Review of Educational Research
June 2022, Vol. 92, No. 3, pp. 413–454
DOI: 10.3102/00346543211052329
Article reuse guidelines: sagepub.com/journals-permissions

Multimedia Design for Learning: An Overview of Reviews With Meta-Meta-Analysis

Michael Noetel and Shantell Griffith
Australian Catholic University

Oscar Delaney and Nicola Rose Harris
University of Queensland

Taren Sanders and Philip Parker
Australian Catholic University

Borja del Pozo Cruz
University of Southern Denmark

Chris Lonsdale
Australian Catholic University

Multimedia is ubiquitous in 21st-century education. Cognitive load theory and the cognitive theory of multimedia learning both postulate that the quality of multimedia design heavily influences learning. We sought to identify how to best design multimedia and review how well those learning theories held up to meta-analyses. We conducted an overview of systematic reviews that tested the effects of multimedia design on learning or cognitive load. We found 29 reviews including 1,189 studies and 78,177 participants. We found 11 design principles that demonstrated significant, positive, meta-analytic effects on learning and five that significantly improved management of cognitive load. The largest benefits were for captioning second-language videos, temporal/spatial contiguity, and signaling. We also found robust evidence for modality, animation, coherence/removing seductive details, anthropomorphics, segmentation, personalization, pedagogical agents, and verbal redundancy effects. Good design was more important for more complex materials, and in system-paced environments (e.g., lectures) than self-paced ones (e.g., websites). Results supported many tenets of both theories. We highlight a range of evidence-based strategies that could be implemented by educators.

Keywords: multimedia learning, presentation design, video, e-learning, online learning

Multimedia is everywhere. So, it is critical educators understand how to design it well. Done well, multimedia leverages both of our key information processing systems—sight and sound—using the strengths of each. Compared with media that only use one information processing channel (sight *or* sound), multimedia can improve learning (Mayer, 2008; Noetel et al., 2021; Rolfe & Gray, 2011). But this is not always the case. Done poorly, multimedia can lead to cognitive overload, needless distraction, and poorer learning (Mayer, 2008, 2009; Sweller et al., 2019; van Merriënboer & Sweller, 2005). How can educators design effective multimedia? What explains the differences between good and bad multimedia?

Previous reviews have either advanced theoretical proposals explaining well-designed multimedia using a selection of empirical studies (e.g., Mayer, 2008; Sweller et al., 2019; van Merriënboer & Sweller, 2005) or conducted meta-analyses on the effects of one discrete, actionable recommendation for multimedia (e.g., remove distracting details; Sundararajan & Adesope, 2020). Our meta-review aims to synthesize both sets of literature, assessing the hypotheses of theoretical papers against the results of empirical meta-analyses. A meta-review (or overview of reviews) functions like a systematic review of primary studies, but instead of comprehensively collating the primary studies on a topic, meta-reviews collate systematic reviews. We first outline the theoretical explanations for effective multimedia, then discuss limitations of existing reviews. We describe our systematic search for systematic reviews of multimedia effects, compare the effectiveness of different design principles, then integrate those results with key theoretical models. Given multimedia is the lingua franca of 21st-century education, we outline how to design good multimedia and explain why good design works.

Multimedia Works Well When Managing Cognitive Load Across Two Processing Systems

Multimedia includes any form of communication that combines words and pictures (Mayer, 2008, 2009). Lectures, videos, textbooks, videoconferences, and online learning almost always contain both words and illustrations. It is almost easier to define multimedia by what the term does not include: podcasts are exclusively words (via audio), photos are just an image, and many academic papers present only text. Almost everything else involves multimedia.

Multimedia varies wildly in its effectiveness. As we all know, just because presenters use videos or PowerPoint slides does not mean they are presenting effectively. Meta-analyses of video and PowerPoint show considerable unexplained heterogeneity—some interventions work very well and others are harmful—which is likely due to variability in how well the media are designed (Baker et al., 2018; Noetel et al., 2021). One key design consideration is how well multimedia manages cognitive load.

Cognitive Load Theory

Cognitive load theory describes how learning is affected by the natural limits of our working memory (Sweller et al., 1998; Sweller et al., 2019; Van Gog et al., 2005; van Merriënboer & Sweller, 2005). We can usually hold only a small number of elements in our mind simultaneously (Kalyuga, 2011; Sweller, 2010). In contemporary approaches to cognitive load theory, the key bottleneck is our ability to manage the *element interactivity* (Sweller, 2010): the complexity of the materials, operationalized by the number of links between concepts. For example, learning flags of foreign countries would be "low element interactivity" because each element could be studied in isolation (the image of the flag and the name of the country). In contrast, learning how the lymphatic system works requires "high element interactivity" because the learner must understand the relationships between many elements (e.g., how the thymus gland, spleen, bone marrow, and lymph nodes work together).

Intrinsic load refers to working memory requirements that are inherent in the task itself (e.g., mentally holding the country name "Chad" and an image of its flag). When a task is inherently too difficult for a learner (e.g., "1,456 × 2,134") it is because the intrinsic load uses more working memory than is available. In this case, we would need pen and paper to free some of our cognitive load. Alternatively, when learning a concept with high element interactivity, we can either change the basic task or try to increase the learner's knowledge (Kalyuga, 2011; Sweller, 2010). When learning the lymphatic system, we could focus on one component of the system (e.g., the spleen) to reduce the number of elements that interact. If done successfully, the learner develops more complex models of each concept (e.g., what a spleen does), meaning the one "element" represents a more complex construct. As a result, increasing the learner's knowledge reduces the amount of intrinsic load placed on the learner, because each "element" becomes more complex (Kalyuga, 2011; Sweller, 2010). Once learners have easily accessible internal models of each component in a system (e.g., complex models of a spleen and of lymph nodes) they can then more easily handle the interactions between each element.

So, intrinsic load is complexity inherent in acquiring knowledge. Extrinsic load, on the other hand, is the complexity that a teacher adds to a learning task that *detracts* from learning (Kalyuga, 2011; Sweller, 2010). If the intrinsic load is high enough, extrinsic load detracts from learning because the learner cannot hold both the relevant and irrelevant element interactions at the same time (Kalyuga, 2011; Sweller, 2010). Poorly designed multimedia overwhelms our cognitive systems by requiring that we connect more elements than we can handle (Kalyuga, 2011; Sweller, 2010). A familiar example of multimedia overload is "death by PowerPoint," when slides flood our visual system with words, while our auditory system is trying to process additional, often conflicting, words (Sweller et al., 2019; van Merriënboer & Sweller, 2005). This is overwhelming because some of these element interactions are extraneous—they take up working memory but do not facilitate learning.

One of the key goals of implementing cognitive load theory is to reduce this extraneous cognitive load (Sweller et al., 1998; Sweller et al., 2019; Van Gog et al., 2005; van Merriënboer & Sweller, 2005). Reducing extraneous load frees up attention that can be redirected to cognitive processes that facilitate learning (Kalyuga, 2011; Sweller, 2010). For complex tasks, removing extraneous load provides space in working memory for the interactions inherent in the learning task (Kalyuga, 2011; Sweller, 2010). For simple tasks, this freed space allows for more sophisticated approaches to learning the content (Kalyuga, 2011; Sweller, 2010). Early models of cognitive load theory identified a third type of load called germane load, which involved cognitive processes not *inherent* in the task but that facilitated learning (Sweller et al., 1998). For example, a learner might not *need* to know a metaphor for how a spleen works, but hearing it described as "an elaborate wetland" (Angier, 2009) may help some learners connect their new understanding of a spleen to an established mental model (i.e., of wetlands). Students may draw concept maps, connecting their new knowledge to existing, related knowledge (Nesbit & Adesope, 2006). Newer models of cognitive load theory do not describe this load as a different type of cognitive load, but instead as a type of intrinsic load for more complex schema construction (Kalyuga, 2011; Sweller, 2010). As above, metaphors or concept maps might not be inherently required for learning the basics of a concept, but they do help develop more sophisticated and robust internal models (Nesbit & Adesope, 2006). These types of intrinsic load could be initiated by the educator (e.g., asking the class to draw a concept map of the lymphatic system) or by sufficiently motivated students (e.g., by coming up with their own metaphors) provided they are not overwhelmed by the extrinsic load. Learning designs that reduce extraneous load allow for more attention to be directed to fruitful learning processes.

Cognitive load can be influenced by many factors, from the goals given to learners, to the presence of worked examples, to the skills of the learner (Sweller et al., 1998; Sweller et al., 2019; Van Gog et al., 2005; van Merriënboer & Sweller, 2005). However, given the ubiquity of multimedia for learning, particular attention has been drawn to the influence of multimedia design on cognitive load (Mayer, 2008). This is also because multimedia needs to manage cognitive load in two separate information processing systems.

Cognitive Theory of Multimedia Learning

The cognitive theory of multimedia learning (Mayer, 2008, 2014) focuses on how cognitive load theory explains variations in the effects of multimedia. The theory identifies how our visual and auditory processing channels respond to information, and how that information is integrated into long-term memory (Mayer, 2008; Mayer et al., 2001). It assumes that our cognitive architecture is built around two parallel systems, one for what we see and one for what we hear. This is called the dual-channel assumption (Mayer, 2009). These systems interact and are cross-functional; for example, we can read text with our eyes that is translated into our phonological loop (used for hearing). However, good multimedia is hypothesized to use each channel in its natural format (seeing images and hearing sounds; Mayer, 2008, 2014). As with cognitive load theory, the cognitive theory of multimedia learning proposes that the attention available for each channel is limited (limited-capacity assumption; Mayer, 2009). One final, "active processing" assumption is that learning involves an active, coordinated cognitive process; mere passive exposure is not sufficient (Mayer, 2009).

Together, these theories propose that learning is optimized when students are given tasks they can handle, when they are free from distracting content, when they leverage both streams of information processing, and when they actively engage in learning (even if only to reduce extraneous processing). Some key benefits from these theories are the many hypothesized interventions to improve learning. These "multimedia design principles" aim to improve learning by reducing extraneous load, by making intrinsic load more manageable, or by leveraging our dual processing channels. The theories also seek to explain when the interventions work and when effects might be weak.

Multimedia Design Principles and Hypothesized Moderators

Modality principle is a useful example of a multimedia design principle that leverages both theories, but where the two theories suggest distinct mechanisms. The modality principle proposes that images should be complemented with audio rather than with text (Mayer, 2009). For example, educators may use videos instead of textbooks for prelearning, or delete words from presentation slides and instead explain images verbally. From the perspective of cognitive load theory, this change reduces extrinsic load because the instructional design requires less effort to connect interacting elements (Mayer & Moreno, 2010). That is, it takes less effort to connect what we hear to what we see than it does to connect text we read with images on a page, so the extra effort is extraneous load (Mayer & Moreno, 2010). From the perspective of the cognitive theory of multimedia learning, the major problem with text accompanying images is that the design overloads the learner's visual channel (Mayer & Moreno, 2010). Both the text and the image may be essential, so is "intrinsic load," but offloading the text to the auditory mode makes that intrinsic load more manageable (Mayer & Moreno, 2010). This example highlights the main ways both theories improve instructional design: both describe the need to reduce extraneous load, the cognitive theory of multimedia learning emphasizes the need to leverage both processing channels (auditory and visual), and cognitive load theory describes strategies that apply outside of multimedia learning (e.g., the goal-free effect, described below).

Strategies to Reduce Extraneous Load

There are a range of strategies both theories use to reduce extraneous load (Mayer, 2008, 2009; Sweller et al., 2019; van Merriënboer & Sweller, 2005). The coherence principle, or "seductive detail effect," suggests educators should avoid content that could distract learners from core content required for the task. They may remove distracting music or footage from a video, remove distracting images from presentation slides, or remove tangentially relevant stories from a lecture. These "seductive details" add extraneous load because they increase element interactivity, but for content that is not relevant (Mayer & Moreno, 2010). Similarly, "personalization" involves educators using learner-focused conversational language so people are not overloaded by complex vocabulary or jargon (Mayer, 2009). Educators might use metaphors to explain abstract concepts, or use common verbs (e.g., "think") instead of abstract noun-formations (e.g., "considerations"; avoiding so-called "nominalization"). Sometimes jargon is intrinsic to the task at hand (e.g., learning what "photosynthesis" means), but where it is not, simple language reduces the complexity of the interacting elements, reducing extrinsic load (Sweller, 2010).

Sometimes the amount of information is difficult to change, but instead educators can make it easier for learners to direct their attention usefully by cueing, timing, or spacing. The signaling (or "cueing") principle helps learners know where to focus so they do not expend effort either searching for ideas or connecting them together (Mayer, 2009). It might involve circling a key point on a video, using a laser pointer to direct attention on slides, or emphasizing a key point with a phrase that focuses learner attention (e.g., "this is important"). Each of these help the learner to focus their attention on the intrinsic elements and ignore the extraneous materials (Mayer, 2009). The contiguity principle serves a similar purpose, but by presenting related material in the same place or the same time (Mayer, 2009). For example, educators might present key points, one at a time, as they explain them, rather than having multiple key points visible at once (temporal contiguity). Alternatively, they might label images directly, rather than using legends, where learners have to connect the legend to the image (spatial contiguity). Like signaling, these help manage extraneous load by both focusing attention on what is most important and by helping the learner to connect interacting elements (e.g., spoken words with the correct section of the slide; Mayer, 2009).

Strategies to Better Manage Intrinsic Load

As described with the modality principle, the cognitive theory of multimedia learning proposes that we can better manage cognitive load by better using both modes of information processing (auditory and visual; Mayer, 2008, 2009). As with moving text to speech (modality principle), educators can also use informative graphics to offload some intrinsic load to the visual system (i.e., multimedia principle; Mayer, 2009). Educators might add visuals to textbooks, add informative images to presentation slides, or display data visually, instead of only using text. Both cognitive load theory and the cognitive theory of multimedia learning propose a range of other approaches for making intrinsic load more manageable. Both discuss how breaking longer content into smaller chunks allows for better processing of the intrinsic load (segmenting principle; Mayer & Moreno, 2010). This may involve converting long recordings into shorter clips, adding breaks or discussions to a workshop/lecture, or using subheadings or short chapters in written work. Similarly, learners might be encouraged to learn simpler components in isolation before moving onto more complex concepts (Sweller, 2010). Cognitive load theory also describes many phenomena that help optimize learning beyond the presentation of multimedia. Educators could provide learners with worked examples so they are not burdened with the task of finding the correct way to do something while *also* learning how to *implement* that approach (Sweller, 2010). In designing learning activities, educators can reduce cognitive load by making activities "goal-free" (Sweller, 2010). If students are asked to "find as many values as possible," instead of "find *x*," they are not encumbered by the goal-state and the discrepancy between their current state and the goal.

