Figure 2 – The benefit profile of background music.



Caption: (a) Pooled random-effects estimates by outcome family; thick bars are 95% confidence intervals and thin lines 95% prediction intervals. (b) Individual attentional effects grouped by measurement modality, with horizontal bars showing modality means. Positive values of g indicate music improved the outcome. Families whose interval spans zero are drawn in grey.

Source: prepared by the author.

NO OUTCOME FAMILY SHOWS A RELIABLE COST

Eighteen of the 26 effects favoured music and 14 excluded zero. Neither of the other two families produced a statistically reliable effect in either direction: content memory pooled to $g = -0.22$ (95% CI [-0.56, 0.11]; $p = 0.19$) and exposure duration to $g = 0.46$ (95% CI [-0.21, 1.13]; $p = 0.18$).

The memory result deserves precision rather than reassurance. Its central tendency is negative, and the interval includes values that would represent a meaningful cost as well as values representing a benefit; under the fitted random-effects distribution, the probability that the true effect in a new study is trivial ($|d| < 0.20$) is only 0.34. What the data rule out is a reliable general memory penalty, not the possibility of a real one in specific configurations. Those configurations, as will be seen, are identifiable.

THE MEMORY COST IS A BOUNDARY CONDITION, NOT A GENERAL PROPERTY

Splitting the memory family by test depth localises the negative signal precisely (figure 3a). Tests requiring effortful verbal reconstruction while narration is present — free recall and problem-solving transfer — pooled to $g =$

-0.46 ($k = 5$). Tests that do not compete for the verbal channel — recognition and matching — pooled to $g = -0.13$ (95% CI [-0.40, 0.14]; $k = 6$), close to zero. The moderator contrast is significant ($b = -0.44$; $p = 0.017$).

Three findings independently confirm that concurrency with speech, rather than music itself, is the operative variable. Music placed in a pretraining segment, where it did not compete with narration, improved retention ($F(1,88) = 8.01$; $p = 0.006$), while the same music inserted inside the narrated content produced no effect at all ($p = 0.53$) (PHAN, 2023). Music was detrimental for the easier of two instructional videos and null for the harder one, where the content itself already occupied attention (MEYERHOFF et al., 2022). And a high-powered replication of the television-news line of research found no recognition differences at all (KOPIEZ; PLATZ; WOLF, 2013).

A further caution against reading the memory literature as uniformly negative: in narrative film, music does not raise or lower accuracy so much as reshape *what* is encoded, through mood congruency (DAMJANOVIC; KAWALEC, 2021). For narrative and documentary work, that is a creative instrument rather than a cost.