How Effective is Emotional Design? A Meta-Analysis on Facial Anthropomorphisms and Pleasant Colors during Multimedia Learning

Cyril Brom (corresponding author)

Faculty of Mathematics and Physics, Charles University

Malostranské nám. 25, 118 00, Prague, the Czech Republic

brom@ksvi.mff.cuni.cz

Tereza Stárková

Faculty of Mathematics and Physics, Charles University

Faculty of Arts, Charles University

starkova@ksvi.mff.cuni.cz

Sidney K. D’Mello

Department of Computer Science & Institute of Cognitive Science, University of Colorado Boulder

sidney.dmello@colorado.edu

Notice. This is the author's version of a work that was accepted for publication in Educational Research Review. Changes resulting from the publishing process, such as structural formatting, and other quality control mechanisms may not be reflected in this document. A definitive version was accepted for publication in Educational Research Review (2018), DOI: 10.1016/j.edurev.2018.09.004. The paper was accepted: 18-SEP-2018.

Keywords. Multimedia learning; emotional design; anthropomorphisms; pleasant color; meta-analysis

Source of funding. C.B. and T.S. were supported by the projects PRIMUS/HUM/03 (Charles University) and 15-14715S (Czech Grant Science Foundation (GA ČR). T. S. was also partly supported by PROGRES Q15. S. K. D. was supported by the NSF (IIS 1523091) and IES (R305A170376). Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.

Acknowledgement. We thank to all authors of studies included in this meta-analysis who provided additional information concerning the studies.

Appendix A. Supplementary data to this article can be found online at https://doi.org/10.1016/j.edurev.2018.09.004. Alternatively, they can be sent by the corresponding author upon request.

Abstract. We conducted a meta-analysis of 33 independent samples (N = 2924) to address whether adding anthropomorphic faces to multimedia graphics and/or adding pleasant colors are effective emotional design approaches. We found significant positive meta-analytic effects for retention (k = 18, d_+ = 0.387), comprehension (k = 14, d_+ = 0.317), and transfer (k = 27, d_+ = 0.327) under a random-effects model. Effects for affective-motivational variables were mixed, with a robust effect for intrinsic motivation (k = 23, d_+ = 0.255), a weaker effect for liking/enjoyment (k = 20, d_+ = 0.109), and a marginal effect for positive affect (k = 15, d_+ = 0.113). The manipulations did not significantly (ps > .227) influence perceptions of learning (k = 11, d_+ = 0.097) or effort (k = 20, d_+ = 0.051), but reduced perceptions of difficulty (k = 14, d_+ = -0.208). Four of the outcome variables (retention, transfer, intrinsic motivation, and perceived effort) were sufficiently heterogeneous. There was no major issue with publication bias, influential cases, or outliers. With one exception, there was no evidence of moderation by experimental contrast, dynamicity of materials, age, language/culture, prior mood, time-on-task, and publication type after adjusting for multiple comparisons. There was provisional evidence that age moderated the effect of the manipulations on intrinsic motivation, such that larger effects were revealed for children compared to older learners. Altogether, anthropomorphisms/colors appear to be useful design principles.

Keywords: Multimedia learning; emotional design; anthropomorphisms; pleasant color; meta-analysis

Highlights

- Anthropomorphisms and pleasant colors should enhance multimedia learning.

- We test this hypothesis using a meta-analysis of 33 independent samples.
- Anthropomorphisms/colors consistently increased learning outcomes.
- Their effects on affective-motivational states were weaker and less robust.
- Anthropomorphisms/colors are beneficial, but the mechanism is unclear.

1. Introduction

How should one design a multimedia learning experience to improve learning? One can, of course, appeal to intuition or folklore. However, a better approach is to follow the cognitive principles of multimedia learning (e.g., Mayer, 2009; Mayer, 2014a; Renkl & Scheiter, 2017). These principles provide empirical guidelines for instructional designers and are derived from established theoretical frameworks, such as the cognitive theory of multimedia learning (Mayer, 2009) and cognitive load theory (Sweller, Ayeres, & Kalyuga, 2011). However, these theories, and the principles alike, primarily focus on the cognitive aspects of learning: ignoring the moderating role of affective-motivational states (D'Mello & Graesser, 2012; Fiedler & Beier, 2014).

At the same time, researchers have sought to improve instruction by leveraging affective-motivational factors (e.g., Garner, Brown, Sanders, & Menke, 1992; Kaplan & Pascoe, 1977; see also Abrami, Leventhal, & Perry, 1982; Yeager & Walton, 2011). This research avenue has been called *emotional design* when applied in the context of multimedia learning (Plass & Kaplan, 2015; Um, Plass, Hayward, & Homer, 2012; see also Mayer, 2014b; see Norman, 2004, for broader use of the term) and has led to the emergence of *emotional design principles*, which complement cognitive principles of multimedia learning.

Adding facial anthropomorphisms to non-human graphical elements and/or adding pleasant colors in instructional presentations, animations, and web-pages (henceforth referred to as anthropomorphisms/colors for short; see Figure 1) are popular emotional design principles. Researchers have theorized that these augmentations should improve learning by elevating affective-motivational states such as enjoyment, flow, situational interest, and intrinsic motivation (cf. Plass & Kaplan, 2015, p. 138; Um et al., 2012, p. 488; see also Norman, 2004; Pekrun & Linnenbrink-Garcia, 2012). Emotional design approaches are distinct in that they directly target learning materials compared to other affective-motivational manipulations such as mood induction procedures (e.g., Knörzer, Brünken & Park, 2016), setting classroom goal structures (e.g., Ke, 2008), changing task instructions (e.g., Hawlitschek & Joeckel, 2017), changing workspace color (e.g., Stone & English, 1998), or providing choice (e.g., Fulmer, D'Mello, & Graesser, 2015).

Are anthropomorphisms/colors effective emotional design manipulations? Whereas some studies have indicated that their inclusion in instructional materials improves learning outcomes (e.g., Mayer & Estrella, 2014; Schneider et al., 2018a; Um et al., 2012), others have failed to show a positive effect on learning (e.g., Heidig, Müller, & Reichelt, 2015; Münchow, 2017, Exp. 3; Park, Knörzer, Plass, & Brünken, 2015). Results have also been inconclusive with respect to the effects of anthropomorphisms/colors on affective-motivational variables with studies reporting positive (e.g., Schneider et al., 2018a), null (e.g., Münchow, 2017, Exp. 3), or mixed results (e.g., Mayer & Estrella, 2014). To provide some clarity, we used meta-analytic techniques to investigate their effects on both learning and affective-motivational states.

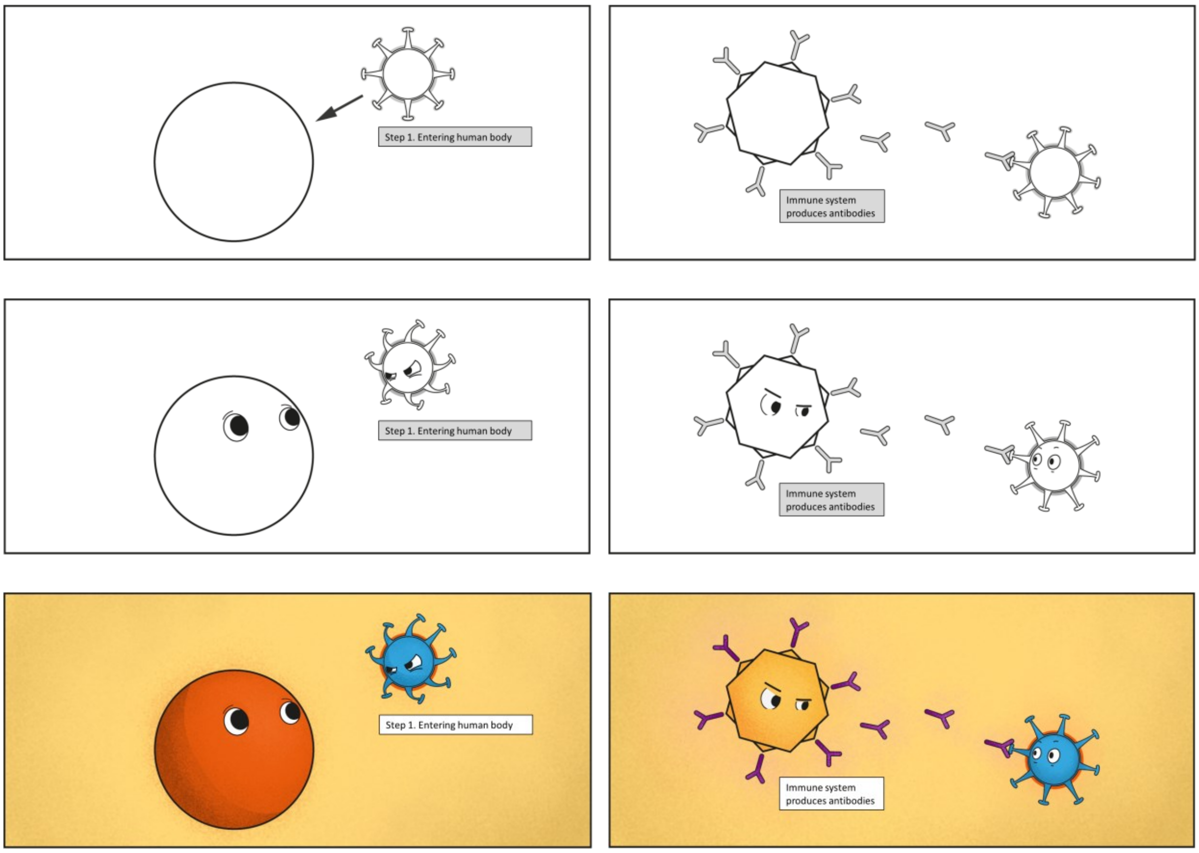

Figure 1. An example manipulation from an unpublished study by Brom and colleagues (with permission), which is an extension of Mayer and Estrella (2014). The graphics depict two steps in the process by which the influenza virus attacks the human body. Top: Two examples of a schematic "baseline" version of the graphics. Middle: black-and-white anthropomorphic version. Bottom: anthropomorphic version with pleasant colors.