Hypothesized Moderators of Effects

Both theories hypothesize that a number of factors increase or decrease these effects of these design principles. The core mechanisms here are that our limited working memory capacity plays a bigger role when (a) the materials are more complex, so intrinsic load is higher; (b) we are newer to an area, so more cognitive load is needed for each concept; and (c) the presentation format is less forgiving, such that we cannot make up for overload through time and effort. Specifically, good design principles are more important for concepts that have high element interactivity (i.e., they are more complex; Sweller, 2010). They are more important for novices who are more likely to become cognitively overloaded and can sometimes reverse when implemented in experts (expertise reversal effect; Sweller, 2010). For example, junior doctors may find anatomy clearer with labels directly on the image (e.g., brain regions), but senior doctors may be so familiar with anatomy that the labels become distracting. Finally, multimedia design may be more important when learners cannot self-pace (Mayer & Moreno, 2010). If learners can pause, rewind, or reread content, then they can somewhat manage their own cognitive load by taking things more slowly if they are overwhelmed (Mayer & Moreno, 2010). Many environments, however, do not allow such learner-control (e.g., lectures, workshops). All of these moderators and design principles are empirically testable hypotheses, and the best data for testing them are from systematic reviews.

Systematic Reviews Robustly Assess When Interventions Work, and Help Clarify Why

In the previous section, we compiled a range of hypothesized multimedia design principles drawn from key theory papers (Mayer, 2008, 2009; Mayer & Moreno, 2010; Sweller, 2010; Sweller et al., 2019; van Merriënboer & Sweller, 2005). These papers are important theoretical contributions, integrating large sets of existing literature into refined theories and actionable recommendations. However, they employed unsystematic methods of determining the strength of proposed effects. For example, Mayer (2008) presented 10 separate meta-analyses of multimedia design principles using the studies conducted in his lab alone. The studies were not drawn from a systematic review of the literature. Rather, Mayer calculated the unweighted median effect size from a group of studies he and his colleagues had conducted. A more robust, gold-standard method of assessing the effects of these interventions is a systematic review and meta-analysis (Alexander, 2020; Pigott & Polanin, 2020). Systematic reviews are more robust because they aim to reproducibly collect all the available evidence on a topic and synthesize it in a way designed to minimize bias (Alexander, 2020; Pigott & Polanin, 2020). Systematic reviews exist for some of these multimedia design principles (e.g., Adesope & Nesbit, 2012; Ginns, 2005; Ginns et al., 2013; Reinwein, 2012; Rey, 2012). Those reviews help resolve controversies—where different theory papers make divergent hypotheses—and help assess the strength of the evidence in support of those hypotheses.

As a concrete example, verbal redundancy has been a controversial design principle, with some saying it helps (Adesope & Nesbit, 2012; Moreno & Mayer, 2002) and others saying it hinders (Mayer, 2008). Verbal redundancy is where multimedia present words via both visual and auditory pathways (e.g., if lecturers were to read quotes also presented on their slides) rather than via one channel (e.g., just reading the quote *or* putting it on slides). Cognitive load theory proposes that this redundancy creates unnecessary duplication that becomes extraneous load (Mayer, 2008; Sweller, 2010). Conversely, the dual-channel assumption of multimedia learning suggests that the two channels could symbiotically facilitate learning (e.g., seeing the text could help learners who missed what the presenter said; Moreno & Mayer, 2002). Adesope and Nesbit's (2012) systematic review synthesized the available studies on verbal redundancy to determine when it helped, when it hindered, and the size of the effects. The overall pooled effect size showed that verbal redundancy modestly improved learning (g = 0.15, 95% confidence interval [CI 0.08, 0.22], k = 57), supporting the dual channel assumption. Specifically, it helped add text to audio presentations but not the other way around (Adesope & Nesbit, 2012). This finding may be because audio is less forgiving than reading text. If learners lose focus or they mis-read something, it is easier to self-correct by rereading than to go back and re-listen.

Regardless, effects are small, particularly when compared with meta-analyses of a related principle: the modality effect (Ginns, 2005; Reinwein, 2012). The modality effect suggests that media are more effective when we hear words and see pictures rather than seeing both. Pooled effect sizes of these effects are moderate-to-large (d = 0.72, 95% CI [0.52, 0.92], k = 39; Ginns, 2005; d = 0.38, 95% CI [0.27, 0.50], k = 86; Reinwein, 2012) and robust to a number of sensitivity analyses. This meta-analytic evidence substantiates stronger causal claims (e.g., "redundancy is helpful") than reviews that either present a subset of confirmatory papers (Sweller et al., 2019) or pool effects from an unrepresentative sample of studies (Mayer, 2008).

Comparing these three meta-analyses (Adesope & Nesbit, 2012; Ginns, 2005; Reinwein, 2012) highlights the need for our overview of reviews. First, the implications for teachers are clearest when interpreting multiple meta-analyses in parallel. Taking the above-mentioned examples, teachers should strive to present content using their voice rather than via text (Ginns, 2005; Reinwein, 2012), but only when there are accompanying images. Teachers using their voice without any visual prompts inhibits learning, so when images are not available, adding text on screen facilitates self-correction (Adesope & Nesbit, 2012). Second, there are hundreds of studies on implementing a range of multimedia design principles. A single systematic review of all these primary studies would be intractable, but an overview of reviews allows for the most robust interventions to be collected and compared. Third, there are frequently multiple meta-analyses on the same topic that can produce substantially different effect sizes (e.g., Ginns, 2005, vs. Reinwein, 2012) or effects on different metrics (e.g., Cohen's d, Hedge's g, Pearson's r, Fisher's z). An overview of reviews allows us to synthesize the effects, compare them on a common metric, and make judgments accounting for the quality of the reviews. These comparisons contribute to theory by determining whether there are common factors that underlie the most powerful effects, or factors that might explain weaker ones. Finally, some moderators of intervention effects—like the expertise reversal effect—are observational by nature (Deeks, 2019). Participants might be randomized within studies, but they are not randomized to one study or another (Deeks, 2019). As a result, an overview of reviews can help determine which of these moderators are robust across reviews, in the same way a review of primary studies can determine whether an intervention is robust across studies.

This Overview Aims to Catalogue, Critique, and Compare Literature on Multimedia Design

In this overview of reviews, we aimed to identify *how* educators can most reliably make learning from multimedia more effective. We collated the best evidence on the various multimedia design principles, compared them on a common metric, and quantified their relative effects using meta-meta-analysis. We also assess whether proposed moderators of effects—like the expertise reversal effect—are robust across multimedia design principles. Reviews of multimedia design are particularly well suited to meta-meta-analysis because each intervention is hypothesized to operate via a similar causal pathway (improved management of cognitive load), most primary studies have a consistent comparison condition (presence vs. absence of a design principle), and reviews report a consistent set of outcomes (standardized measures of learning or cognitive load). This allowed us to assess the influence of these design principles across all kinds of learning environments where multimedia are present (lectures, videoconferences, videos, textbooks, schools, universities, online, face-to-face), allowing us to make more generalizable conclusions. We hypothesized that learning is increased by multimedia design that aims to better manage cognitive load.

Method

Presentation in this article is aligned with the PRISMA 2020 statement (Page et al., 2021) and the Reporting Standards for Research in Psychology (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008). Where possible, we adhered to recommended processes for high-quality systematic reviews of educational research (Alexander, 2020; Pigott & Polanin, 2020).

Eligibility Criteria

We selected studies on the basis of prespecified inclusion criteria:

1. *Design*: We included systematic reviews, meta-analyses, or similar reviews characterized by Grant and Booth (2009), which includes rapid reviews, scoping reviews, state-of-the-art reviews, systematic searches with review, systematized reviews, mapping reviews, or systematic maps. All these methods have focused research questions and reproducible methods for sourcing and evaluating research, so are less likely to be subject to biases like selective reporting. For these reasons, we excluded non–systematic reviews, primary studies, theory papers, or narrative reviews (e.g., Sweller, 2010).
2. *Interventions*: We included any *modification* to a multimedia intervention designed to influence learning or cognitive load. We excluded reviews only comparing the effects of multimedia against other instructional

technologies (e.g., multimedia vs. nothing, multimedia vs. tutoring; e.g., Baker et al., 2018; Noetel et al., 2021).

3. *Comparisons*: We included reviews that compared multimedia with and without a multimedia design feature (e.g., with and without signaling). To be eligible, the participants did not need to be randomized in all primary studies (e.g., we included reviews that contained pre-post or quasi-experimental primary studies).
4. *Outcome*: We only included reviews that reported learning, achievement, or cognitive load. We excluded reviews that *only* reported effects on other outcomes (e.g., health; Hirschey et al., 2020; or enjoyment), but extracted all outcomes from reviews that met criteria.
5. *Participants/context*: Reviews on any participants were eligible for inclusion. Similarly, reviews on learning in any context (e.g., schools, universities, adult learning, public health) were eligible for inclusion; as presented below, results were moderated by the context in which they were observed (where reported).
6. Finally, reviews published in any language were eligible, but searches were limited to post-1989 when systematic forms of review were introduced (Smith et al., 2011).

Information Sources and Search Strategy

To ensure it was comprehensive (Alexander, 2020), we generated our search strategy using the titles and abstracts of an initial sample of target systematic reviews (Adesope & Nesbit, 2012; Ginns, 2005; Ginns et al., 2013; Reinwein, 2012; Rey, 2012). We also used the names of all interventions listed in papers outlining the cognitive theory of multimedia learning (e.g., "signaling," "temporal contiguity"; Mayer, 2008, 2009, 2014, 2019). We generated a list of terms that identified all target papers (Eriksen & Frandsen, 2018) optimizing for sensitivity while maintaining specificity. The list of search terms were as follows:

- *Design*: (meta-anal* *or* metaanal* *or* meta-regress* *or* "systematic review" *or* "rapid review" *or* "scoping review" *or* "State-of-the-art review" *or* "Systematic search and review" *or* "Systematized review" *or* "Mapping review" *or* "Systematic map") *and*
- *Intervention*: (Multimedia *or* Video *or* "Modality effect" *or* "Segmenting effect" *or* "Modality principle" *or* "Segmenting principle" *or* "Spatial contiguity" *or* "Temporal contiguity" *or* "Signalling effect" *or* "Signaling effect" *or* "Signalling principle" *or* "Signaling principle" *or* "Verbal Redundancy" *or* Pretraining *or* Pre-training *or* "Coherence effect" *or* "Coherence principle" *or* "Personalization effect" *or* "Personalisation effect" *or* "Personalization principle" *or* "Personalisation principle" *or* "Spatial ability") *and*
- *Outcome*: (Learn* *or* Achiev* *or* Cognitive)

We entered this set of terms into four databases: ERIC, PsycINFO, Scopus, and Web of Science (translations in Supplementary File 1, available in the online version of this article). We conducted searches in January 2020 and updated searches in December 2020.

Study Selection

We removed duplicates using the deduplication tool in *SR-Accelerator* (Rathbone et al., 2015). To screen against inclusion criteria, two reviewers independently screened titles and abstracts in duplicate using *abstrackr* (Wallace et al., 2012), with conflicts resolved by moving the article to full-text screening. Two reviewers then independently assessed each full-text against the inclusion/exclusion criteria in duplicate, with conflicts resolved through discussion or consultation with a third reviewer. Reviewers logged reasons for excluding any potentially relevant reviews. One reviewer completed forward and backward citation searching for papers that may have been missed by the search strategy (Pigott & Polanin, 2020).

Data Items and Collection Process

Two reviewers developed, piloted, and revised a data extraction form. The form extracted details regarding the review methods (i.e., inclusion/exclusion criteria, search strategy, search terms, databases searched), review results (i.e., search yield, cumulative participants, author summaries of results), and results of each meta-analysis including moderation/sensitivity analyses (i.e., pooled effect size estimates with variances; number of studies and participants informing effect size; participants, intervention, comparison, and outcome for each effect). We independently extracted all data items in duplicate and resolved disagreements via discussion. A third author adjudicated, when required.

Quality Assessment of Reviews

We assessed the quality of the included reviews using an abbreviated list of quality criteria drawn from AMSTAR2 (Shea et al., 2017). The full AMSTAR2 tool was developed on the basis of a scoping review of tools for assessing the quality of systematic reviews, followed by iterative workshops with content experts (Shea et al., 2017). This abbreviated list has been used to assess core quality requirements for systematic reviews, such as those discussed by Alexander (2020) and others (Pigott & Polanin, 2020): comprehensive search, clear inclusion/exclusion, duplicate screening, duplicate quality assessment, presentation of each included study, assessment of heterogeneity, and assessment of publication bias. These quality assessment items are judged to be the most important in assessing the reliability and validity of systematic reviews (National Heart, Lung, and Blood Institute, n.d.) and deliberately omit AMSTAR2 criteria that are seldom applied in education (e.g., searching trial registries). The "comprehensive search" criteria assessed whether the search strategy was likely to find the majority of studies meeting inclusion criteria. The "clear inclusion/exclusion" criteria assessed the reproducibility of the systematic review's eligibility criteria. The "duplicate screening" and 'duplicate quality assessment" criteria assessed whether tasks prone to consequential errors or judgement are done in duplicate. The "presentation of each included study" criteria assessed the transparency of the results from the systematic review, used to inform conclusions. Finally, the “assessment of heterogeneity” and “assessment of publication bias” criteria assessed how well the reviews accounted for these two sources of uncertainty in conclusions. We did not tally different criteria into a summative score because each criteria carry different risks that should not be weighted equally (Higgins et al., 2011). We instead presented the results of each criteria for each review.