1.1. Theoretical background

Why have emotional design studies in general, and anthropomorphism/color studies in particular, yielded inconclusive findings? One reason is that the manipulations can have conflicting effects on learning processes (e.g., Leutner, 2014; Mayer, 2014b). According to contemporary theories of learning (e.g., Moreno, 2005; Mayer, 2009; Sweller, Ayeres, & Kalyuga, 2011), learners must actively engage in mental processes: such as selecting relevant bits of information and integrating them with prior knowledge to construct mental models that are eventually consolidated into long-term memory. This active learning process requires managing limited working memory resources. Cognitive load theory (CLT; Sweller, Ayeres, & Kalyuga, 2011) posits two types of cognitive load imposed during learning: intrinsic and extraneous (see Kalyuga, 2011 for a discussion on germane load). Intrinsic load is derived from the complexity of the learning task with respect to the learner's prior knowledge and strategy use. Learners need to allocate sufficient resources to accommodate intrinsic load for learning to occur. Extraneous load arises from the suboptimal design of the instructional materials and requires learners to process information irrelevant to the central learning task. Because extraneous load consumes cognitive resources needed to accommodate intrinsic load, it should be minimized via good instructional design. Thus, according to CLT, it is better to avoid emotional design elements because they mainly increase extraneous load (see, e.g., Rey, 2012; Um et al., 2012 for a discussion). They might also increase levels of affective-motivational states to the point of distraction (e.g., Pekrun & Linnenbrink-Garcia, 2012; Um et al., 2012); further reducing cognitive resources available for managing intrinsic load.

In contrast, the cognitive-affective theory of learning from media (CATLM; Moreno, 2005), which is a recent expansion of the cognitive theory of multimedia learning (Mayer, 2009), would posit that positive-activating affective-motivational states triggered by emotional design manipulations would also trigger deeper cognitive processing; essentially increasing cognitive resources devoted to the task (see Figure 2). As such, emotional design principles would be beneficial to both learning and affective-motivational states.

In light of this discussion, an effective emotional design manipulation should elevate learning-centered, positive-activating affective-motivational states (e.g., enjoyment, interest), while not increasing *extraneous* cognitive load to a level that would inhibit learning.

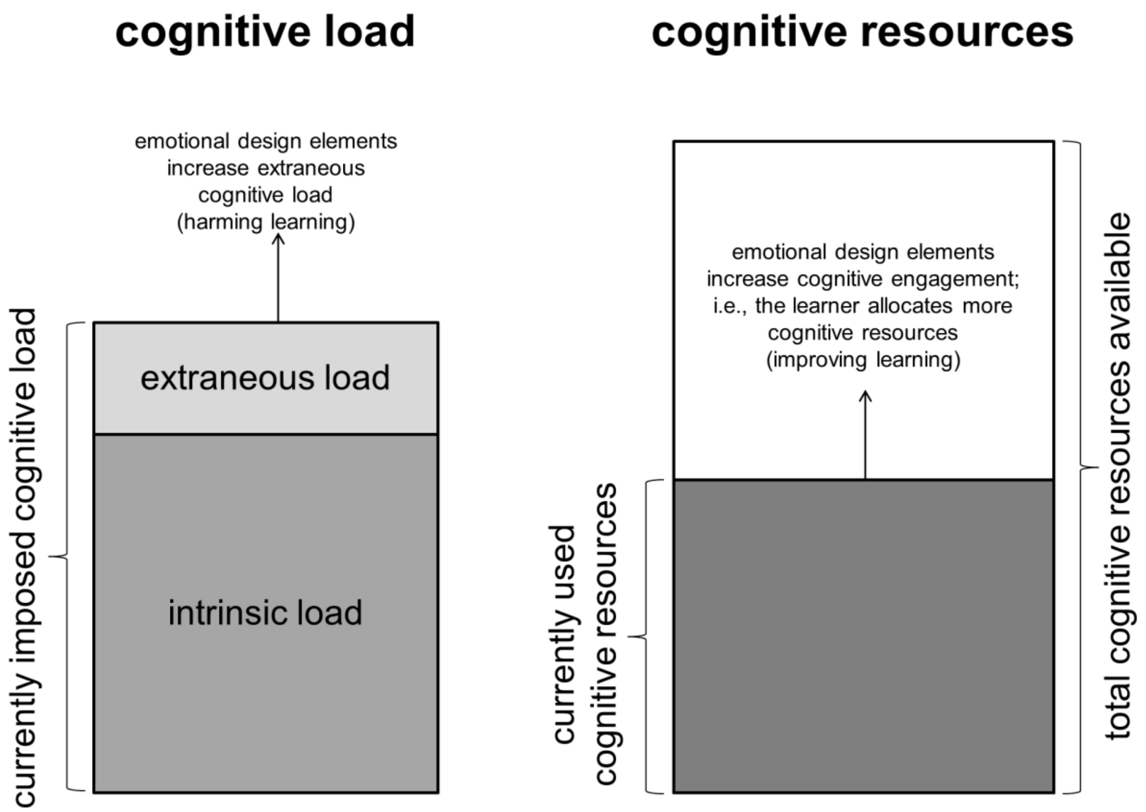

Figure 2. Opposing effects of emotional design elements on cognitive load and cognitive resources available for learning.

1.2. Anthropomorphisms and colors as emotional design manipulations

What are effective emotional design manipulations? Let us first consider two that would not apply. In particular, it is well known that embellishing texts or pictures with elements that capture interest but are not central to the learning content (i.e., seductive details) increases extraneous cognitive load and may inhibit learning (see Garner et al., 1992; Rey, 2012; but see also Schneider, Nebel, & Rey, 2016). Some argue that game design elements, such as narratives, interactivity, or points, are all examples of seductive details (e.g., Adams et al., 2012; Mayer, 2014c). Adding seductive details is not a promising emotional design principle.

An alternative approach is to “minimally” augment instructional materials to elevate positive-activating affective-motivational states while minimally increasing cognitive load (cf. Brom, Děchtěrenko et al., 2017; Mayer & Estrella, 2014). Such augmentations should not change (or should not change much) the number of informational units in a text, the number of elements in an accompanying picture, or the complexity of interactions among informational units and image elements. Changing the style of instructional texts from formal to conversational – the personalization principle (Mayer, 2009) – is an example of a minimalistic textual manipulation. Although it can enhance learning (see Ginns, Martin, & Marsh, 2013), it is unclear if the improvements are due to affective-motivational factors (see Brom, Hannemann et al., 2017). Because provisional evidence indicates that conversational texts increase perceived friendliness, but not affective-motivational states like interest (see Ginns et al, 2013), we would not yet categorize the personalization principle as an emotional design principle.

A more promising approach is to consider “minimalistic” alterations to graphics, such as changing color (e.g., Heidig et al., 2015; Plass, Heidig, Hayward, Homer, & Um, 2014), typeface (font) (e.g., Kumar, Muniandy, & Yahaya, 2016), sharpness of edges/corners (e.g., Um et al., 2012; Münchow, Mengelkamp, & Bannert, 2017), or embedding elements of humor in the graphic (e.g., Brom et al., 2016). The most frequent and salient of such manipulations involve facial anthropomorphisms (schematic versus anthropomorphic) and/or color (black-and-white/grey-scale versus pleasant/aesthetic colors) (see Figure 1). It is theorized that facial anthropomorphisms can positively activate learners (e.g., Um et al., 2012); presumably because facial expressions communicate emotions (e.g., Ekman & Rosenberg, 1997) and due to baby-face bias (i.e., positively-valenced reactions toward “baby-like” rounded objects with infantile facial features (Lorenz, 1970)). Because color influences human emotion, cognition, and behavior (e.g., Valdez & Mehrabian, 1994; Weller &

Livingston, 1988; Wolfson & Case, 2000), pleasant/aesthetic colors should increase positive affective-motivational states (e.g., Plass et al., 2014; Um et al., 2012; see also Heidig et al., 2015). Negative color manipulations are also possible (e.g., Kumar et al., 2016), but emotional design research has mainly focused on pleasant color manipulations; i.e., using predominantly warm, bright, saturated colors (Plass et al., 2014; Um et al., 2012).

We would consider adding facial anthropomorphisms to non-human elements (e.g., Um et al., 2012; Mayer & Estrella, 2014) (see Figure 1) a "minimalistic" manipulation because it should not impose too much extraneous load as face processing is fast and automatic (e.g., Crouzet & Thorpe, 2011) and a non-human element with an added face can still be considered one information "chunk". Similarly, pleasant colors should also have a minimal influence on cognitive load if selected using good visual design principles (i.e., avoiding gaudy and garish colors).