Summary Measures and Synthesis of Results

The primary outcome was the pooled effect size of the multimedia design principle on learning. We also extracted effects on transfer and on other proximal learning outcomes, like cognitive load or learning speed. We conducted a meta-meta-analysis to allow for comparison between the pooled effect sizes using the *metasem* package (Cheung, 2014a) and *msemtools* (Conigrave, 2019) in *R* (R Core Team, 2020) to construct a maximum likelihood estimate of the pooled meta-meta-analytic effect size using all available information (Cheung, 2014b). We used random-effects multilevel meta-analyses because these analyses are robust to effect sizes that are clustered (Moeyaert et al., 2016); in our review, effect sizes on the same design principle were not independent because the studies informing each effect were overlapping (e.g., one study on signaling might report both retention and transfer scores). That is, samples were sometimes dependent within design principles (because experiments sometimes reported multiple outcomes) but were independent between design principles (because experiments tested one design principle at a time). We chose random effects models to account for heterogeneity in effects from different design principles. Our analyses, therefore, met all the assumptions for a second-order meta-analysis: random effects between and within meta-analyses, inverse sampling weights, and between-sample independence (Hennessy et al., 2019; Schmidt & Oh, 2013). These models are the current best practice for pooling meta-analytic findings, accounting for both within- and between-review variance (Hennessy et al., 2019; Schmidt & Oh, 2013).

First, we converted all reported, pooled effect sizes to a common metric, Hedges’s *g*, which corrects for biases in small sample sizes (Hedges, 1981). We performed calculations using the *compute.es* package (Del Re, 2020). If reviews failed to report the number of participants, we used the median number of participants per study across all reviews, then used the number of studies to estimate the pooled sample size. Because multiple systematic reviews on the same topic have overlapping primary studies, we followed the Cochrane Collaboration advice to select the largest of those reviews for each multimedia design principle to avoid double-counting older primary studies (Pollock et al., 2018). Given reviews each assessed different design principles, there was no overlap between reviews and therefore no correction required for overlap (e.g., Hennessy & Johnson, 2020). Where reviews reported multiple effect sizes for learning (e.g., retention and transfer) we calculated pooled effect sizes, nesting effects within reviews. As described above, we used multilevel random effects meta-analysis, clustering effects within interventions to control for covariance between multiple measures of the same outcome. We then conducted moderation analyses for the most commonly reported moderators:

1. The type of learning outcome (retention test, transfer test, learning [general; for example, pooled effect of retention and transfer])
2. The pacing of the media (self-paced, like a website, or system-paced, like a TV show)
3. The prior knowledge of the learner (low or high, as coded by reviewers)
4. The level of "element interactivity" (complexity of the multimedia, with more complex media having more interactions; low or high, as coded by reviewers)
5. The education level of the learner (school or university/adult)
6. The presentation mode of the media (e.g., online or via paper)
7. The subject domain (science, mathematics, humanities, etc.)

For moderation analyses, we calculated the percentage of variance in outcomes explained by the moderator using R^2 (Cheung 2014a). To assess whether this change in variance was significant, we used the likelihood ratio test (Cheung, 2014a, 2014b).

Results

The PRISMA flow diagram (Figure 1) outlines the study selection process. Through database searching, we identified 3,982 possible reviews. After removing duplicates, two reviewers independently screened each of 1,526 titles and abstracts, three of which were found through forward and backward citation searching. After screening titles and abstracts, two reviewers independently reviewed each of 102 full-text articles. Consensus reasons for exclusion are documented in Online Supplementary File 2. Following this full-text screening, we included 29 systematic reviews. These reviews included a combined total of 1,189 studies and 79,415 participants. The characteristics of each included review are available in Table 1.

Effect of Design Principles on Learning

We dual-extracted 674 pooled effect sizes (Online Supplementary File 3) from 23 reviews that presented quantitative data. To avoid double-counting any primary studies and to maintain independence between reviews, we chose the most comprehensive review when conducting each analysis.

To assess whether retention and transfer measures could be validly meta-meta-analyzed, we first assessed whether the specific outcome measure moderated the effects of multimedia design principles. From all included reviews, we extracted retention, transfer, and mixed retention/transfer measures, even if the review reported a pooled estimate (of both retention and transfer). To avoid double-counting older primary studies, we selected the biggest review for each design principle that reported these effects. The specific learning outcome (retention vs. transfer) did not moderate the effects of these design principles ($N_{PooledEffects} = 23$, $R_2^2 = 0.12, p = 0.51$). The effects of each design principle, separated by outcome, are displayed in Figure 2. The pooled effects of multimedia design principles were almost identical for measures of retention ($g = 0.42$, 95% CI [0.32, 0.52]) and transfer ($g = 0.43$, 95% CI [0.31, 0.54]).

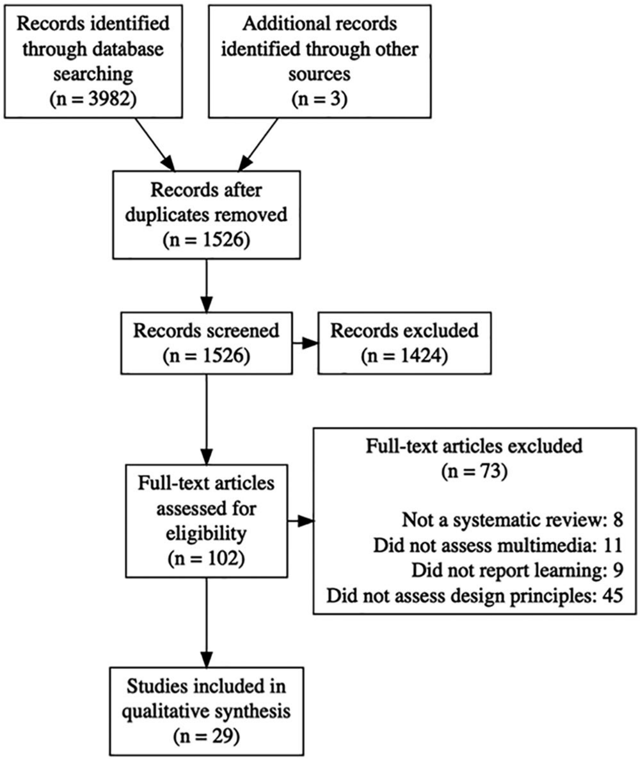

FIGURE 1. *PRISMA flow diagram.*

Because the specific learning outcome did not moderate effects, we calculated pooled retention and transfer measures where they were not available. To assess the relative influence of each design principle, the 11 largest reviews pooled a combined 808 effect sizes from 66,553 participants. The average effect of multimedia design principles on learning was moderate and statistically significant ($g = 0.38$, 95% CI [0.27, 0.49], $k = 808$). The specific design principle explained most of the variance in learning outcomes ($N_{PooledEffects} = 16$, $R_2^2 = 1.00$, $p <$ 0.001). As shown in Figure 3, some design principles had much larger effects than others. Specifically, reviews found moderate-large benefits from captioning second-language videos, integrated designs, the modality effect, and signaling. Effects were weak and positive for verbal redundancy and pedagogical agents. Most other effects were small to moderate. The anticipated effects of each design principle are explored in detail below.

Effects of Design Principles on Cognitive Load

Overall, multimedia design interventions led to small improvements in management of cognitive load ($g = 0.22$, 95% CI [0.04, 0.40], $k = 68$). As with effects

TABLE 1

Characteristics of included reviews, including scope, yield, and key findings

Author, year	Review focus	Review scope	*n*	*k*	Summary of findings and conclusions
Captioning second-language video					
Montero Perez et al., 2013	Captioning of video for second-language learning vs. no captions	Experimental studies comparing video with and without captions. Ten key words in 6 databases, 10 language journals and manual search of references.	NR	18	Captioning led to a significantly increased effect on comprehension ($g = 0.99$) and vocabulary learning ($g = 0.87$). Captioning enhances the ability to encode the information, increasing recall specifically, listening comprehension, and vocabulary acquisition. The findings support the dual-coding theory which states that the use of bimodal encoding leads to increased recall.
Integrativity/contiguity: *keeping relevant information close together in space (spatial contiguity) and time (temporal contiguity)*					
Ginns, 2006	Contiguity for cognitive load and learning	Randomized experiments that compared split vs. contiguous media. Four keywords in 2 databases and manually searched reference lists.	2,375	50	Increasing contiguity increased learning (overall $d = 0.85$). Effects were significant for both spatial ($d = 0.72$) and temporal contiguity ($d = 0.78$). Contiguity was more important for materials where learners needed to connect many dots (high element interactivity; $d = 0.78$) compared with simple materials ($d = 0.28$). Integrating information across time or space overworks cognitive load, thus negatively affects learning ability.
Schroeder & Cenkci, 2018	Spatial contiguity for learning	Experimental studies on spatial contiguity vs. split attention. Four key words in 5 databases; manual search of Mayer and Fiorella (2014)and Ginns (2006).	2,426	58	Integrated learning designs enhanced performance ($g = 0.63$) when compared to spatially distant designs. Integrated instructional designs appear to decrease the extraneous processing required, increasing learning outcomes.
Schroeder & Cenkci, 2020	Mediators of spatial contiguity effects on learning	Published, English experiments spatial contiguity multimedia with split-attention media. Needed to report measures of cognitive load. Used 4 keywords across 5 databases.	NR	41	Interpreting effect sizes suggest that spatial contiguity decreases mental effort during testing, perceived task difficulty, intrinsic load, extrinsic load, and increased the capacity for germane load. The reviewers used a vote-count to make conclusions and argued the final result is most "consistent" but did not conduct a meta-analysis. Based on that consistency, they argued integrated designs do not reduce extraneous load but instead facilitate germane load (integrative cognitive processing).
Signaling: *cues that direct attention to important content, like colors, arrows, animations, or text*					
Alpizar et al., 2020	Signaling for learning	Randomized trials of multimedia learning with and without signaling/cuing. Three keywords across 7 databases	2,726	29	Signaling increased learning ($d = 0.38$), particularly from multimedia ($d = 0.63$), for low-prior knowledge learners ($d = 0.47$), and particularly when using dynamic (e.g., moving/animated) signals ($d = 0.43$).
Richter et al., 2016	Signaling for learning	Experimental studies of multimedia with and without supporting signals. Seven keywords in 5 databases; citation searching.	2,464	27	Signaling increased learning ($r = 0.17$). Participants with low prior knowledge got more benefit from signaling ($r = 0.19$) than those with high prior knowledge ($r = -0.08$). Signaling highlights important information for learners, without altering the content, decreasing cognitive load from extraneous materials.

(continued)

TABLE 1 (CONTINUED)

Author, year	Review focus	Review scope	n	k	Summary of findings and conclusions
S. Schneider et al., 2018	Signaling for learning	Experimental studies of multimedia with and without supporting signals. Nineteen keywords in 5 databases; citation searching.	12,201	103	Signaling increased transfer (g = 0.53), retention (g = 0.33), and reduced cognitive load (g = 0.25). It increased learning-relevant fixations (g = 0.39) and motivation (g = 0.13). This effect was relatively consistent across diverse educational levels and disciplines, in both high- and low-prior knowledge of learners, and within both system- and learner-paced presentations. Text signaling generally improved retention and transfer to a greater extent than graphic signaling.
Xie et al., 2016	Signaling or contiguity for learning	Experimental studies testing the effects of signaling or "spatiotemporal cues" (likely to be "contiguity") on any outcome. Nine search terms across 6 databases.	3,910	43	Primary analyses appear to present results of both temporal/spatial contiguity and signaling combined. Results show large benefits for spatiotemporal cues (spatial or temporal contiguity; g = 0.73), and small-moderate benefits of signaling (physical cues; g = 0.31). Overall, design strategies led to increases in retention (g = 0.53), transfer (g = 0.36), and increased the learner's fixation time (g = 0.50) and the number of fixations (g = 0.70).
Xie et al., 2017	Signaling for learning	Experimental studies on multimedia with vs. without cues. Nine keywords in 7 databases; citation searching.	3,597	32	Signaling reduced cognitive load (d = −0.11), increased retention (d = 0.27) and transfer (d = 0.34). Signaling appears to reduce extraneous cognitive load.
Modality effect: *presenting words via speech instead of printed text (usually in combination with graphics)*					
Ginns, 2005	Modality effect	Experimental studies that compare audio to textual presentation of information. Three keywords over 3 databases and manual search of reference lists.	1,887	43	Replacing text with audio improved knowledge transfer (d = 0.76) and alleviated cognitive load (d = 0.62). This was only true in system-paced presentations (d = 0.93). Otherwise, effects appeared too generalized across various types of presentations, educational levels, and disciplines. With textual-only information the learner must process the entire material using visual cache. In contrast, using an audio-visual format reduced cognitive load because learners can use the visual cache for graphics and the phonological loop for words.
Reinwein, 2012	Modality effect (inc. controls for publication bias)	Study uses papers from Ginns's (2005) systematic review and a few additional papers determined by a nonsystematic search.	3,316	86	Modality effect was robust but small when accounting for publication bias. Effect is eliminated when students can self-pace, likely because self-paced reading reduces the chance of overload. It is stronger for short passages, likely because when students read long-passages, they can self-pace. Effects are bigger in dynamic visualizations, again probably because students can read static visualizations at their own pace.
Pleasant colors/anthropomorphics: *for example, faces on inanimate objects, like enzymes*					
Brom et al., 2018	Pleasant colors or anthropomorphics for learning	Publicly available (quasi-) experimental studies that manipulated anthropomorphisms and/or pleasant/aesthetic colors. Outcomes also included affective/motivation states. Twenty-one key words in 4 databases; manually searched reference lists of 15 studies.	2,924	33	Anthropomorphism and adding pleasant colors improved retention (d = 0.39), comprehension (d = 0.32), and transfer (d = 0.33). Anthropomorphism and adding pleasant colors had a more reliable effect for intrinsic motivation (d = 0.26) than liking/enjoyment (d = 0.11) or generalized positive affect (d = 0.11); materials were perceived to be less difficult (d = −0.21). Emotional design is a useful tool to enhance learning, and appears more beneficial for younger learners.

(continued)

TABLE 1 (CONTINUED)

Author, year	Review focus	Review scope	*n*	*k*	Summary of findings and conclusions
Segmenting: *breaking longer learning activities into shorter, learner-paced chunks*					
Rey et al., 2019	Segmenting for retention and transfer	Experimental, English-language studies on segmented vs. unsegmented materials. Five keywords in 5 databases, plus citation searching.	NR	56	Segmenting enhances retention ($g = 0.32$), transfer ($g = 0.36$), and cognitive load ($g = 0.23$), compared to nonsegmented multimedia materials. Segmenting allows learners to chunk information more coherently, providing optimum time to process the information at pace the individual needs.
Personalization: *conversational rather than formal language*					
Ginns et al., 2013	Conversational language for learning, interest, friendliness, and difficulty	(Quasi-)experimental studies comparing conversational vs. formal language in media where a learning outcome was reported. Effectively 3 key words (+6 duplicates) across 3 databases.	3,312	22	Conversational language increased both retention ($g = 0.30$) and transfer ($g = 0.54$). It also increased perceived friendliness ($g = 0.46$) and effective cognitive processing ($g = 0.62$). It did not improve student interest ($g = 0.15$). This was consistent for both "polite" and "personal" language. There was some evidence of publication bias.
Removing seductive details: *interesting but irrelevant digressions or images*					
Rey, 2012	Removing seductive details for retention and transfer	Experimental studies with vs. without seductive detailing. Two key words over 2 databases.	5,169	39	Removing seductive details improved retention ($d = 0.30$) and transfer ($d = 0.48$). Materials that incorporate seductive details reduce available cognitive load by taking up space that could otherwise be directed to the core materials.
Sundararajan & Adesope, 2020	Removing seductive details for learning	Publicly available experiments with seductive details vs. control condition. Three keywords in 7 databases and alerts set using search terms in Google Scholar.	7,521	58	Removing seductive details improved learning ($g = 0.33$), particularly when those details were persistently on-screen ($g = 0.43$) and for those without prior experience in the subject ($g = 0.52$).
Animation					
Berney & Bétrancourt, 2016	Animated vs. static graphics	Publicly available studies (inc. PhD theses) comparing animated and static graphics. Used 11 key words in 5 databases.	7,036	61	Animations yield better learning outcomes than using static graphics ($g = 0.23$). Effects were larger for factual knowledge ($g = 0.336$, $p = .001$) than conceptual knowledge ($g = 0.162$, $p = .006$). They appeared to work better when combined with audio commentary, or when learners have little control over the pacing.
Höffler & Leutner, 2007	Animated vs. static graphics	Experimental studies comparing animated and static visualizations, with no or minimal interactivity. Ten key words in 5 databases and citation searching.	3,564	26	Noninteractive animations were more effective at facilitating learning than static images ($d = 0.37$). Animations with representation information ($d = 0.40$) are better than static pictures; however, decorational animations have no significant learning benefit ($d = 0.05$). Animations worked better in procedural and motor learning tasks over factual knowledge settings. Computer generated vs. video animations seem similarly efficacious.
Kaushal & Panda, 2019	Animated vs. static graphics, moderated by student spatial ability and prior knowledge	Experimental studies reported moderating effects of individual differences (spatial ability and prior knowledge) on animation for learning. Unclear keywords in 2 databases.	NR	37	Animation improved learning for both high spatial ability learners ($d = 0.34$) and low spatial ability learners ($d = 0.58$). Low prior knowledge learners did not benefit from animations ($d = -0.16$). There was a positive effect for high prior knowledge learners ($d = 0.49$). The review did not provide a theoretical explanation for the findings.

(continued)

TABLE 1 (CONTINUED)

Author, year	Review focus	Review scope	*n*	*k*	Summary of findings and conclusions
Pedagogical agents: *on-screen anthropomorphized characters or human faces that communicate with learners*					
Castro-Alonso et al., 2021	Pedagogical agents for learning	Experimental studies with and without pedagogical gesturing agents (written in English, published between 2012 and 2019). Fourteen keywords across 5 databases.	2,104	32	Pedagogical agents led to small improvements in learning (g = 0.20; 95% CI [0.07, 0.34]). Effects appeared stronger for 2D agents than 3D agents, presumably because 3D was a seductive detail.
Davis, 2018	Pedagogical agents as methods of signaling	Experimental studies with and without pedagogical gesturing agents. Four keywords in 10 databases and manually searched previous reviews and meta-analyses.	3,841	20	Gesturing pedagogical agents improved transfer (g = 0.39) and retention (g = 0.28). There were minimal effects on cognitive load (g = 0.13). Benefits were relatively consistent across participant age groups (i.e., undergraduate and middle school students) and agent characteristics.
Kim, 2005	Multimedia design principles that moderate effects of pedagogical agents	Experimental studies on multimedia design principles in studies that include pedagogical agents; unclear search terms across 6 databases.	NR	12	Pedagogical agents appeared to increase learning (pooled *d*s between 0.29 and 0.58) but there were no significant differences between agents that used text (d = 0.29), spoke (d = 0.53), were animated (d = 0.52), or had "social meaning" (e.g., had a gender or ethnicity; d = 0.58). As a result, it is not clear how to optimize the benefits from pedagogical agents.
Schroeder et al., 2013	Pedagogical agents for learning	Publicly available experiments with vs. without pedagogical agents. The agent must have been informational, not just decorative. Use 2 key words in 8 databases and manually searched the 2011 meeting of American Education Association.	3,088	43	Pedagogical agents significantly benefit learning (g = 0.19). This was higher for mathematics/science than other subjects, and higher among school students (g = 0.56) than university students, but did not depend on how anthropomorphized the agent was. Agents using text (g = 0.51) were more effective than those using voice (g = 0.12). Reviewers argued that young people are more influenced by social agency: the feeling of social interaction between the learner and the agent. They argued data did not support the split-attention principle because learning was not impeded by the agent's appearance on the screen.
Verbal redundancy: *adding words on screen when presenting audio (e.g., speech)*					
Adesope & Nesbit, 2012	Verbal redundancy	Publicly available experiments (inc. cohorts) that compared exclusively written or spoken presentations with those of written+spoken presentations. Used 2 key words in 5 databases.	3,452	33	Verbal redundancy enhanced learning outcomes (g = 0.15), but only when adding visual presentation of key words to audio presentations (g = 0.29), not when adding audio to text presentations (g = −0.04). Benefits were proportional to study and treatment duration, and consistent in both learner- and system-paced presentations, classroom and laboratory settings, and between participants of various levels of education and language fluency.

(continued)

TABLE 1 (CONTINUED)

Author, year	Review focus	Review scope	n	k	Summary of findings and conclusions
Cognitive conflict: *eliciting knowledge-gaps; identifying and refuting misconceptions*					
Rahim et al., 2015	Cognitive conflict added to multimedia	Any study that used cognitive conflict in multimedia learning. Three keywords in 4 databases.	NR	7	Cognitive conflict appeared to be helpful, and appeared best implemented by the teacher facilitating a discussion among students about their current understanding, and then providing interesting, relevant, relatable information that challenges the students' misconception. Review authors argue conflict is more successful when it attracts attention by identifying with students, evoke curiosity as a motivator, and lightly/comfortably correcting misconceptions.
All/mixed multimedia design principles					
Alemdag & Cagiltay, 2018	Eye-movement as a mediator of multimedia's effect on learning	Studies assessing the mediating effects of eye movement on learning from multimedia design effects; 6 keywords across 4 databases	NR	58	Five multimedia learning principles were supported by eye-tracking studies: signaling led to more attention on important elements; omitting text (applying the modality principle) meant people attended more to images, and more attention to images improved learning; seductive details took attention from key words and pictures, which in turn inhibited learning; spatial contiguity led to more focus on key visuals, and better integration between visuals, both of which generally improved learning; learners directed greater attention to animated (dynamic) visuals than static ones; finally, greater attention was allocated to text than to pictures, but fixation on visuals led to better learning.
Amaka & Goeman, 2017	Multimedia attributes in blended and online learning	Peer-reviewed experiments: online or blended learning vs. control. Searched 18 databases, with unreported keywords; manual search of reference lists.	2,861	24	Effects of blended/online learning appeared stronger for rich, interactive, flexible, easy to navigate interventions, particularly with the instructor being present.
Cromley & Lawrence, 2018	Multimedia simulation for deep learning (e.g., transfer)	Peer-reviewed experiments: stimulation multimedia for maths or science vs. control. Eleven keywords in 2 databases; manual search of reference list and contacted specific authors.	7,498	48	Simulations enhance transfer, compared to traditional teaching ($d = 0.57$) and active control (not theoretically driven simulations; $d = 0.64$). The review did not provide any theoretical explanation for findings.
Kang, 2019	Multimedia for second-language learning	Empirical studies of multimedia with listening comprehension as the dependent variable. Five keywords in 3 databases and 2 keywords targeting 13 journals.	NR	50	Multimedia has a moderate-large benefit on second-language learning, specifically when captioned ($d = 1.69$), when self-paced or slowed ($d = 0.67$), and when video (vs. audio alone; $d = 0.34$).

Note. Effect sizes reported here are those directly extracted from the reviews (unconverted to a common metric); N = total number of participants summarized by the review; k = the number of effects; NR = not reported; unless otherwise indicated in *Review Scope*, reviews tended to also require that studies reported quantitative measures of learning.

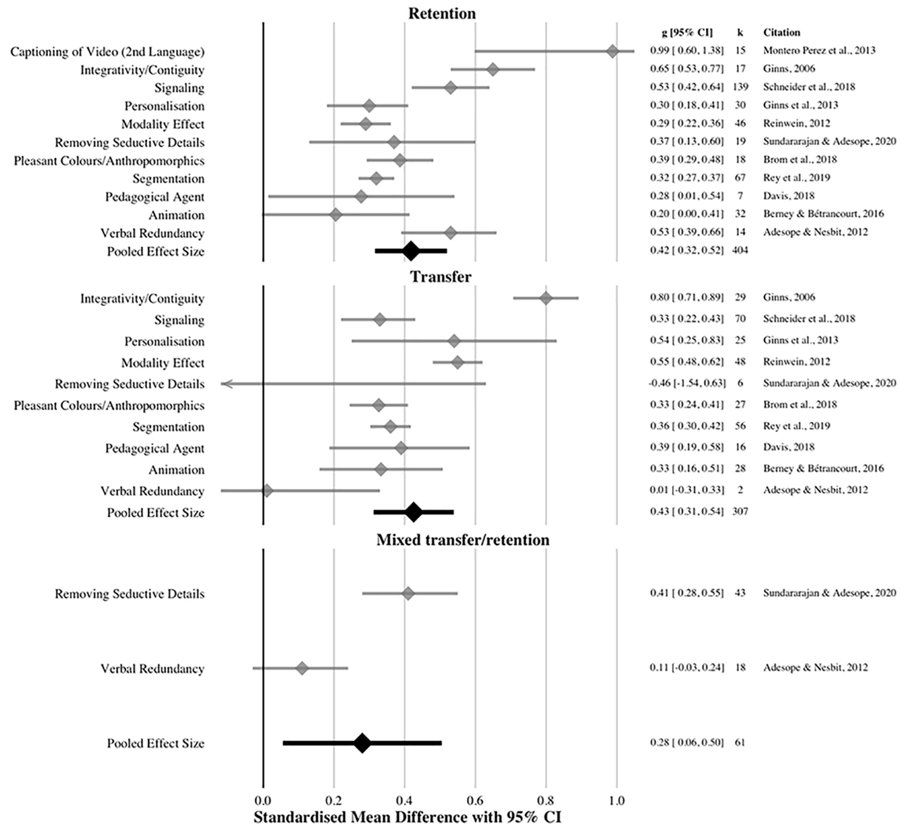

FIGURE 2. *Effects of each design principle on learning, separated by outcome.*

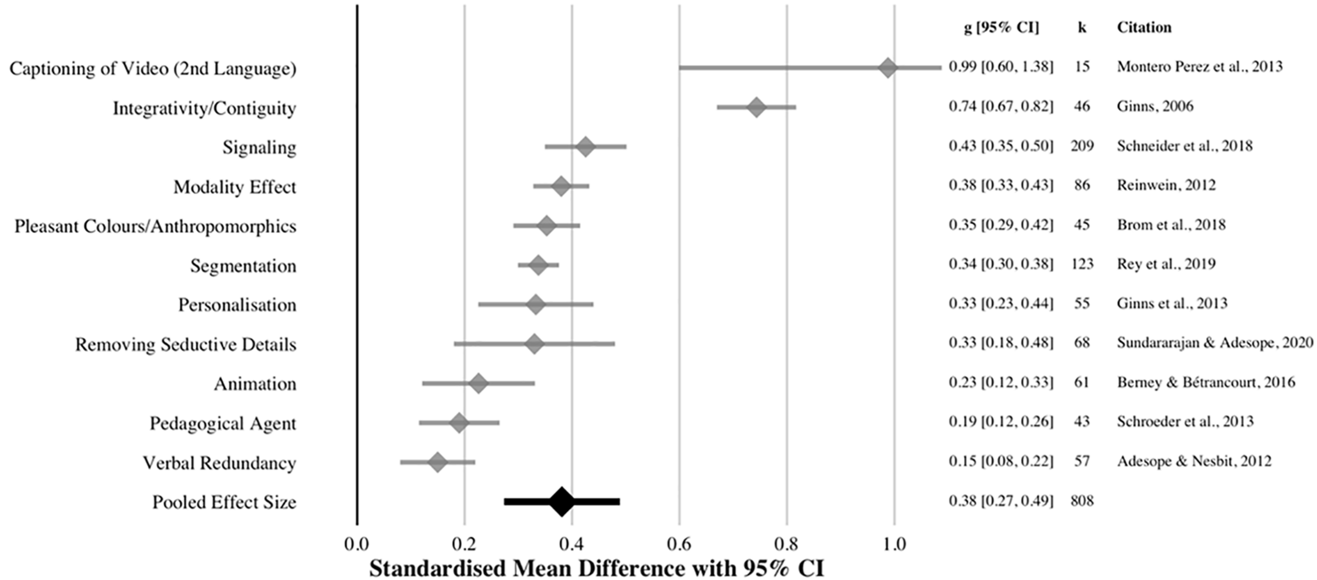

FIGURE 3. *Pooled, meta-meta-analytic estimates of intervention effects on learning.*