1.3. Boundaries of anthropomorphisms and pleasant/aesthetic color principles

Where does one principle end and another begin? For anthropomorphisms, a key defining feature is that the anthropomorphic elements should be depicted on an existing, non-anthropomorphic graphical object (i.e., it is *not* a new graphical object). Therefore, manipulating the presence/absence of a pedagogical agent (e.g., Moreno, Mayer, Spires, & Lester, 2001) is not an anthropomorphism manipulation. Comparisons of different agent appearances (e.g., changing a neutral agent appearance to a positive appearance; see Domagk, 2010; also see Guo & Goh, 2016; Liew, Zin, Sahari, 2017), would also not count unless one lacks anthropomorphic features.

Another important dividing line lies in the targeted mental processes. This is especially important for color. For example, there has been a long history of research on using color for signaling purposes and for increasing realism of the instructional materials (e.g., Dwyer, 1971; reviewed in Dwyer & Lamberski, 1982; Pett & Wilson, 1996). Although not all applications of color for these purposes facilitate learning (see Dwyer & Lamberski, 1982), color can aid intermediate processes, such as helping learners locate corresponding parts of text and picture if in the same color (e.g., Oczelik, Karakus, Kursun, & Cagiltay, 2009). Color can also act as a memory cue (e.g., see Kanner, 1968). For example, color coding can help retention during learning of nonsense syllables (Van Buskirk, 1932).

We do not consider these manipulations as reflecting *emotional design* because they primarily target attention-cognitive-memory processes whereas *emotional design* should target *emotions,* particularly positive emotions (cf. Stark, Brünken, & Park, 2018, p. 187). Therefore, we exclude negative color manipulations (e.g., Kumar et al., 2016) and random color manipulations. The latter were frequently employed in older research examining the cost-effectiveness of switching from monochrome instructional films to their colorful counterparts (e.g., Kanner & Rosenstein, 1960). These studies generally produced null/mixed results with respect to achievement (see Kanner, 1968; Dwyer & Lamberski, 1982), likely because the colors tended to be superfluous features of the presentation medium rather than an instructional feature targeting specific emotional or motivational processes (cf. Dwyer & Lamberski, 1982).

1.4. Current study

The effects of facial anthropomorphisms and/or pleasant/aesthetic colors on both affective-motivational states and learning outcomes have been mixed (see above). Therefore, we used meta-analytic techniques to provide some clarity on the effectiveness of these two manipulations. We hypothesized that there would be positive overall effects of pleasant colors, facial anthropomorphisms, and a combination thereof, on learning outcomes (Hypothesis 1) and affective-motivational variables (Hypothesis 2).

We note that the meta-analysis cannot identify the precise mechanism by which colors/anthropomorphisms influence learning outcomes because available studies generally do not conduct mediation or similar analyses (but see Heidig et al., 2015; Münchow et al., 2017; Um et al., 2012). The results can, however, inform the search for possible mechanisms. In particular, a positive effect of colors/anthropomorphisms on both learning outcomes and affective-motivational states would suggest that these manipulations *may* influence learning through affective-motivational states (though influence through hidden variables is still possible). If we assume that affective-motivational states are positively related to learning outcomes, then a positive effect of the manipulations on the former but not the latter would suggest that the manipulations also increased levels of a third variable (possibly extraneous cognitive load), which countered the beneficial effects of the affective-motivational states.

If the manipulations increase learning outcomes, but not affective-motivational states, this would suggest that they act as classic cognitive, rather than emotional design, manipulation. For example, people learn better when learning materials include cues that highlight or help organize relevant information (see Schneider, Beege, Nebel, & Rey, 2018; van Gog, 2014). Because color can be used as such a cue (see van Gog, 2014), a beneficial effect of color might result from a reduction in cognitive load rather than an increase in positive affective-motivational states (despite this not being the intention). Similarly, eyes and gaze from facial anthropomorphisms (see Figure 1) can also serve as signaling cues given that faces can convey complex socially-relevant information (see Frischen, Bayliss, & Tipper, 2007; see also van Wermeskerken & van Gog, 2017). Anthropomorphisms are ostensibly also more salient and more memorable compared to schematic baselines.

Across-the-board null effects might imply that these augmentations were too subtle to influence learning processes – at least based on existing measures, which might be insensitive to subtle changes (e.g., the self-reports used to measure affective-motivational states). It could also suggest moderation. One obvious moderator is the experimental contrast: the difference between experimental and control condition (i.e., color alone versus anthropomorphisms alone versus a combination of both). Another is age as these augmentations can be considered “cute” and could yield more positive effects for young children compared to older learners. It is also possible that the visual emotional design augmentations would have different effects on participants with different language/cultural backgrounds based on some parallel evidence regarding the personalization principle (Brom, Děchtěrenko et al., 2017). Time-on-task might also be a moderator since the positive effect of the manipulations might dissipate over longer study periods (novelty effect). Alternatively, schematic, monochromatic baselines might be unbearably boring as time-on-task increases. Pacing (system-paced versus self-paced), dynamicity (static versus dynamic content), and prior knowledge (low versus high) are potential moderators because these factors can influence cognitive load (see Mayer, 2014a), thereby interfering with the effectiveness of emotional design augmentations. Finally, prior mood is a potential moderator because emotional design manipulations may be less effective for learners already in a positive mood (i.e., a ceiling effect). We investigate the effect of these and other methodological moderators as an exploratory aim.

2. Method

2.1. Search process, inclusion/exclusion criteria, and sample coding

We searched Google Scholar, PsychINFO, ERIC, and ScienceDirect using the query: ("emotional design") AND ("learning gains" OR "learning gain" OR "posttest" OR "post-test" OR "learning outcome" OR "learning outcomes"). In additional searches, the term "emotional design" was replaced by the following terms: a) (("anthropomorphisms" OR "anthropomorphism") AND "multimedia learning"), b) ("pleasant colors" OR "pleasant color" OR "aesthetic colors" OR "aesthetic color") AND "multimedia learning"), c) ("pleasant colours" OR "pleasant colour" OR "aesthetic colours" OR "aesthetic colour") AND "multimedia learning") (the search terms (b) and (c) differ only in British vs. US spelling of "color"). This yielded 15 articles/manuscripts/theses that satisfied the inclusion criteria (see below). We identified an additional four studies by examining the reference lists of the 15 articles, searching for further studies that cited the published studies, and by reaching out to relevant researchers. The search was limited to manuscripts published since January 1990. The search ended on January 18, 2018 with the exception that one additional study was published during the peer review process of this manuscript. Altogether, this meta-analysis included 20 manuscripts.

We included studies that:

1) were based on (quasi-) experimental design, i.e., they had a control condition;
2) manipulated facial anthropomorphisms and/or pleasant/aesthetic colors in at least one factor. We included studies that made "minor" alterations to other elements as part of the manipulation (rather than alterations related to an established design principle like seductive details or the personalization effect). We considered the following to be "minor": font type; sharpness/roundness of edges/corners/user interface elements; the level of the graphical elements' humor in two small images (neutral versus positive) in one study (Brom et al., 2016); facial expressions in human characters included in both compared versions (neutral versus positive) in one study (Uzun & Yıldırım, 2018); additional arrows in the non-anthropomorphic version replacing gaze direction or visible movement direction in the anthropomorphic version (Brom et al., unpublished manuscript; Mayer & Estrella, 2014); a slight change in layout, but not sufficient to violate the spatial contiguity principle (Mayer, 2009, Ch. 7) in one condition and not the other (Haaranen et al., 2015; Nurminen et al., 2017). The following between-group changes were not considered "minor": addition of an agent image (e.g., Andrey, Brunisholz, Dos Reis, & Molinari, 2016), addition of background sound (e.g., Doolittle &

Altstaedter, 2009, Exp. 1; Uzun & Yıldırım, 2018, condition (4), see below), addition of extraneous elements (e.g., additional image; Doolittle & Altstaedter, 2009, Exp. 1), and change of style of the instructional texts (formal versus conversational) (e.g., Andrey et al., 2016);

3) targeted positive-activating rather than negative-activating or deactivating (cf. Pekrun & Linnenbrink-Garcia, 2012) affective-motivational states;
4) investigated learning outcomes or performance outcomes;
5) included statistics for computing standardized effect sizes or provided this information via email queries. We excluded studies, or some of their variables, if the authors did not respond to repeated queries.

We specifically excluded studies that:

1) did not use pleasant colors to target positive-activating affective motivational states (i.e., used negative colors to target negative states or random colors as a superfluous element) (e.g., Kanner & Rosenstein, 1960);
2) manipulated colors within the surrounding context (e.g., color of walls in the lecture room vs. a graphical element germane to the learning content) (e.g., Stone & English, 1998);
3) manipulated appearance of pedagogical agents (e.g., changing the agent's appearance from neutral to positive or manipulating the agent's gender) (e.g., Domagk, 2010).