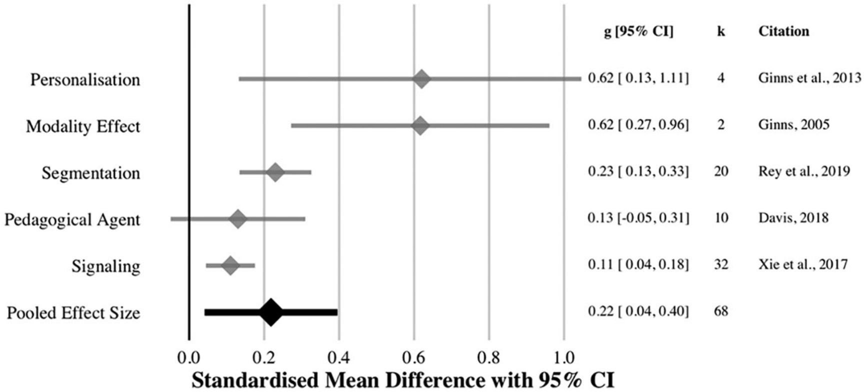

FIGURE 4. *Pooled, meta-meta-analytic estimates of intervention effects on cognitive load.*

on learning, effects on cognitive load were moderated by the multimedia design principle ($N_{PooledEffects} = 5$, $R_2^2 = 1.00$, $p = 0.01$). As outlined in Figure 4, a small number of primary studies found large effects personalization ($g = 0.62$, 95% CI [0.13, 1.11], $k = 4$; Ginns et al., 2013) and the modality effect ($g = 0.62$, 95% CI [0.27, 0.96], $k = 2$; Ginns, 2005). Larger reviews found small effects for for segmentation ($g = 0.23$, 95% CI [0.13, 0.33], $k = 20$; Rey et al., 2019), signaling ($g = 0.11$, 95% CI [0.04, 0.18], $k = 32$; Xie et al., 2017), and pedagogical agents ($g = 0.13$, 95% CI [−0.05, 0.31], $k = 10$; Davis, 2018).

Synthesis of Review Findings for Each Design Principle

Effects of Captioning Videos

There were large effects were for captioning videos on second-language learning ($g = 0.99$, 95% CI [0.60, 1.38], $k = 15$; Montero Perez et al., 2013). While two reviews assessed this phenomenon (Kang, 2019; Montero Perez et al., 2013), the review by Montero Perez et al. (2013) was rated more favorably on four out of seven quality assessment criteria. This review found captioning led to large benefits for both comprehension ($g = 0.99$, 95% CI [0.60, 1.38], $k = 15$; Montero Perez et al., 2013) and vocabulary acquisition ($g = 0.87$, 95% CI [0.58, 1.15], $k = 10$; Montero Perez et al., 2013).

Effects of Contiguity

Three reviews explored the benefits of contiguity (a.k.a., integration; Ginns, 2006; Schroeder & Cenkci, 2018; Schroeder & Cenkci, 2020). Overall, there were large benefits of having materials presented contiguously ($g = 0.74$, 95% CI [0.67, 0.82], $k = 46$; Ginns, 2006).

Spatial contiguity is where related material is put close together rather than in different parts of the media (e.g., labels *on* a diagram, rather than in a legend). The larger of these reviews found that spatial contiguity led to significant improvements in learning compared with spatially distant designs ($g = 0.63$, 95% CI

[0.55, 0.71], $k = 58$; Schroeder & Cenkci, 2018). The spatial contiguity principle was also validated by eye-tracking studies (Alemdag & Cagiltay, 2018). These showed contiguity led to more focused attention and better integration between visuals and text, both of which improved learning (Alemdag & Cagiltay, 2018). Spatial contiguity appears to improve management of cognitive load, reducing intrinsic and extrinsic load and increasing the opportunity for germane load (Schroeder & Cenkci, 2020). Although Schroeder and Cenkci (2020) described the increases in germane load as the most robust, they did not conduct a meta-analysis to establish the relative pooled effect sizes.

Temporal contiguity, like spatial contiguity, refers to the benefits of presenting connected information together, but in time, rather than in space (e.g., presenting one bullet-point at a time as the presenter talks to each rather than all bullets at once). Moderation analyses from reviews have found strong benefits of this design principle too ($g = 0.78$, 95% CI [0.64, 0.92], $k = 13$; Ginns, 2006).

Effects of Signaling

Five reviews have explored the benefits of signaling (Alpizar et al., 2020; Richter et al., 2016; S. Schneider et al., 2018; Xie et al., 2016; Xie et al., 2017). Signals are cues that help the learner know where to direct their attention (e.g., arrows, highlighted text, a laser pointer). Signaling has been extensively studied and was shown to increase learning ($g = 0.43$, 95% CI [0.35, 0.50], $k = 209$; S. Schneider et al., 2018) and reduce cognitive load ($g = 0.25$, 95% CI [0.04, 0.45], $k = 27$; S. Schneider et al., 2018). Most other reviews found similar results. Signaling increased learning-relevant fixations ($g = 0.39$, 95% CI [0.09, 0.68], $k = 14$; S. Schneider et al., 2018) and motivation ($g = 0.13$, 95% CI [0.04, 0.22], $k = 13$; S. Schneider et al., 2018). It was also validated by eye-tracking studies that showed these elements helped learners focus on key details, which then improved learning (Alemdag & Cagiltay, 2018).

Effects of Changing Modality

Students learn better when hearing text alongside visuals (e.g., in a textbook) than when reading text alongside visuals (e.g., in a textbook). Two reviews have explored this phenomenon (“modality effect”; Ginns, 2005; Reinwein, 2012). The larger of these reviews found moderate benefits on learning ($g = 0.38$, 95% CI [0.33, 0.43], $k = 86$; Reinwein, 2012). Effects were smaller, but still significant, when statistically accounting for publication bias ($g = 0.20$, 95% CI [0.15, 0.25], $k = 86$; Reinwein, 2012). These benefits were also validated by eye-tracking studies (Alemdag & Cagiltay, 2018). When text and images were displayed, learners preferentially attended to text, but did so at the expense of attending to images, which reduced their learning (Alemdag & Cagiltay, 2018). Omitting text for spoken words meant people attended more to images, which increased their learning (Alemdag & Cagiltay, 2018).

Effects of Pleasant Colors or Anthropomorphics

One review explored the benefits of using either pleasant colors or anthropomorphics in multimedia (Brom et al., 2018). Anthropomorphics are human features that an instructor might add to a graphics, like putting an angry face on an image of a virus entering the body. This review pooled pleasant colors and anthropomorphics because they may operate by similar mechanisms (i.e., increased cognitive engagement; Brom et al., 2018); however, as a result, we could not disentangle their distinct effects. Pooled together, these interventions lead to moderate improvements in learning (g = 0.35, 95% CI [0.29, 0.42], k = 45; Brom et al., 2018). Anthropomorphism and colors had small benefits for intrinsic motivation (g = 0.25, 95% CI [0.17, 0.34], k = 23; Brom et al., 2018) and they made materials feel less difficult (g = 0.21, 95% CI [0.08, 0.33], k = 14; Brom et al., 2018).

Effects of Segmentation

Segmenting materials involves breaking the lesson or video into meaningful chunks. For example, rather than presenting one long lecture on multimedia design principles, segmenting would involve a series of smaller lessons, one on each design principle. Rey et al. (2019) found segmenting had moderate benefits for learning (g = 0.34, 95% CI [0.30, 0.38], k = 123; Rey et al., 2019). It also reduced the cognitive load students experienced (g = 0.23, 95% CI [0.13, 0.33], k = 20; Rey et al., 2019). Students learned better when instructors segmented learning for students (g = 0.41, 95% CI [0.32, 0.50], k = 32; Rey et al., 2019)—as opposed to allowing students to segment for themselves (g = 0.20, 95% CI [0.11, 0.28], k = 32; Rey et al., 2019)—presumably because instructors can more easily ensure that the chunks are meaningful and appropriately sized. It took students longer to learn segmented content than when it was unsegmented (g = 0.92, 95% CI [0.82, 1.02], k = 19; Rey et al., 2019), suggesting they used time between segments to consolidate their learning.

Effects of Personalization

One review explored the benefits of personalization: changing language to be either simpler, more polite, or more related to the learner (Ginns et al., 2013). This review showed that all of these modifications were equally effective, and they increase learning (g = 0.33, 95% CI [0.23, 0.44], k = 55; Ginns et al., 2013), and reduced cognitive load (g = 0.62, 95% CI [0.13, 1.11], k = 4; Ginns et al., 2013).

Effects of Removing Seductive Details

Two reviews explored the benefits of removing seductive, irrelevant details (a.k.a., "coherence principle"; Rey, 2012; Sundararajan & Adesope, 2020). Removing these details has moderate benefits for learning (g = 0.33, 95% CI [0.18, 0.48], k = 68; Sundararajan & Adesope, 2020). Effects were bigger when seductive details were otherwise displayed persistently onscreen (g = 0.43, 95% CI [0.29, 0.57], k = 47; Sundararajan & Adesope, 2020) as opposed to being transient (g = 0.12, 95% CI [−0.33, 0.57], k = 18; Sundararajan & Adesope, 2020). Materials that incorporate seductive details appear to reduce available cognitive load by taking up space that could otherwise be directed to the core materials: eye-tracking studies that showed seductive details drew attention away from core information, and doing so consistently led to poorer learning (Alemdag & Cagiltay, 2018).

Effects of Animation

Three reviews compared the effects of animated, dynamic graphics, compared with static images (Berney & Bétrancourt, 2016; Höffler & Leutner, 2007; Kaushal & Panda, 2019). Berney and Bétrancourt (2016) conducted the most comprehensive of these reviews with 61 primary studies (vs. 26 primary studies, Höffler & Leutner, 2007; 22 primary studies, Kaushal & Panda, 2019). Their review found significant learning benefits (g = 0.23, 95% CI [0.12, 0.33]). Eye tracking studies found that learners directed more attention toward animated graphics than static ones (Alemdag & Cagiltay, 2018). Animations were only helpful when the animations communicated something meaningful, like how gears work (g = 0.40, 95% CI [0.34, 0.46], k = 59; Höffler & Leutner, 2007). Decorational animations had no learning benefits (g = −0.05, 95% CI [−0.17, 0.07], k = 17; Höffler & Leutner, 2007).

Effects of Pedagogical Agents

A pedagogical agent is a multimedia design feature in which there is an artificial "teacher" to guide the learner. Four reviews explored the benefits of these pedagogical agents (Castro-Alonso et al., 2021; Davis, 2018; Kim, 2005; Schroeder et al., 2013). Some reviewers argue that pedagogical agents are not seductive details because they increase learning (g = 0.19, 95% CI [0.12, 0.27], k = 43; Schroeder et al., 2013). But, when only using studies where the agent gestured toward what was important—that is, when the agent served a signaling function—effects of the agent were stronger (g = 0.28, 95% CI [0.01, 0.54], k = 7; Davis, 2018). Learners also preferred these agents to those that did not serve a signaling function (g = 0.44, 95% CI [0.08, 0.81], k = 8; Davis, 2018). Agents that were presented in three dimensions showed no benefits (g = 0.11, 95% CI [-0.06, 0.27], k = 21; Castro-Alonso et al., 2021), presumably because those agents are more distracting than those in two dimensions (g = 0.38, 95% CI [0.16, 0.60], k = 11; Castro-Alonso et al., 2021).

Effects of Verbal Redundancy

Verbal redundancy refers to the effects of communicating words both visually and aurally. For example, what happens when text is added to slides accompanying a lecture (adding text to audio) or when narration is added to written text (adding audio to text)? One review explored this phenomenon (Adesope & Nesbit, 2012). Verbal redundancy enhanced learning outcomes overall (g = 0.15, 95% CI [0.08, 0.22], k = 57; Adesope & Nesbit, 2012), but only when adding text to audio (g = 0.29, 95% CI [0.20, 0.39], k = 34; Adesope & Nesbit, 2012), not when adding audio to text (g = -0.04, 95% CI [-0.14, 0.06], k = 23; Adesope & Nesbit, 2012).

Other Design Principles

A number of other reviews explored multimedia principles that were either less extensively explored (by the extant literature), or less well defined (by reviewers). Rahim et al. (2015) described how cognitive conflict (eliciting knowledge-gaps; identifying and refuting misconceptions) may lead to increases in learning from multimedia (pooled effects not reported). Cromley and Lawrence (2018) argued that “theoretically driven” multimedia simulations were more effective than those which were not theoretically driven (pooled effects not reported). In addition to the effects of captioning (described earlier), Kang (2019) found that students learned second-languages better from video than audio alone ($g = 0.34$, 95% CI [0.23, 0.45], $k = 20$).

Robustness of Other Moderators From Included Reviews

Many reviews assessed both theoretical and emergent moderators of intervention effects. The key theoretical moderators were the pacing of the material (learner vs. system paced), the complexity of the materials (high vs. low element interactivity), and the prior knowledge of the learner (high vs. low). The pacing of the materials was a significant moderator of effects ($N_{PooledEffects} = 22$, $R_2^2 = 0.34$, $p = 0.02$). Design principles were more important when the multimedia is system-paced, like in a lecture ($g = 0.41$, 95% CI [0.33, 0.49]), than when learner-paced, like on a website ($g = 0.27$, 95% CI [0.19, 0.35]). Element interactivity does not refer to interactivity between the learner and the materials, but the complexity of the materials, with higher interactivity meaning more concepts for the learner to connect. The element interactivity moderated the effects of design principles ($N_{PooledEffects} = 4$, $R_2^2 = 1.00$, $p < 0.001$). Multimedia design principles were more important in complex media ($g = 0.70$, 95% CI [0.59, 0.81]) than more simple media ($g = 0.20$, 95% CI [0.02, 0.39]). The prior knowledge of the learner did not moderate effects ($N_{PooledEffects} = 30$, $R_2^2 = 0.14$, $p = 0.14$). Emergent moderators included the education level of participants (e.g., school vs. university), the media presentation (e.g., paper vs. computer), and the subject domain (e.g., STEM vs. social science). Neither the educational level of learner ($N_{PooledEffects} = 30$, $R_2^2 = 0.05$, $p = 0.95$), the presentation format ($N_{PooledEffects} = 8$, $R_2^2 = 0.98$, $p = 0.24$), nor the subject being studied ($R_2^2 = 60$, $R_2^2 = 0.15$, $p = 0.22$) consistently moderated results. Forest plots demonstrating all of these moderation effects are available in Online Supplementary Figures S1 through S6.

Quality Assessment of Included Reviews

The results of our quality assessment are described in Table 2. Most reviews presented a focused question (28/29) with clearly specified inclusion criteria (25/29). Most also appeared to conduct a comprehensive search for studies that addressed those questions (23/29). Only two explicitly reported duplicate study screening (Alemdag & Cagiltay, 2018; Berney & Bétrancourt, 2016). Only one review reported a formal quality assessment (Adesope & Nesbit, 2012), and that quality assessment was done in duplicate. Most reviews explicitly described each included study (23/29). Most reviews assessed publication bias (20/29) and possible sources of heterogeneity (22/29).

Discussion

With this overview of reviews, we aimed to collate evidence from systematic reviews on how to optimize multimedia for learning. We were able to find meta-analyses supporting many of the multimedia design principles hypothesized by cognitive load theory and the cognitive theory of multimedia learning. Some of these design principles produced robust, moderate-to-strong benefits for learning

TABLE 2

Results from quality assessment of the included reviews

First author, year	Focused question	Inclusion clearly specified	Comprehensive search	Duplicate screening	Duplicate quality assessment	All studies described	Publication bias assessed	Heterogeneity assessed
Adesope, 2012	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes
Alemdag, 2018	Yes	Yes	Yes	Yes	NA	No	No	No
Alpizar, 2020	Yes	Yes	Yes	No	NA	No	Yes	Yes
Amaka, 2017	No	No	Yes	Unclear	NA	Yes	No	No
Berney, 2016	Yes	Yes	Unclear	Yes	NA	Yes	Yes	Yes
Brom, 2018	Yes	Yes	Yes	Unclear	NA	Yes	Yes	Yes
Castro-Alonso, 2021	Yes	Yes	Yes	No	NA	Yes	Yes	Yes
Cromley, 2018	Yes	Yes	Yes	Unclear	NA	Yes	No	No
Davis, 2018	Yes	Yes	Yes	Unclear	NA	Yes	Yes	Yes
Ginns, 2005	Yes	Unclear	Unclear	No	NA	Yes	No	Yes
Ginns, 2006	Yes	Yes	Yes	No	NA	Yes	No	Yes
Ginns, 2013	Yes	Yes	Yes	Unclear	NA	Yes	Yes	Yes
Höffler, 2007	Yes	Yes	Yes	Unclear	NA	Yes	Yes	Yes
Kang, 2019	Yes	Unclear	Yes	No	NA	Yes	No	No
Kaushal, 2019	Yes	Yes	Yes	Unclear	NA	Yes	No	No
Kim, 2005	Yes	Yes	No	No	NA	Unclear	Yes	Yes
Montero Perez, 2013	Yes	Yes	Yes	Unclear	NA	Yes	Yes	Yes
Rahim, 2015	Yes	Yes	No	Unclear	NA	Yes	No	No
Reinwein, 2012	Yes	Yes	No	No	NA	Yes	Yes	Yes
Rey, 2012	Yes	Unclear	Unclear	Unclear	NA	Yes	Yes	Yes
Rey, 2019	Yes	Yes	Yes	Unclear	NA	Yes	Yes	Yes
Richter, 2016	Yes	Yes	Yes	Unclear	NA	Yes	Yes	Yes
Schneider, 2018	Yes	Yes	Yes	Unclear	NA	Yes	Yes	Yes
Schroeder, 2013	Yes	Yes	Yes	Unclear	NA	Yes	Yes	Yes
Schroeder, 2018	Yes	Yes	Yes	No	NA	No	Yes	Yes
Schroeder, 2020	Yes	Yes	Yes	No	NA	No	No	No
Sundararajan, 2020	Yes	Yes	Yes	Unclear	NA	No	Yes	Yes
Xie, 2016	Yes	Yes	Yes	Unclear	NA	Yes	Yes	Yes
Xie, 2017	Yes	Yes	Yes	Unclear	NA	Yes	Yes	Yes

Note. NA = not applicable because review did not conduct quality assessment of included studies.

(e.g., contiguity, signaling). As hypothesized, interventions also led to better management of cognitive load (e.g., signaling, modality effect). We used meta-meta-analyses to assess the reproducibility of three hypothesized moderators of design principles (pacing, prior-knowledge, and element interactivity). We found multimedia design principles were more important for complex materials (i.e., high element interactivity) than simple ones, and more important in system-paced environments than self-paced ones. Finally, we found robust support for multimedia design features that were seldom highlighted by previous theoretical papers (e.g., animation, anthropomorphics, captioning) and some emerging design principles that do not yet have meta-analytic support (e.g., cognitive conflict).

Findings Support Cognitive Load Theory and the Cognitive Theory of Multimedia Learning

Some multimedia design principles rely most heavily on the dual-channel assumption of multimedia learning. That is, the amount of information is relatively constant, but the media more effectively leverages both auditory and visual channels. For example, captioning second-language videos had a strong positive effect, arguably due to the way it facilitates encoding via both visual and auditory pathways (Montero Perez et al., 2013). Similarly, we found moderate, robust benefits from implementing the modality principle (replacing text + images with voice + images; Reinwein, 2012). With this principle, the amount of information is relatively constant, which is why effects might be lower than for captioning, where information is added. Similarly, we saw small benefits from adding animations (Berney & Bétrancourt, 2016) and only if they communicated meaning (Höffler & Leutner, 2007). Again, the animations provided benefit only if they added information beyond the static image, for instance, by helping learners understand how components of the image move and interact (Berney & Bétrancourt, 2016; Höffler & Leutner, 2007).

For these design principles, people learn better because adding richness to the visual stream leverages both channels of information processing (Mayer, 2008, 2009). Dual-channel presentations (voice + images) also allow information to be presented for easy integration; that is, it reduces the extraneous load required to find connections between interacting elements (Mayer & Moreno, 2010; Sweller, 2010). It is easier for a learner to receive a voice and image simultaneously than to read text and view an image simultaneously (i.e., temporal contiguity is easier; Schroeder & Cenkci, 2018), because in the latter, the learner must scan between related elements on the page. It is easier for them to hear how gears move while watching them interact, than to see a static image and have to imagine them turning. The importance of design principles like these might help explain why media that typically use voice *and* images (e.g., video) do better than media that do not (e.g., textbooks; Noetel et al., 2021). The positive effect of redundancy also shows the importance of leveraging both channels. Where cognitive load theorists often presume that both reading and hearing words adds extraneous load (Mayer, 2008; Sweller et al., 1998), our meta-review found small positive benefits of verbal redundancy (Adesope & Nesbit, 2012). The increased cognitive load of redundant text appears to be offset by the benefit of using both channels. Combining the strong benefits from captioning with the modest benefits of verbal redundancy, it may be that text on the screen is intrinsic, not extraneous, when it *facilitates* element interactivity aligned to learning objectives (Sweller, 2010). In some cases, like learning second languages from videos, connecting the caption to speech is tightly aligned with the learning objectives. In many other cases, like learning the jargon associated with a concept, having the key words duplicated visually helps connect those words with the spoken description of the concept. It may be only in cases of very bad design—so-called death by PowerPoint—where verbal redundancy is destructive: Where the effort required to connect the speech to written text is so high that it leads to cognitive overload.

Interventions that reduced extraneous load were robustly useful for improving learning. The strategies designed to direct attention toward important concepts all demonstrated similar effects. Highlighting important information somewhat reduced extraneous cognitive load (Xie et al., 2017) and led to moderate increases in learning (S. Schneider et al., 2018). Removing seductive details increased learning with similar effect sizes (Sundararajan & Adesope, 2020). Pedagogical agents were often seductive details (Castro-Alonso et al., 2021) and seldom reduced cognitive load (Davis, 2018), but modestly increased learning (Schroeder et al., 2013), likely because they often serve a signaling function (Davis, 2018). When used well, all of these strategies reduce extraneous cognitive load by focusing attention toward elements inherent in learning and away from elements that add extraneous load. As a result, it is consistent with theory that all strategies would demonstrate similar effect sizes.

Strategies to make intrinsic load more manageable consistently increased learning, also with similar effects. Segmenting modestly reduced cognitive load and led to moderate increases in learning (Rey et al., 2019). Using simpler, more informal language increased learning by the same amount, and reduced cognitive load (Ginns et al., 2013). Across design principles, effects were larger when multimedia were system-paced than when self-paced (e.g., Reinwein, 2012; Rey, 2012), because when self-paced, learners can more easily use their own strategies to manage high levels of cognitive load. Overall, the results of our meta-review support the notion that providing learners with ways of managing the intrinsic load (i.e., allowing self-pacing, segmenting materials, using personalized language) increased learning.

Our moderation analyses generally found support for the role of element interactivity in cognitive load (Sweller, 2010). The benefits of these design principles were usually larger for complex materials because—compared with simple materials—they were more likely to cause cognitive overload without good design (e.g., Ginns, 2005, 2006). That is, multimedia design is most important for concepts with high element interactivity, because those with low element interactivity are less likely to exceed our cognitive capacities (Sweller, 2010). In a similar vein, some of the largest pooled effect sizes were for forms of contiguity (i.e., presenting interacting elements in the same time and space). This design principle explicitly facilitates element interactivity, reducing the effort required for learners to search for links between interacting elements. In addition, findings are consistent with new conceptualizations of germane load as a form of intrinsic load made possible when students are motivated and extraneous load is low (Kalyuga, 2011; Sweller, 2010). Brom et al. (2018) found that pleasant colors and anthropomorphics modestly increased learning despite the risk of them being seductive details. They argue these improve learning because they promote increased cognitive engagement, meaning students are more likely to use free cognitive load for germane processes. Similarly, emerging research on "cognitive conflict" (e.g., presenting students with misconceptions then correcting them) may also increase learning (Rahim et al., 2015) because such conflict uses free cognitive capacity for germane processing. These findings are all consistent with theory: educators should reduce extraneous load and instead use that cognitive capacity for germane processes that promote deeper learning, or motivate students to do this themselves (Kalyuga, 2011; Sweller, 2010).

This interpretation helps explain our findings regarding the expertise reversal effect. Most theory papers hypothesized that prior knowledge should nullify the benefits of good multimedia design (Mayer, 2008, 2009; Sweller et al., 2019; van Merriënboer & Sweller, 2005). This moderation would exist because, compared with novice learners, knowledgeable learners would require less intrinsic load for the same task, and they can better manage their own cognitive load. For example, they could identify where they should be focusing without needing signaling. Unlike previous narrative reviews (Kalyuga, 2007), we did not find this effect was robust across meta-analyses. Some reviews found larger benefits for low prior-knowledge learners, as hypothesized (Kaushal & Panda, 2019; Sundararajan & Adesope, 2020), but others found the reverse (Rey et al., 2019; Richter et al., 2016). It is possible that these conflicting results may be explained by the idiosyncratic fit between some design principle and high versus low prior-knowledge learners. That is, there may be some reason that segmentation/signaling benefits high prior-knowledge learners more than low prior-knowledge learners, but that the reverse is true for animation/removing seductive details. However, this difference was not significant in our analysis and thus may be due to sampling error. We found the pooled effects of multimedia design principles were, on average, equivalent for high and low prior-knowledge learners. Good educators calibrate the challenge to the knowledge of the learner. Most educators would know when they are working with a high prior-knowledge sample (e.g., postgraduate medicine) rather than a low prior-knowledge one (e.g., primary school mathematics). We think it is self-evident that most educators would offer the high prior-knowledge learners more difficult tasks, closer to their zone of proximal development (Schnotz & Kürschner, 2007). While it is more likely that complex materials do benefit more from good multimedia design, we hypothesize that the prior knowledge of the learner is less important. That is, the expertise-reversal hypothesis—that good multimedia design helps nonexperts and hurts experts—was not supported by our analyses.

Similarly, reviews often addressed whether design principles were moderated by the subject being studied or the level of education. Neither of these factors consistently moderated effects. We argue that the reasons are similar to that of prior knowledge. Some subjects and levels of education may indeed be more prone to cognitive overload. But, good educators calibrate their teaching to the challenge involved in the subject. As a result, while there could indeed be some reason why signaling is stronger in universities than primary schools (S. Schneider et al., 2018), we are hesitant to overinterpret findings that may be better explained by sampling error. Findings may also be explained by the high chance of Type I error that is typical of social science meta-analyses, where multiple testing is rife (Cafri et al., 2010). Specifically, most of the meta-analyses included in this overview used a $p < 0.05$ critical value while conducting dozens of tests for moderation.

Nevertheless, this overview highlights the importance of using systematic reviews and meta-analyses to estimate effect sizes. Effect sizes of these design principles were generally small to moderate (Funder & Ozer, 2019), which is more conservative than previous reviews of multimedia design principles (e.g., Mayer, 2008). In those nonsystematic reviews, pooled effect sizes were very large ($d > 1.0$). In contrast, effects from the meta-analyses we found were more consistent with expected effects in psychology and education (Funder & Ozer, 2019; Hattie, 2008; M. Schneider & Preckel, 2017). So, while most were still significant and in the hypothesized direction, educators should have more conservative expectations for the effects of multimedia design on learning.