We treated multiple experiments within each study as separate samples. Some studies included two- or three-factorial designs. If a factor was unrelated to the manipulation of interest in a between-subjects design, each level of the secondary factor was considered an independent sample. For example, Um and colleagues (2012) used a 2 (grey-scale schematic graphics versus colorful graphics with anthropomorphisms) × 2 (neutral versus positive prior mood induction) between-subject design. Here, the neutral and positive mood groups would be considered to be two samples. Some studies included additional experimental conditions which were irrelevant to the present question and were thus excluded. For example, Uzun and Yıldırım (2018) used four groups: (1) no-sound with a grey-scale non-anthropomorphic version (i.e., control); (2) no-sound with a colorful non-anthropomorphic version; (3) no-sound with a colorful anthropomorphic version; and (4) sound with a colorful anthropomorphic version. The last group (with sound) was excluded from the meta-analysis.

Altogether, we included 33 independent samples ($N = 2924$), though sample size varies as a function of the dependent variable. An overview of studies is provided in Supplementary Materials. We note that previous influential multimedia learning meta-analyses have used samples of similar size (*k*s of 15 – 43 - e.g., Gegenfurtner, Quesada-Pallarès, & Knogler, 2014: digital simulation-based training; Ginns et al., 2013: personalization principle; Ginns, 2005: modality effect; Takacs, Swart, & Bus, 2014: children's comprehension of stories presented by multimedia versus print-like materials).

2.2. Coding of study variables

We coded the following study variables based on the consensus of two coders:

a) *Experimental contrast*. We coded this variable as a contrast between experimental and control conditions. We used four categories: color (i.e., only pleasant colors were added in the experimental condition), anthropomorphisms (i.e., only facial anthropomorphisms were added), both (i.e., both pleasant colors and facial anthropomorphisms were added), and combined (i.e., when we collapsed experimental conditions with different manipulation types – see Section 2.4). For example, Münchow and colleagues (2017) compared grey-scale hypermedia materials to a colorful version. This was coded as "color". Park and colleagues (2015) compared two colorful versions of an animation, one with and the other without anthropomorphisms. This was coded as "anthropomorphisms". Mayer and Estrella (2014) compared black-and-white, non-anthropomorphic slides to their colorful counterparts with anthropomorphisms; this was coded as “both”. Finally, in studies with multiple experimental groups and a single control group, the experimental groups were collapsed (this was required to preserve independence – see Section 2.4). The experimental contrast of these studies was coded as follows. It was coded as “color” in the study by Heidig and colleagues (2015), which included multiple experimental conditions (and a single control group) to investigate the effect of various color combinations (and a usability factor irrelevant for present purposes). It was coded as “anthropomorphisms” in the study by Schneider et al. (2018b), which included two experimental conditions (moderate and high levels of anthropomorphisms) and a single no-anthropomorphisms group. Finally, other studies with multiple experimental conditions used different manipulation types in their experimental conditions (e.g., colorful non-anthropomorphic version in group 2 and colorful anthropomorphic version in group 3 in the study by Uzun and Yıldırım, 2018; see Section 2.1). These were therefore coded as “combined”.

b) *Topic*. After preliminary screening, we collapsed the topics into two broad categories: natural sciences (primarily biology) and technical content (computer science, engineering, and mathematics). The topic for one study (Miller, 2011: American sign language) could not be accommodated within these categories, so topic was not coded for this study.

c) *Age*. Due to an insufficient number of studies with younger students, we created three categories: 1) college students with backgrounds primarily in psychology and/or educational sciences; 2) other college students (two study samples with high school students 17-18 years of age, i.e., Grade 11 and higher, was included here: Brom et al., 2016; Schneider et al. 2018b, Exp. 3); 3) younger students (<16 years old; up to Grade 10).

d) *Language/Culture*. Based on where the experiment was conducted, we coded for German, US, Chinese, and Czech. Two Finnish samples, two Malaysian samples, and one Turkish sample were included into an "other" group.

e) *Time-on-task.* We coded studies as (1) short (up to 15 min), (2) medium (15 – 40 min), and (3) long (> 40 min); however, as only one study was longer than 40 minutes, the latter two categories were collapsed.

f) *Pacing*. We distinguished between system-paced versus self-paced materials.

g) *Dynamicity*. We coded text-and-picture slides as static and all other cases, such as animations or hypermedia with animations, as dynamic.

h) *Prior mood manipulation.* We coded studies as positive mood manipulation, neutral mood manipulation, and no mood manipulation.

i) *Publication type*. We coded whether the study was published in a peer-reviewed journal article or "other" (a thesis, a conference paper, an unpublished manuscript).

We did not code for prior knowledge because only nine samples (27%) used an actual knowledge pre-test; the remainder used perceived prior knowledge questionnaires or a combination of pre-test and perceived knowledge questions.

2.3. Coding of dependent measures

We coded three broad categories of dependent measures: learning outcome measures, affective-motivational variables, and learner perception variables.

Learning outcomes. We coded *retention* (memorizing key facts; e.g. "Reproduce what you saw in the animation on how a biological wastewater treatment plant works."), *comprehension* (understanding of key concepts; e.g., "Why are azo dyes dangerous?"), and *transfer* (ability to use learnt concepts in new situations; e.g., "What would happen if a fungus first appeared in the treatment plant and then bacteria? Write down all consequences that come to mind."). Coding mostly followed the primary authors' classification of their learning measures with the following exceptions: Ng and colleagues (2017) measured "problem-solving performance" – this was coded as comprehension; Miller (2011) investigated memorization of sign language – this was coded as retention; Nurminen and colleagues (2017) used a combined retention + transfer measure – this study was excluded from the analysis pertaining to learning outcomes. Most studies provided immediate post-test scores, but a handful also provided delayed scores ($k = 2$; 6.06%) or immediate learning gains (i.e., post – pre; $k = 5$; 15.15%); these latter measures were not analyzed.

Affective-motivational states. Some meta-analyses have combined various affective-motivational measures under an umbrella label, for example, motivation (e.g., Wouters et al., 2013). Because there is some evidence that different affective-motivational states may be differentially related to learning (Brom, Děchtěrenko et al., 2017), we distinguished among generalized positive affect, intrinsic motivation, and liking/enjoyment. Due to an insufficient number of studies with available data, we did not code flow ($k = 3$, 9.09%) and attention/concentration ($k = 3$, 9.09%).

Generalized positive affect is related to various positively-valenced, activating feelings; for instance, feelings of excitation, activity, attentiveness, or enthusiasm (Watson & Tellegen, 1985). It was primarily measured with the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) or a short version of "Scales for Assessing Positive/Negative Activation and Valence in Experience Sampling Studies" (PANAVA-KS; Schallberger, 2005). Generalized positive affect was usually measured before and after the learning session, resulting in post and change (post – pre) measures. We analyzed the change measure as it was reported more often ($k = 15$ vs. 11) than the post measure.

Intrinsic motivation refers to "doing something because it is inherently interesting or enjoyable" (Ryan & Deci, 2000; p. 55). In the included studies, intrinsic motivation was measured by established questionnaires (e.g., Isen & Reeve, 2005) or researcher-created short

(1-3 item) questionnaires, generally probing the desire to continue studying with the present materials and/or to use similar materials in the future in a different domain.

Liking/enjoyment can be defined as an activity-related affective state experienced when the learning activity or materials are positively valued and when the activity is sufficiently controllable by the learner (Pekrun, 2006; p. 323). It was typically measured by researcher-created, 1-3 item questionnaires assessing appeal, likability, and/or enjoyment of the lesson/materials.

Learning perception variables. We coded *perceived effort* (i.e., how much effort the learners thought they invested into learning), *perceived difficulty* (i.e., how difficult did learners find the topic/learning/materials), and *perceived learning* (i.e., how much learners thought they learnt). The first two are proxies of cognitive load, or its subcomponents (DeLeeuw & Mayer, 2008), but there are arguments that they may not be appropriate for this purpose (de Jong, 2010). Unfortunately, contemporary cognitive load questionnaires (e.g., Leppink et al., 2014) or objective methods, such as dual-task measures (Brünken, Seufert, Paas, & 2010), have rarely been used in emotional design research (see Schneider et al. 2018b, for an exception). These three variables were selected because they were reported in multiple studies.

2.4. Data treatment

We used the metafor package (Viechtbauer, 2010) in R for all analyses. We computed the standardized mean effect (SDM – Cohen's *d*) for treatment vs. control; positive effects indicate an advantage of emotional design manipulations over control. We chose to work with *d* rather than Hedge's *g* because the individual studies had adequate sample sizes (Borenstein, Hedges, Higgins, & Rothstein, 2009). If a study had multiple experimental groups (e.g., Uzun & Yıldırım, 2018: see Sec. 2.1), but a single control group, these groups were collapsed by computing pooled statistics prior to computing the effect sizes. This is the recommended approach because it preserves independence and prevents inflating the sample size (because the control group is only counted once) (Borenstein et al., 2009). This was done for three studies with two experimental groups (Brom et al., unpublished manuscript, sample with eye-tracker as a factor; Schneider et al., 2018b; Uzun & Yıldırım, 2018); two others with three experimental groups (Plass et al., 2014, Exp. 2; Gong et al., 2012, Exp. 2); and one with eight experimental groups (Heidig et al., 2015). One study (Mayer & Estrella, 2014, Experiments 1 and 2) separately reported results for appeal and enjoyment; these were combined following guidelines for combining dependent measures (we assumed a correlation of .5 among the two measures) (Borenstein et al., 2009).

3. Results

We used random effect models to estimate the meta-analytic effect and to model heterogeneity in the effect size distributions. Publication bias was assessed using the rank correlation test (Begg & Mazumdar, 1994), Egger's (1997) regression (using standard errors as the predictor), and via visual inspection of funnel plots. When publication bias was identified, we used a trim-and-fill analysis (Duval, 2005) to estimate the number of missing studies and adjust the meta-analytic effect accordingly. Table 1 provides a summary of the results (k refers to the number of samples and d_+ is the meta-analytic estimate). Unless indicated otherwise, we used two-tailed tests with an alpha of .05.

Table 1. Summary of random-effects models

	Sample		Meta-analytic estimate (d_+)			Heterogeneity			Publication Bias	
Measure	***k***	***n***	**Estimate (SE)**	**95% CI [LB, UB]**	***Z***	***Q* (df)**	τ^2	I^2 **(%)**	**Egg. *Z***	**K. tau**
Learning Outcomes										
Retention	18	1759	0.387 (0.107)	[0.177, 0.597]	3.61**	65.9 (17)**	0.149	74.4	1.69†	0.281
Comprehension	14	1404	0.317 (0.065)	[0.190, 0.444]	4.89**	13.6 (13)	0.008	14.5	1.68†	0.275
Transfer	27	2281	0.327 (0.063)	[0.203, 0.452]	5.17**	47.9 (26)**	0.046	45.0	2.08*	0.219
Transfer $_{\text{trim-and-fill}}$	28	-	0.316 (.064)	[0.191, 0.440]	4.97**	51.1 (27)**	0.048	45.3	-	-
Affective-motivational										
Liking/Enjoyment	20	1474	0.109 (0.053)	[0.005, 0.212]	2.06*	21.9 (19)	0.000	0.01	0.19	0.042
Positive affect	15	1407	0.113 (0.060)	[-0.005, 0.232]	1.88†	16.6 (14)	0.002	4.50	1.10	0.181
Intrinsic motivation	23	2023	0.255 (0.086)	[0.086, 0.424]	2.95**	67.0 (22)**	0.109	66.4	0.76	0.178
Learning perceptions										
Perceived effort	20	1215	0.051 (0.143)	[-0.228, 0.331]	0.36	80.2 (19)**	0.319	81.4	0.39	0.105
Perceived difficulty	14	967	-0.208 (0.074)	[-0.353, -0.063]	-2.80**	15.9 (13)	0.014	17.9	-1.30	-0.165
Perceived learning	11	739	0.097 (0.080)	[-0.060, 0.254]	1.21	11.6 (10)	0.006	8.46	0.603	0.055

***p* < .01; **p* < .05; †*p* < .10; Positive affect = Generalized positive affect gain; Egg. = Egger's regression; K. tau = Kendall's tau rank correlation coefficient.
k = number of samples; *n* = sample size across studies.

3.1. Meta-analytic effects

Learning outcomes. There were statistically significant positive effects for retention ($k = 18$, $d_+ = 0.387$; Figure 3), comprehension ($k = 14$, $d_+ = 0.317$; Figure 4), and transfer ($k = 27$, $d_+ = 0.327$; Figure 5). There was significant heterogeneity for retention and transfer but not for comprehension ($p = .406$). Further, the I^2 statistic, which measures between-study variability due to heterogeneity (residual heterogeneity) vs. chance (unaccounted variability), indicated substantial heterogeneity for retention (74.4%) and transfer (45.0%), and a lower amount for comprehension (14.5%).

Egger's regression was marginally significant for retention ($p = .092$) and comprehension ($p = .094$) and was significant for transfer ($p = .038$). The rank correlation test was nonsignificant for all three outcomes ($p > .112$). These results suggest that publication bias was unlikely for retention and comprehension, but possible for transfer. A trim-and-fill analysis resulted in the addition of one study for transfer (Figure 6), but the meta-analytic estimate remained significant $d_{+(\text{trim-and-fill})} = 0.316$ and was similar to the previously reported effect ($d_+ = 0.327$). Thus, Hypothesis 1 was supported.

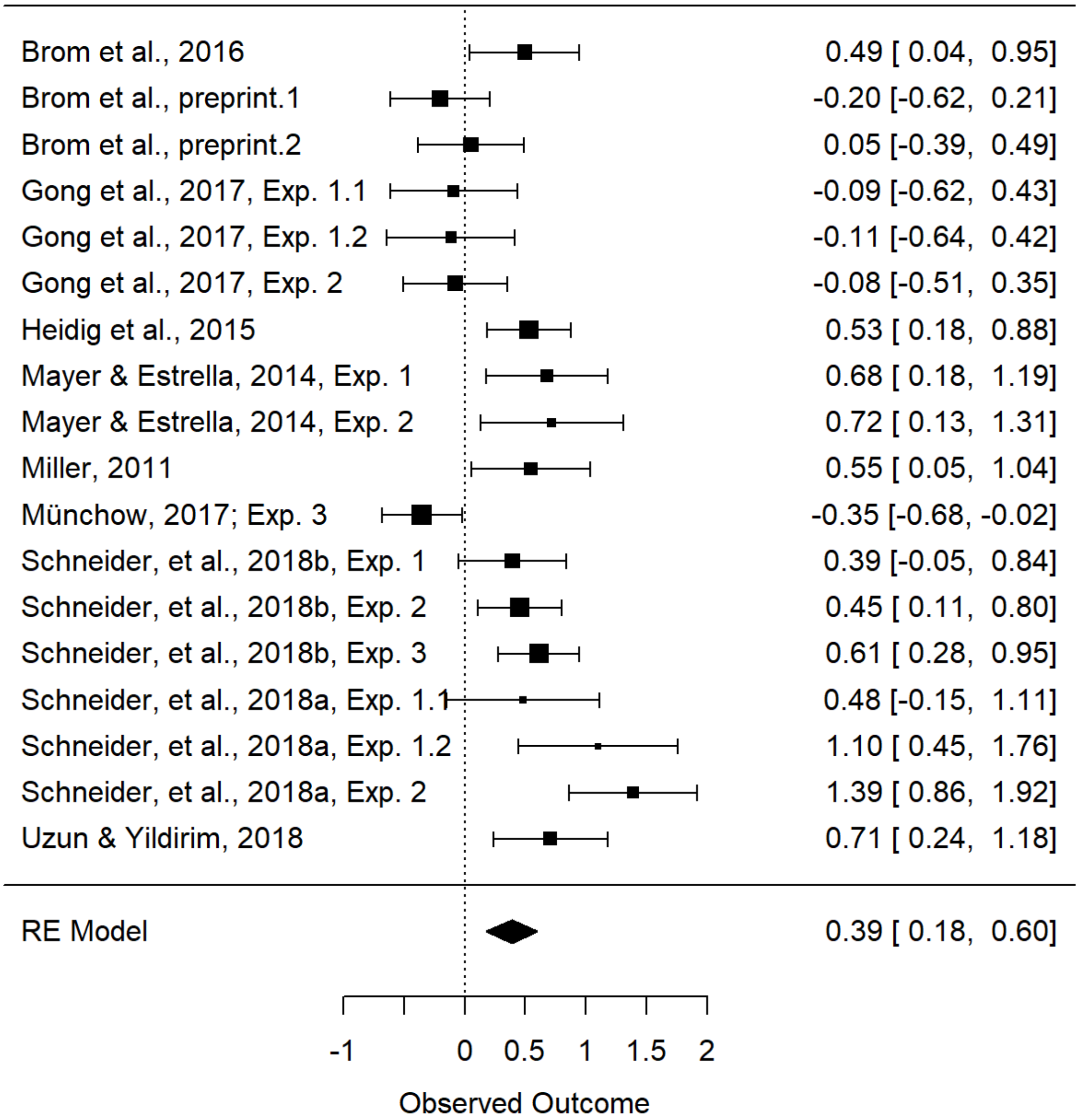

Figure 3. Forest plot for retention.