Practical Implications for Educators and Instructional Designers

Many of these design principles have robust support from systematic reviews and meta-analyses. More broadly, we found support for the premises underlying the cognitive theory of multimedia learning: that learners integrate information more effectively when it is presented via both sight and sound, as long as care is taken to reduce the chance of cognitive overload. Specifically, this means most educators could improve learning if they

- Speak to meaningful pictures, rather than relying on text, so learners are using both visual and auditory systems in the optimal way (visual system for pictures, auditory for sound)
- Where meaningful images are not possible, use some words presented visually (e.g., on slides) so the learner can use both visual and auditory systems in some way
- Present related words and pictures as close together as possible in both time and space, so learners are not using working memory to connect related material
- Help the learner know what is important using verbal emphasis, a pointer, arrows, highlighting (etc.), so learners are focused on key information
- Remove words or images that are not directly relevant to the core learning objectives so learners are not wasting mental resources attending to and processing irrelevant content
- Make visuals attractive by using pleasant color palettes that use color purposefully (e.g., use a contrast color to indicate something important) so visuals are engaging
- Animate visuals meaningfully, where possible, so the learner understands how something moves and how it works without having to fill in the motion of a static visual themselves
- Make abstract concepts more concrete by giving them agency and human features (e.g., with an “aggressive” virus drawn with a scowl, “attacking” the immune system)

- Ensure all additions—like colors, animations, and human features—are meaningful, not distracting (e.g., avoid using many colors, avoid distracting animations)
- Use simple, personal language that connects the learning concept to the learners' experiences, so they can construct the new knowledge in light of what they already know
- Break learning into meaningful chunks rather than long, uninterrupted presentations, so learners can consolidate each learning objective before moving on
- Include a person/character on screen only if they help direct attention to what is important, and even then consider if a more simple signal would be less distracting

Most of these features would be applicable across learning domains (e.g., STEM, social science), across ages (e.g., school, university, adult professional learning), and across uses of multimedia (e.g., online, face-to-face). Some other design features might only be useful in some contexts, like captioning of video for those learning second languages. Nevertheless, given the ubiquity of multimedia, the strategies above are useful for most educators to consider when designing how they present materials.

Limitations of This Overview of Reviews

Although our overview of reviews allowed us to explore more interventions than a review of primary studies, it was by no means exhaustive. Multimedia are only one component of how educators teach, and should be considered in light of the other evidence-based educational strategies that enhance learning (Hattie, 2008). Similarly, we only included reviews that focused on learning and cognitive load. In doing so, we excluded reviews that only addressed outcomes like student satisfaction, motivation, and engagement, which are important outcomes for both institutions and educators (Lawson & Lawson, 2013; Lazowski & Hulleman, 2016; Spooren et al., 2013). While we excluded no reviews on the basis of that criteria, we excluded a number that reported only health outcomes (e.g., Chandan et al., 2020; Ruiz-Roca et al., 2020). Health may be an important area where multimedia directly influences health and wellbeing (e.g., health promotion). Future reviews could identify how well the findings from this review translate to behavior-change outcomes sought in other domains (e.g., health or marketing).

By focusing on modifications to multimedia, we excluded reviews that only assessed person-level moderators. For example, meta-analyses have shown that higher spatial ability increases the benefits of visualizations, particularly when visualizations are abstract, system-paced, and lack signaling (Höffler, 2010). Similarly, nonsystematic reviews have collated evidence that supports the expertise reversal effect, where higher prior-knowledge reduces the benefit of good multimedia design (Kalyuga, 2007). While we used moderation effects reported in reviews to explore some of these questions, we were reliant on reviewers having assessed them in their meta-analyses. Future meta-analyses could robustly test these hypotheses by collecting all the primary studies that test these moderators (e.g., expertise reversal) and identify whether the phenomena are robust across all studies and design principles.

By focusing on review-level evidence, we were also unable to estimate the benefit from multimedia design principles where no review has been conducted, despite a range of primary studies. For example, there is some evidence that dialogues between learner and educator work well, even compared with a coherent exposition by a teacher (Chi et al., 2017). A systematic review of all studies on multimedia would have been intractable, with 1,189 primary studies identified by this overview. However, at our level of analysis, we necessarily excluded interventions like these that show promise but lack systematic reviews directly assessing their efficacy. We hope researchers continue to assess promising design principles like these via focused systematic reviews of the primary studies.

By analyzing reviews instead of primary studies, we were also beholden to the methodological decisions of previous reviewers. For overviews of reviews to make strong conclusions, the included meta-analyses should have comparable outcomes using a comparable reference point (Snook et al., 2009). As with systematic reviews of primary studies, they require strong methods and relatively homogeneous samples to calculate meaningful pooled effects (Pigott & Polanin, 2020). In general, our included reviews of multimedia design principles were an excellent fit for an overview of reviews because most reviews and primary studies reported similar outcomes, with similar comparisons (presence vs. absence of design principles). However, without extracting effect sizes of the 1,189 primary studies, we could not ensure consistency of analytical decisions across the included reviews. Different analytical decisions can lead to different interpretations from the same dataset (Silberzahn et al., 2018). While manuals and standards aim to reduce the variability in meta-analytic methods (e.g., Higgins et al., 2019; Pigott & Polanin, 2020), researchers conducting meta-analyses still make consequential decisions (e.g., which moderators to include, how to assess publication bias). Our review could not ensure each included review made the same decisions. Granted, meta-analyses of primary studies are vulnerable to the same weakness because reviewers cannot control for all decisions made by original authors. We focused interpretations on more expansive reviews and those with more robust quantitative methods, but beyond that, we could not statistically account for all the variability between reviews on the basis of methodological rigor. As described below, there was a large amount of variability between the quality of reviews, and like any systematic review, our review could not fully correct for those design variations.

Limitations of Included Reviews

The included reviews did not meet many established criteria for quality (Alexander, 2020; Pigott & Polanin, 2020). Few reviews explicitly reported that they had two authors screen studies independently (i.e., duplicate screening). This is important to ensure the sensitivity of the systematic search because adding a reviewer increases reliability, meaning reviews are more likely to include all relevant studies (Shea et al., 2017). Most studies assessed publication bias and heterogeneity. However, the methods of assessing publication bias were inconsistent. Many relied on methods prone to Type I and Type II errors, like

Trim-and-Fill, Egger's regression, or the rank correlation test (Pustejovsky & Rodgers, 2019). Few used selection methods (Hedges & Vevea, 1996), which are less prone to these biases (Pustejovsky & Rodgers, 2019). Similarly, the methods of assessing heterogeneity were underpowered and prone to Type I error (Cafri et al., 2010), as noted earlier. We seldom identified reviews that were prospectively registered, which can reduce the risk of these biases (Greco et al., 2013; Pigott & Polanin, 2020).

Most egregiously, we only found one review that assessed the quality of the included primary studies (Adesope & Nesbit, 2012), and that quality assessment was brief (e.g., "Was reliability reported?"). While we acknowledge this is not common practice in education research (Pigott & Polanin, 2020), it is well established in other fields that the quality of the primary studies affects the quality of the meta-analytic estimate (Higgins et al., 2019; Shea et al., 2017). Many well-established biases (e.g., unblinded designs) are hard to avoid in education, but until we have strong evidence that they do not affect meta-analytic estimates for educational interventions, we should presume that educational research is not immune to experimental bias. Therefore, we recommend that future reviews implement a gold-standard assessment of the quality of primary studies (currently RoB 2; Sterne et al., 2019) drawn from meta-meta-analytic estimates of what design choices inflate effect sizes (e.g., unblinded designs, no prospective registration). Without these safeguards, reviewers cannot assess whether or not their reviews are succumbing to the "garbage in, garbage out" phenomenon (Taylor et al., 2021).

Conclusions

In this overview of reviews, we found robust support for many multimedia design principles informed by the cognitive theory of multimedia learning and cognitive load theory. Specifically, we found that interventions that leveraged our two channels of information processing (visual and auditory) improved learning (e.g., captioning second-language video; modality effect). We also found benefits for strategies that reduced extraneous cognitive load (e.g., contiguity, signaling, segmentation, personalization, removing seductive details). Many such interventions have robust support for their causal models, with randomized experiments showing improvements in both learning and hypothesized mediators (e.g., cognitive load, eye movements). While the quality of the reviews was inconsistent, the consistency of the results and strength of the effects provide strong support for these models of multimedia learning. Results show good multimedia design can have a transformative effect on a range of educational outcomes. As a result, mastering these design principles is a critical competency for all educators.

ORCID iDs

Michael Noetel https://orcid.org/0000-0002-6563-8203

Oscar Delaney https://orcid.org/0000-0002-8792-0744

Philip Parker https://orcid.org/0000-0002-4604-8566

Chris Lonsdale https://orcid.org/0000-0002-2523-5565

References

References marked with an asterisk indicate reviews meeting inclusion criteria for the meta-review.

*Adesope, O. O., & Nesbit, J. C. (2012). Verbal redundancy in multimedia learning environments: A meta-analysis. *Journal of Educational Psychology*, *104*(1), 250–263. https://doi.org/10.1037/a0026147

*Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, *125*, 413–428. https://doi.org/10.1016/j.compedu.2018.06.023

Alexander, P. A. (2020). Methodological guidance paper: The art and science of quality systematic reviews. *Review of Educational Research*, *90*(1), 6–23. https://doi.org/10.3102/0034654319854352

*Alpizar, D., Adesope, O. O., & Wong, R. M. (2020). A meta-analysis of signaling principle in multimedia learning environments. *Educational Technology Research and Development: ETR&D*, *68*(5), 2095–2119. https://doi.org/10.1007/s11423-020-09748-7

*Amaka, I. H., & Goeman, K. (2017). Selecting media for effective learning in online and blended courses: A review study. *Journal of Educational Multimedia and Hypermedia*, *26*(1), 29–59. https://www.learntechlib.org/primary/p/173341/

Angier, N. (2009, August 3). Finally, the spleen gets some respect. *The New York Times*. https://www.nytimes.com/2009/08/04/science/04angier.html

APA Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *The American Psychologist*, *63*(9), 839–851. https://doi.org/10.1037/0003-066X.63.9.839

Baker, J. P., Goodboy, A. K., Bowman, N. D., & Wright, A. A. (2018). Does teaching with PowerPoint increase students' learning? A meta-analysis. *Computers & Education*, *126*, 376–387. https://doi.org/10.1016/j.compedu.2018.08.003

*Berney, S., & Bétrancourt, M. (2016). Does animation enhance learning? A meta-analysis. *Computers & Education*, *101*, 150–167. https://doi.org/10.1016/j.compedu.2016.06.005

*Brom, C., Stárková, T., & D'Mello, S. K. (2018). How effective is emotional design? A meta-analysis on facial anthropomorphisms and pleasant colors during multimedia learning. *Educational Research Review*, *25*, 100–119. https://doi.org/10.1016/j.edurev.2018.09.004

Cafri, G., Kromrey, J. D., & Brannick, M. T. (2010). A meta-meta-analysis: Empirical review of statistical power, Type I error rates, effect sizes, and model selection of meta-analyses published in psychology. *Multivariate Behavioral Research*, *45*, 239–270. https://doi.org/10.1080/00273171003680187

*Castro-Alonso, J. C., Wong, R. M., Adesope, O. O., & Paas, F. (2021). Effectiveness of multimedia pedagogical agents predicted by diverse theories: A meta-analysis. *Educational Psychology Review*. Advance online publication. https://doi.org/10.1007/s10648-020-09587-1

Chandan, S., Arora, S., Mohan, B. P., Khan, S. R., Chandan, O. C., Kassab, L. L., & Murali, A. R. (2020). Multimedia based education on bowel preparation improves adenoma detection rate: Systematic review & meta-analysis of randomized controlled trials. *Digestive Endoscopy*, *33*(5), 730–740. https://doi.org/10.1111/den.13809

Cheung, M. W.-L. (2014a). metaSEM: an R package for meta-analysis using structural equation modeling. *Frontiers in Psychology*, *5*, 1521. https://doi.org/10.3389/fpsyg.2014.01521

Cheung, M. W.-L. (2014b). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*(2), 211–229. https://doi.org/10.1037/a0032968

Chi, M. T. H., Kang, S., & Yaghmourian, D. L. (2017). Why students learn more from dialogue- than monologue-videos: Analyses of peer interactions. *Journal of the Learning Sciences*, *26*(1), 10–50. https://doi.org/10.1080/10508406.2016.1204546

Conigrave, J. (2019). *msemtools: Routines, tables, and figures for metaSEM analyses* (Version 0.9.8) [Computer software]. Github. https://github.com/JConigrave/msemtools

*Cromley, J. G., & Lawrence, L. E. M. (2018). Multimedia simulations that foster transfer: Findings from a review of the literature. In R. Z. Zheng (Ed.), *Strategies for deep learning with digital technology: Theories and practices in education* (pp. 93–110). Nova Science.