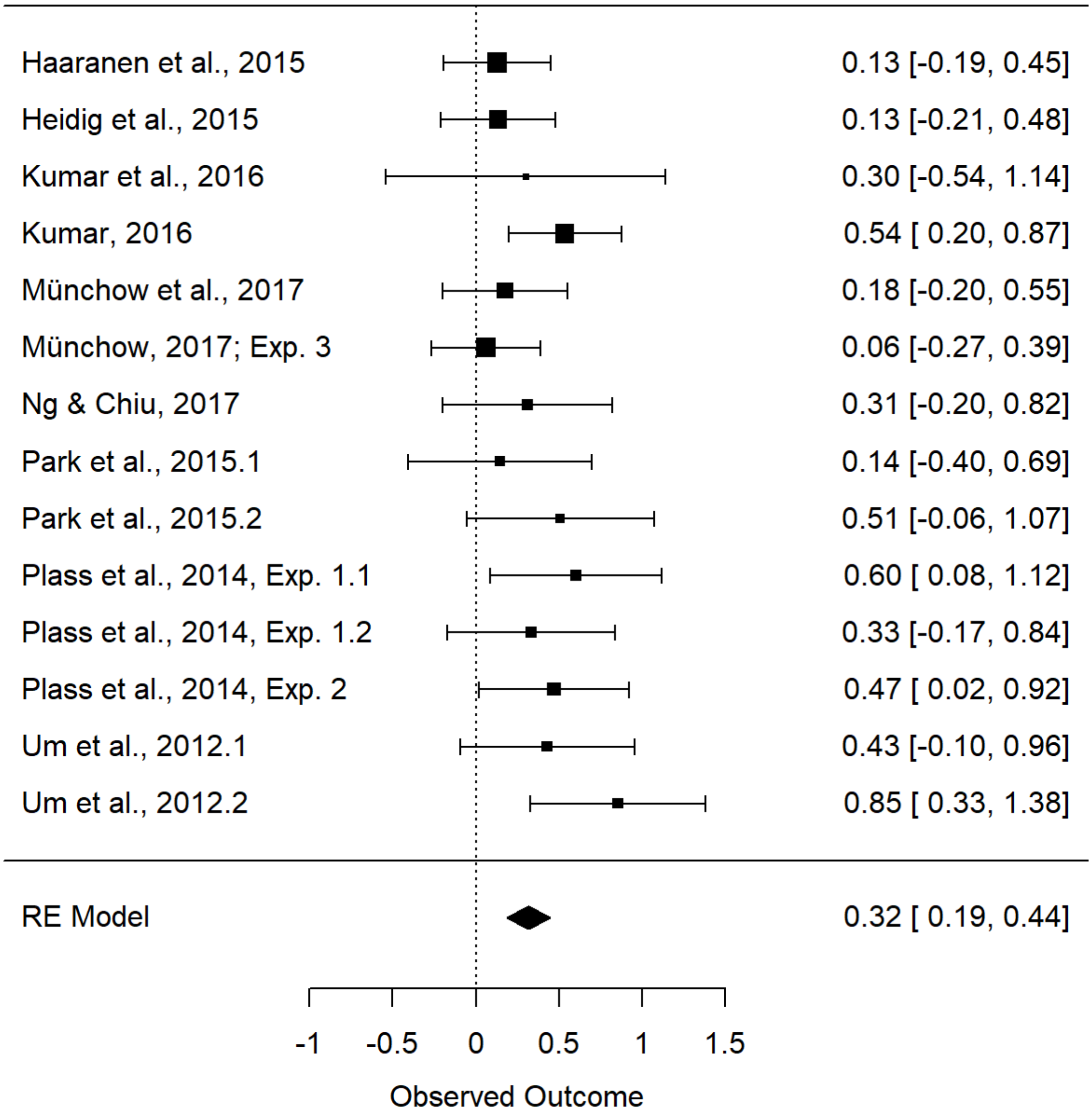

Figure 4. Forest plot for comprehension.

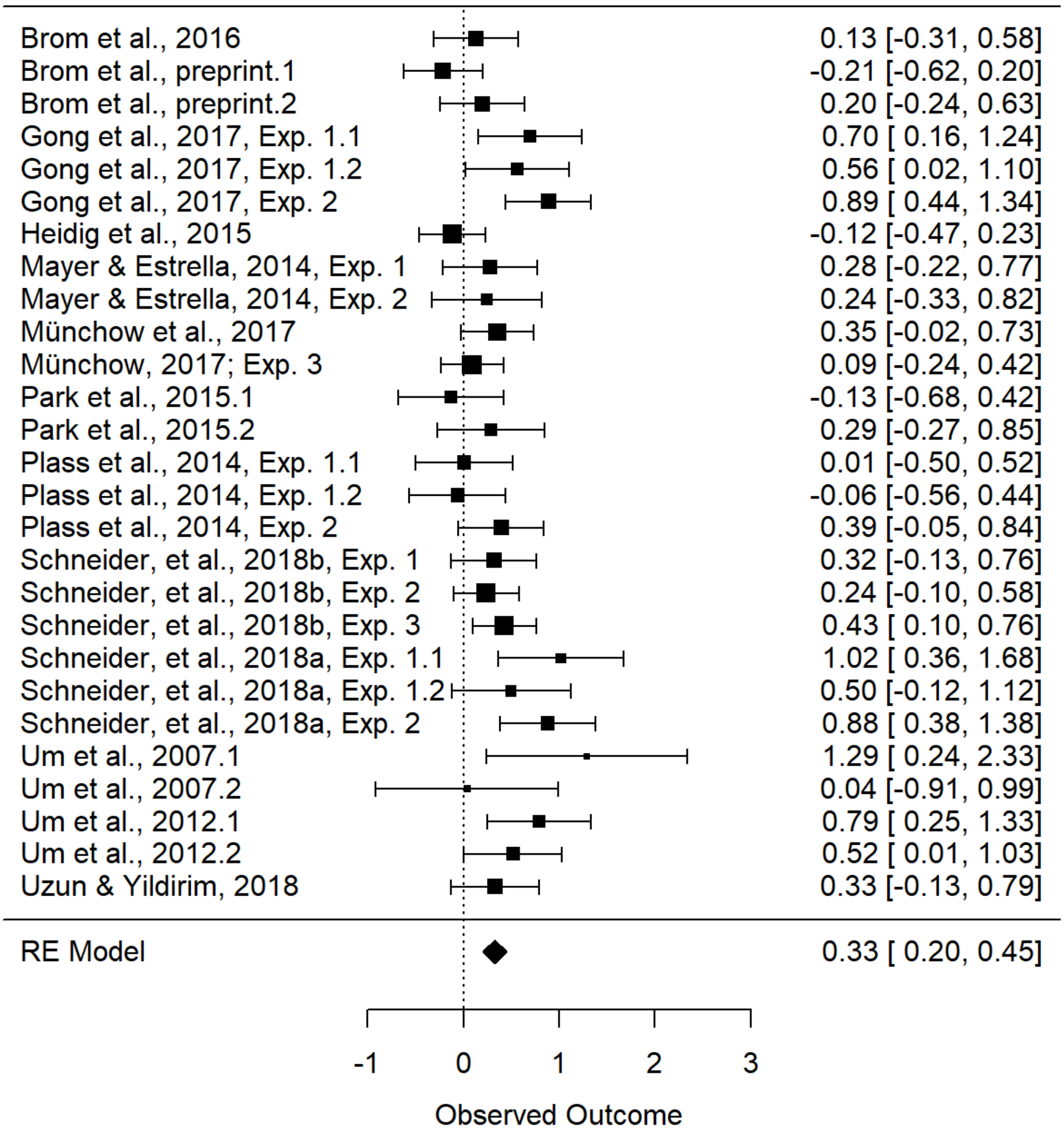

Figure 5. Forest plot for transfer.

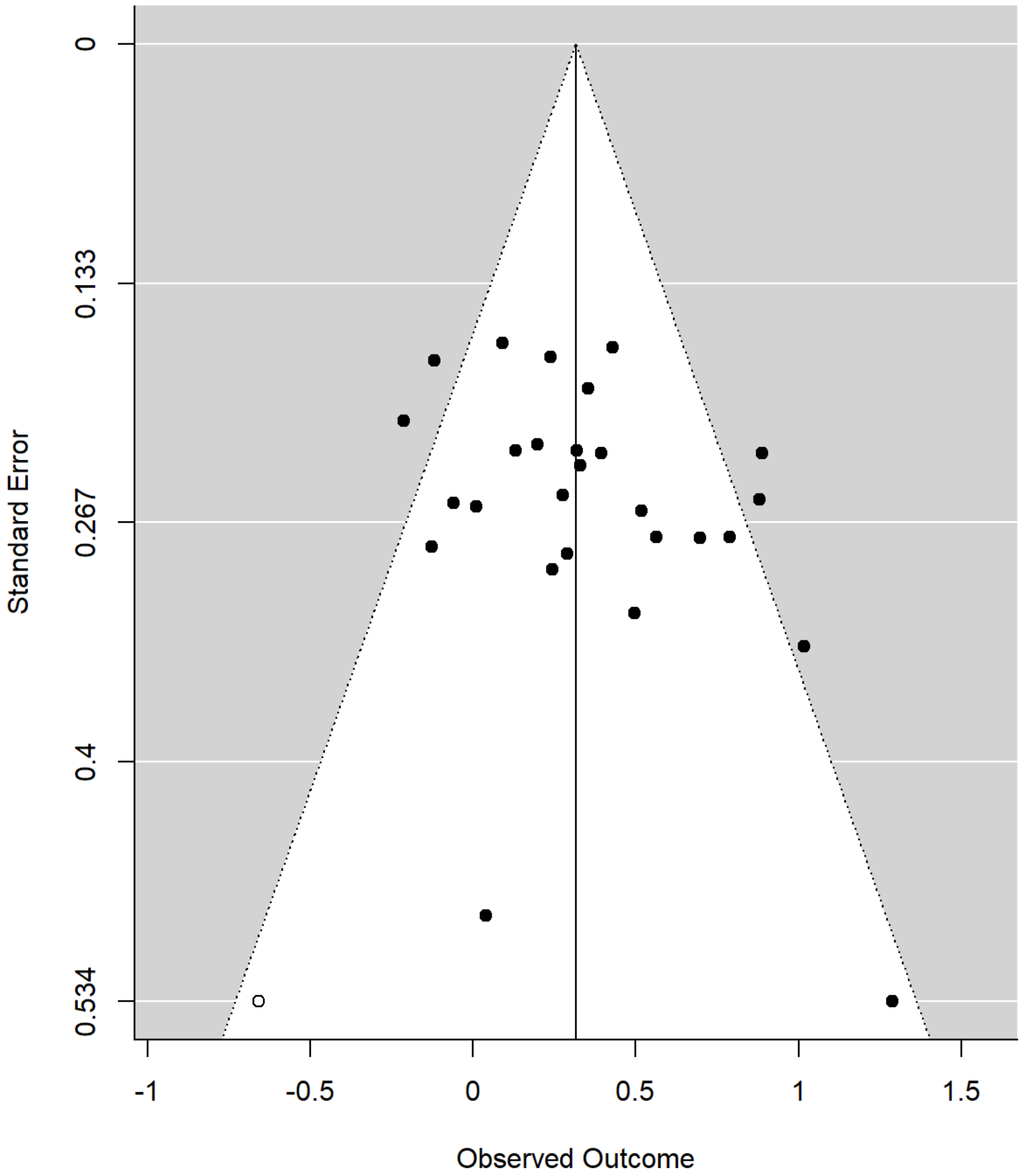

Figure 6. Funnel plot with one study (empty circle) added via trim-and-fill analysis for transfer.

Affective motivational variables. There was a significant positive meta-analytic effect for liking/enjoyment (k = 20, d_+ = 0.109; Figure 7), but the test for heterogeneity was not significant (p = .289) and the I^2 was basically zero (0.006%). Similarly, the effect for generalized positive affect gain was marginally positive (k = 15, d_+ = 0.113; p = .061; Figure 8), but was not significantly heterogeneous (p = .279; I^2 = 4.50%). There was a significant effect for intrinsic motivation (k = 23, d_+ = 0.255; Figure 9), and the test for heterogeneity was also significant with an I^2 of 66.4%. Neither Egger's regression (ps > .269) nor the rank correlation test (ps > .247) were significant for any of these variables, suggesting that publication bias was unlikely. Thus, Hypothesis 2 has limited support for liking/enjoyment and generalized positive affect (due to weak effects) and more robust support for intrinsic motivation.

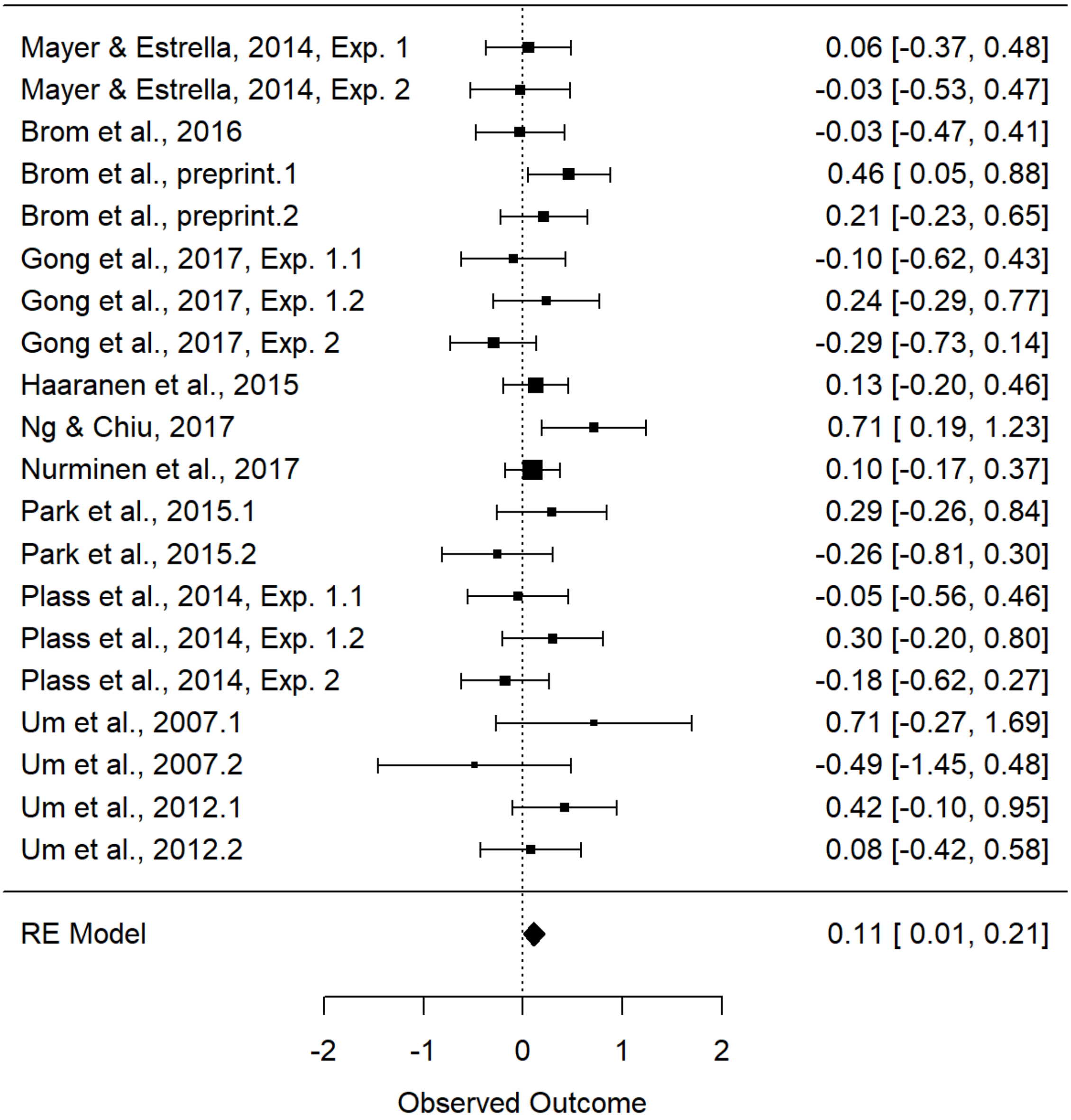

Figure 7. Forest plot for liking/enjoyment.

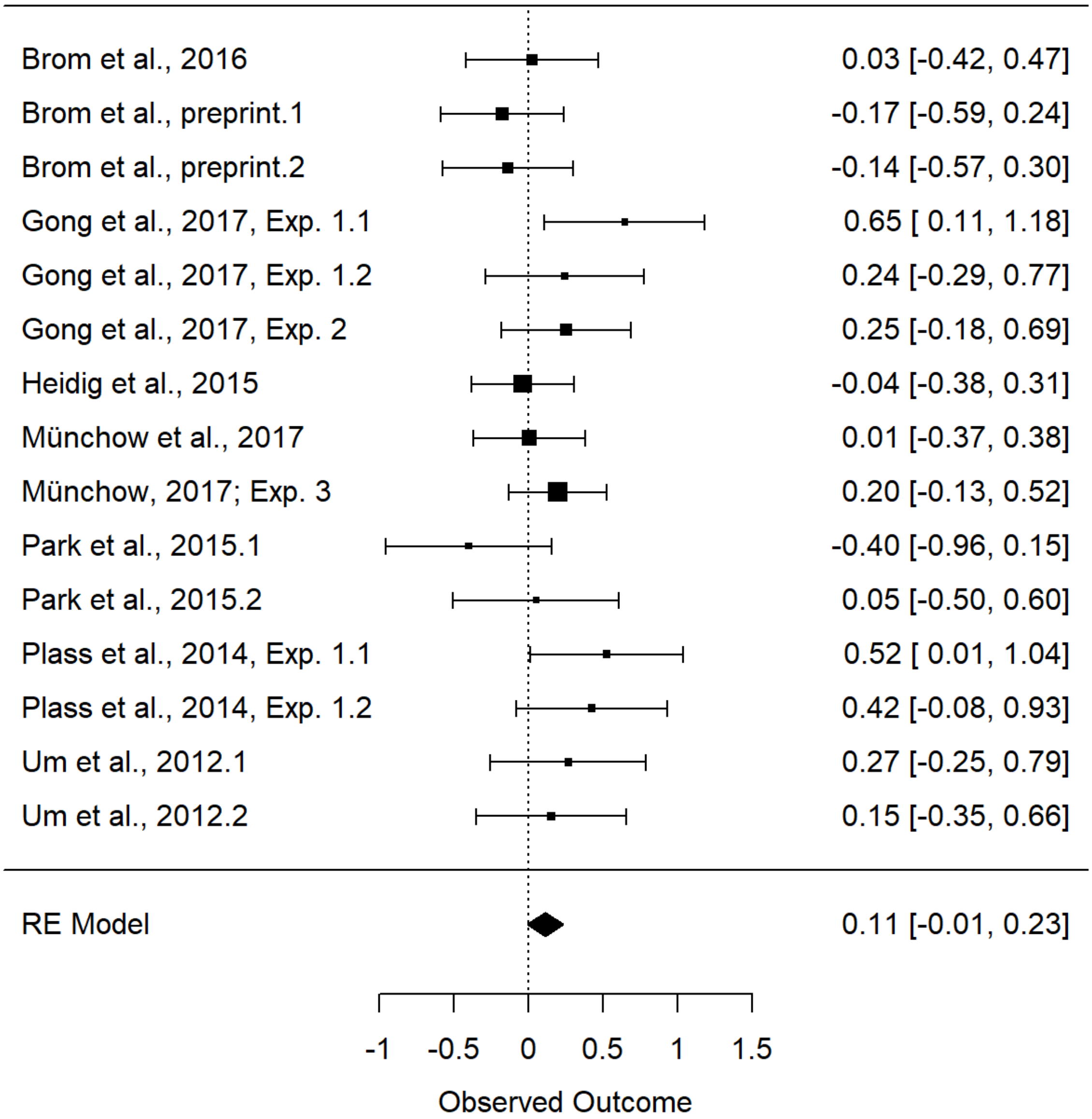

Figure 8. Forest plot for generalized positive affect gain.