*Davis, R. O. (2018). The impact of pedagogical agent gesturing in multimedia learning environments: A meta-analysis. *Educational Research Review*, *24*, 193–209. https://doi.org/10.1016/j.edurev.2018.05.002

Deeks, J. J. (2019). Chapter 10: Analysing data and undertaking meta-analyses In: J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions, Version 6*. https://training.cochrane.org/handbook/archive/v6/chapter-10

Del Re, A. C. (2020). Package "compute.es": Compute effect sizes (Version 0.2-5) [Computer software]. https://cran.r-project.org/web/packages/compute.es/compute.es.pdf

Eriksen, M. B., & Frandsen, T. F. (2018). The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: A systematic review. *Journal of the Medical Library Association: JMLA*, *106*(4), 420–431. https://doi.org/10.5195/jmla.2018.345

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

*Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, *15*(4), 313–331. https://doi.org/10.1016/j.learninstruc.2005.07.001

*Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, *16*(6), 511–525. https://doi.org/10.1016/j.learninstruc.2006.10.001

*Ginns, P., Martin, A. J., & Marsh, H. W. (2013). Designing instructional text in a conversational style: A meta-analysis. *Educational Psychology Review*, *25*(4), 445–472. https://doi.org/10.1007/s10648-013-9228-0

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, *26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Greco, T., Zangrillo, A., Biondi-Zoccai, G., & Landoni, G. (2013). Meta-analysis: Pitfalls and hints. *Heart, Lung and Vessels*, *5*(4), 219–225. https://www.ncbi.nlm.nih.gov/pubmed/24364016

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge. https://doi.org/10.4324/9780203887332

Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics: A Quarterly Publication Sponsored by the American Educational Research Association and the American Statistical Association*, *6*(2), 107–128. https://doi.org/10.3102/10769986006002107

Hedges, L. V., & Vevea, J. L. (1996). Estimating effect size under publication bias: Small sample properties and robustness of a random effects selection model. *Journal of Educational and Behavioral Statistics*, *21*(4), 299–332. https://doi.org/10.3102/10769986021004299

Hennessy, E. A., & Johnson, B. T. (2020). Examining overlap of included studies in meta-reviews: Guidance for using the corrected covered area index. *Research Synthesis Methods*, *11*(1), 134–145. https://doi.org/10.1002/jrsm.1390

Hennessy, E. A., Johnson, B. T., & Keenan, C. (2019). Best practice guidelines and essential methodological steps to conduct rigorous and systematic meta-Reviews. *Applied Psychology. Health and Well-Being*, *11*(3), 353–381. https://doi.org/10.1111/aphw.12169

Higgins, J. P. T., Altman, D. G., & Sterne, J. A. C., & Cochrane Statistical Methods Group and the Cochrane Bias Methods Group. (2011). Assessing risk of bias in included studies. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions* (Vol. 5.1.1). The Cochrane Collaboration. https://handbook-5-1.cochrane.org/chapter_8/

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley. https://training.cochrane.org/cochrane-handbook-systematic-reviews-interventions

Hirschey, R., Bryant, A. L., Walker, J. S., & Nolan, T. S. (2020). Systematic review of video education in underrepresented minority cancer survivors. *Cancer Nursing*, *43*(4), 259–268. https://doi.org/10.1097/NCC.0000000000000829

*Höffler, T. N. (2010). Spatial ability: Its influence on learning with visualizations—A meta-analytic review. *Educational Psychology Review*, *22*, 245–269. https://doi.org/10.1007/s10648-010-9126-7

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, *17*(6), 722–738. https://doi.org/10.1016/j.learninstruc.2007.09.013

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509–539. https://doi.org/10.1007/s10648-007-9054-3

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review*, *23*(1), 1–19. https://doi.org/10.1007/s10648-010-9150-7

*Kang, T. (2019). The effectiveness of multiple media tools in L2 listening: A meta-analysis. In Information Resources Management Association (Ed.), *Computer-assisted language learning: Concepts, methodologies, tools, and applications* (pp. 585–616). IGI Global. https://doi.org/10.4018/978-1-5225-7663-1.ch028

*Kaushal, R. K., & Panda, S. N. (2019). A meta analysis on effective conditions to offer animation based teaching style. *Malaysian Journal of Learning and Instruction*, *16*(1), 129–153. https://eric.ed.gov/?id=EJ1219792

*Kim, S. (2005). The effect of multimedia design elements on learning outcomes in pedagogical agent research: A meta-analysis. *Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology*, *963*. https://dl.acm.org/doi/proceedings/10.5555/1562524

Lawson, M. A., & Lawson, H. A. (2013). New conceptual frameworks for student engagement research, policy, and practice. *Review of Educational Research*, *83*(3), 432–479. https://doi.org/10.3102/0034654313480891

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation interventions in education: A meta-analytic review. *Review of Educational Research*, *86*(2), 602–640. https://doi.org/10.3102/0034654315617832

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *The American Psychologist*, *63*(8), 760–769. https://doi.org/10.1037/0003-066X.63.8.760

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

Mayer, R. E. (2014). Incorporating motivation into multimedia learning. *Learning and Instruction*, *29*, 171–173. https://doi.org/10.1016/j.learninstruc.2013.04.003

Mayer, R. E. (2019). Searching for the role of emotions in e-learning. *Learning and Instruction*, *70*, 101213. https://doi.org/10.1016/j.learninstruc.2019.05.010

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 279–315). New York: Cambridge University Press.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, *93*(1), 187–198. https://doi.org/10.1037/0022-0663.93.1.187

Mayer, R. E., & Moreno, R. (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 131–152). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.009

Moeyaert, M., Ugille, M., Natasha Beretvas, S., Ferron, J., Bunuan, R., & Van den Noortgate, W. (2016). Methods for dealing with multiple outcomes in meta-analysis: A comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology*, *20*(6), 559–572. https://doi.org/10.1080/13645579.2016.1252189

*Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, *41*(3), 720–739. https://doi.org/10.1016/j.system.2013.07.013

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology*, *94*(1), 156–163. https://doi.org/10.1037/0022-0663.94.1.156

National Heart, Lung, and Blood Institute. (n.d.). *Study quality assessment tools*. https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, *76*(3), 413–448. https://doi.org/10.3102/00346543076003413

Noetel, M., Griffith, S., Delaney, O., Sanders, T., Parker, P. D., del Pozo Cruz, B., & Lonsdale, C. (2021). Video improves learning in higher education: A systematic review. *Review of Educational Research*, *91*(2), 204–236. https://doi.org/10.3102/0034654321990713

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., & Moher, D. (2021). Updating guidance for reporting systematic reviews: Development of the PRISMA 2020 statement. *Journal of Clinical Epidemiology*, *134*, 103–112. https://doi.org/10.1016/j.jclinepi.2021.02.003

Pigott, T. D., & Polanin, J. R. (2020). Methodological guidance paper: High-quality meta-analysis in a systematic review. *Review of Educational Research*, *90*(1), 24–46. https://doi.org/10.3102/0034654319877153

Pollock, M., Fernandes, R. M., Becker, L. A., Pieper, D., & Hartling, L. Others. (2018). Chapter V: Overviews of reviews. *Cochrane Handbook for systematic reviews of interventions, Version 6*. https://training.cochrane.org/handbook/current/chapter-v

Pustejovsky, J. E., & Rodgers, M. A. (2019). Testing for funnel plot asymmetry of standardized mean differences. *Research Synthesis Methods*, *10*(1), 57–71. https://doi.org/10.1002/jrsm.1332

R Core Team. (2020). R: A language and environment for statistical computing (Version 3.6.3) [Computer software]. *R Foundation for Statistical Computing*. https://www.R-project.org/

*Rahim, R. A., Noor, N. M., & Zaid, N. M. (2015). Meta-analysis on element of cognitive conflict strategies with a focus on multimedia learning material development. *International Education Studies*, *8*(13), 73–78. https://doi.org/10.5539/ies.v8n13p73

Rathbone, J., Carter, M., Hoffmann, T., & Glasziou, P. (2015). Better duplicate detection for systematic reviewers: Evaluation of Systematic Review Assistant-Deduplication Module. *Systematic Reviews*, *4*, 6. https://doi.org/10.1186/2046-4053-4-6

*Reinwein, J. (2012). Does the modality effect exist? And if so, which modality effect? *Journal of Psycholinguistic Research*, *41*(1), 1–32. https://doi.org/10.1007/s10936-011-9180-4

*Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review*, *7*(3), 216–237. https://doi.org/10.1016/j.edurev.2012.05.003

*Rey, G. D., Beege, M., Nebel, S., Wirzberger, M., Schmitt, T. H., & Schneider, S. (2019). A meta-analysis of the segmenting effect. *Educational Psychology Review*, *31*(2), 389–419. https://doi.org/10.1007/s10648-018-9456-4

*Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, *17*, 19–36. https://doi.org/10.1016/j.edurev.2015.12.003

Rolfe, V. E., & Gray, D. T. (2011). Are multimedia resources effective in life science education? A meta-analysis. *Bioscience Education*, *18*(1), 1–14. https://doi.org/10.3108/beej.18.5

Ruiz-Roca, J. A., Martínez-Izquierdo, A., Mengual-Pujante, D., López, E. P. F., & López-Jornet, P. (2020). Is YouTube a useful tool for oral care in patients with Parkinson’s disease? *Special Care in Dentistry*, *40*(5), 464–469. https://doi.org/10.1111/scd.12489

Schmidt, F. L., & Oh, I.-S. (2013). Methods for second order meta-analysis and illustrative applications. *Organizational Behavior and Human Decision Processes*, *121*(2), 204–218. https://doi.org/10.1016/j.obhdp.2013.03.002

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, *143*(6), 565–600. https://doi.org/10.1037/bul0000098

*Schneider, S., Beege, M., Nebel, S., & Rey, G. D. (2018). A meta-analysis of how signaling affects learning with media. *Educational Research Review*, *23*, 1–24. https://doi.org/10.1016/j.edurev.2017.11.001

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, *19*(4), 469–508. https://doi.org/10.1007/s10648-007-9053-4

*Schroeder, N. L., Adesope, O. O., & Gilbert, R. B. (2013). How effective are pedagogical agents for learning? A meta-analytic review. *Journal of Educational Computing Research*, *29*(1), 1–39. https://doi.org/10.2190/EC.49.1.a

*Schroeder, N. L., & Cenkci, A. T. (2018). Spatial contiguity and spatial split-attention effects in multimedia learning environments: A meta-analysis. *Educational Psychology Review*, *30*(3), 679–701. https://doi.org/10.1007/s10648-018-9435-9

*Schroeder, N. L., & Cenkci, A. T. (2020). Do measures of cognitive load explain the spatial split-attention principle in multimedia learning environments? A systematic review. *Journal of Educational Psychology*, *112*(2), 254–270. https://doi.org/10.1037/edu0000372

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, *358*, j4008. https://doi.org/10.1136/bmj.j4008

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., . . . Nosek, B. A. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, *1*(3), 337–356. https://doi.org/10.1177/2515245917747646

Smith, V., Devane, D., Begley, C. M., & Clarke, M. (2011). Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC Medical Research Methodology*, *11*(1), 15. https://doi.org/10.1186/1471-2288-11-15

Snook, I., O'Neill, J., Clark, J., O'Neill, A.-M., Openshaw, R., & Others. (2009). Invisible learnings? A commentary on John Hattie's book—"Visible learning: A synthesis of Over 800 meta-analyses relating to achievement". *New Zealand Journal of Educational Studies*, *44*(1), 93. https://doi.org/10.3316/INFORMIT.467818990993648

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, *83*(4), 598–642. https://doi.org/10.3102/0034654313496870

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., . . . Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials [Review of *RoB 2: A revised tool for assessing risk of bias in randomised trials*]. *BMJ*, *366*, l4898. https://doi.org/10.1136/bmj.l4898

*Sundararajan, N., & Adesope, O. (2020). Keep it coherent: A meta-analysis of the seductive details effect. *Educational Psychology Review*, *32*, 707–734. https://doi.org/10.1007/s10648-020-09522-4

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Taylor, K. S., Mahtani, K. R., & Aronson, J. K. (2021). Summarising good practice guidelines for data extraction for systematic reviews and meta-analysis. *BMJ Evidence-Based Medicine*. https://doi.org/10.1136/bmjebm-2020-111651

Van Gog, T., Ericsson, K. A., Rikers, R. M., & Paas, F. (2005). Instructional design for advanced learners: Establishing connections between the theoretical frameworks of cognitive load and deliberate practice. *Educational Technology Research and Development: ETR&D*, *53*(3), 73–81. https://doi.org/10.1007/BF02504799

van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*(2), 147–177. https://doi.org/10.1007/s10648-005-3951-0

Wallace, B. C., Small, K., Brodley, C. E., Lau, J., & Trikalinos, T. A. (2012). Deploying an interactive machine learning system in an evidence-based practice center: Abstrackr. *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, 819–824. https://doi.org/10.1145/2110363.2110464

*Xie, H., Wang, F., Zhou, Z., & Wu, P. (2016). Cueing effect in multimedia learning: A meta-analysis. *Xin Li Xue Bao. Acta Psychologica Sinica*, *48*(5), 540. https://doi.org/10.3724/sp.j.1041.2016.00540

*Xie, H., Wang, F., Hao, Y., Chen, J., An, J., Wang, Y., & Liu, H. (2017). The more total cognitive load is reduced by cues, the better retention and transfer of multimedia learning: A meta-analysis and two meta-regression analyses. *PloS One*, *12*(8), e0183884. https://doi.org/10.1371/journal.pone.0183884

Authors

MICHAEL NOETEL is a senior lecturer in the School of Health and Behavioural Sciences at the Australian Catholic University, 1100 Nudgee Road, Banyo, Queensland 4014, Australia; email: *michael.noetel@acu.edu.au*. He received his doctorate in psychology from Australian Catholic University. His major research interests include online learning, motivation, and research synthesis.

SHANTELL GRIFFITH is a postgraduate student in social work at the Australian Catholic University, 1100 Nudgee Road, Banyo, Queensland 4014, Australia; email: *shantell.griffith@myacu.edu.au*. She received her bachelor of psychological science from the Australian Catholic University.

OSCAR DELANEY is an undergraduate philosophy and biology student at the Faculty of Humanities and Social Sciences, University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia; email: *o.delaney@uq.net.au*. He has research interests in climate adaptation and plant genetics.

NICOLA ROSE HARRIS is a postgraduate medicine student at the University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia; email: *rosienharris@gmail.com.* She has an interest in psychology and neuroscience. She completed her bachelor of science with a major in biomedical science at the University of Queensland.

TAREN SANDERS is the deputy program leader for the Motivation and Behaviour Research Program in the Institute for Positive Psychology and Education at Australian Catholic University, 33 Berry St, North Sydney, NSW 2060, Australia; email: *taren.sanders@acu.edu.au*. He received his doctorate from Western Sydney University. His major research interest includes screen use and physical activity of children and young people.

PHILIP PARKER is the deputy director of the Institute for Positive Psychology and Education at Australian Catholic University, 33 Berry St, North Sydney, NSW 2060, Australia; email: *philip.parker@acu.edu.au.* He received his doctorate in educational psychology from the University of Sydney. His major research interest includes educational inequality, developmental transitions, and educational attainment.

BORJA DEL POZO CRUZ is an associate professor in the Department of Sports Science and Clinical Biomechanics at the University of Southern Denmark, Odense, Syddanmark, Denmark; email: *BdelPozoCruz@health.sdu.dk*. He received his doctorate from the University of Pablo de Olavide. His major research interest includes health promotion, physical activity, and analytical methods.

CHRIS LONSDALE is the associate dean of research for the Faculty of Health Sciences, and program leader in the Institute for Positive Psychology and Education at Australian Catholic University, 33 Berry St, North Sydney, NSW 2060, Australia; email: *chris.lonsdale@acu.edu.au*. He received his doctorate from the University of Otago. His major research interest includes the motivation underpinnings of behavior in education and health contexts.