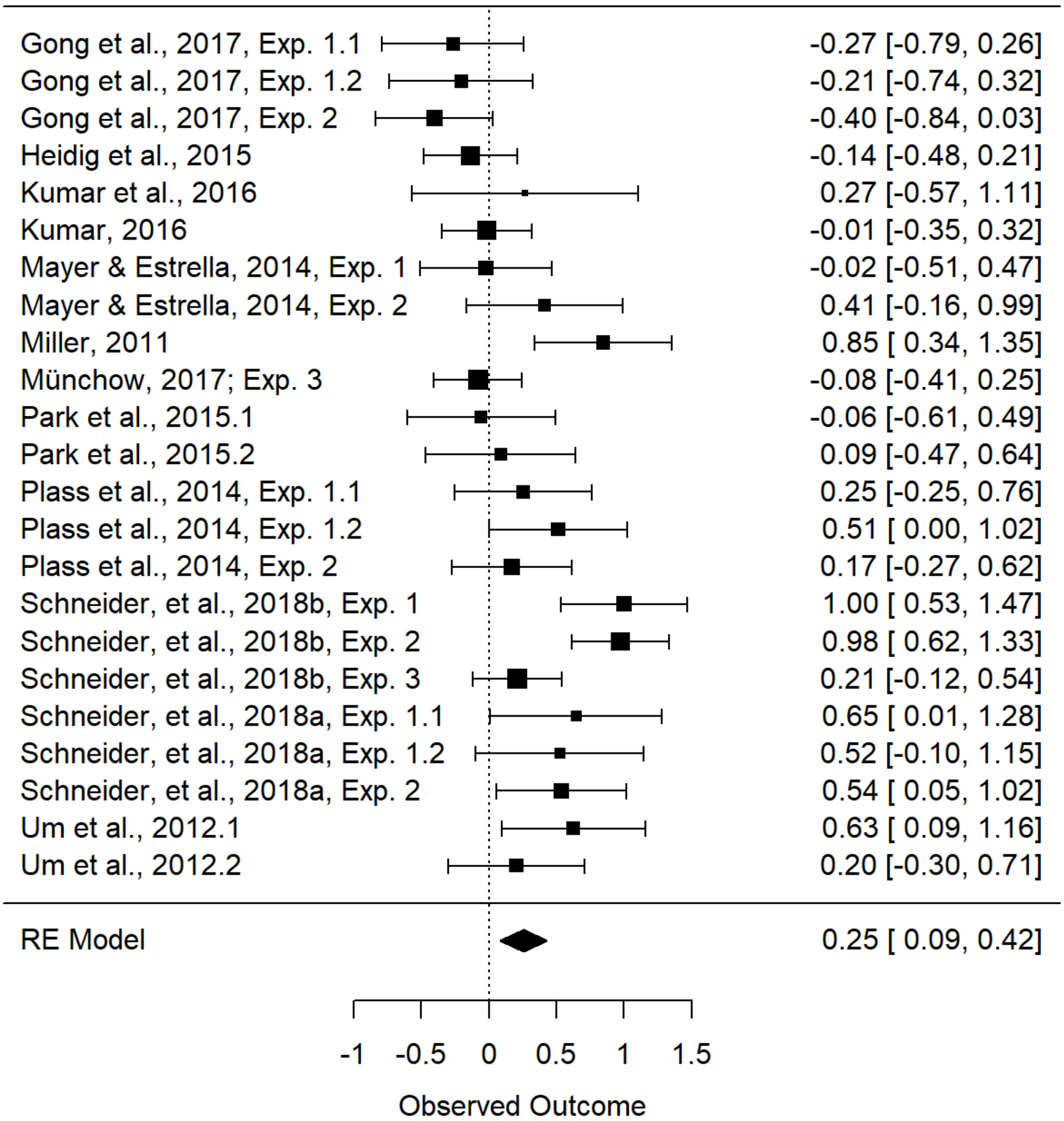

Figure 9. Forest plot for intrinsic motivation.

Learning perceptions. The meta-analytic effect for perceived effort (k = 20, d_+ = 0.051; Figure 10) was not significant (p = .719), but there was significant and substantial heterogeneity (I^2 = 81.4%). In contrast, there was a significant positive meta-analytic effect for perceived difficulty (k = 14, d_+ = -0.208; Figure 11), but the test for heterogeneity was not significant (p = .256; I^2 = 17.9%). There was no significant effect for perceived learning (k = 11, d_+ = 0.097; p = .227; Figure 12), nor was there sufficient heterogeneity (p = .313; I^2 = 8.46%). Egger's regression (*ps* > .194) and the rank correlation test (*ps* > .451) were non-significant for all three learning perception measures, suggesting publication bias was unlikely.

Influential case analyses. We used leave-one-out (LOO) analyses and DFBETAS (assessing a change in standard deviation of estimate after eliminating each study) to identify influential cases that might have biased the results. Specifically, we compared the mean meta-analytic effect with all studies included to the minimum and maximum mean effects obtained after removing each study. The LOO analyses did not suggest any overly influential cases for any of the measures and all DFBETAS were < 1 (Table 2).

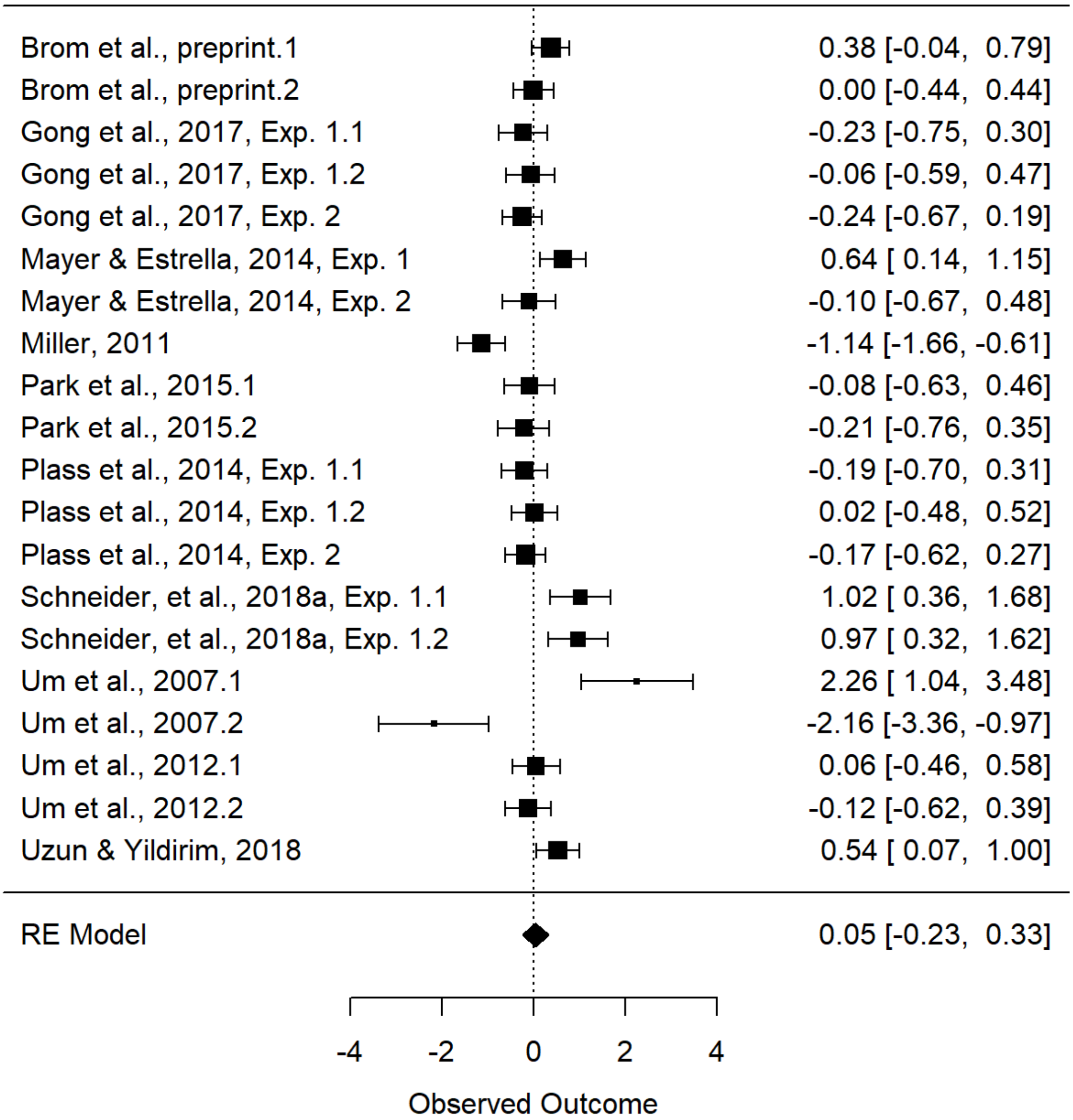

Figure 10. Forest plot for perceived effort.

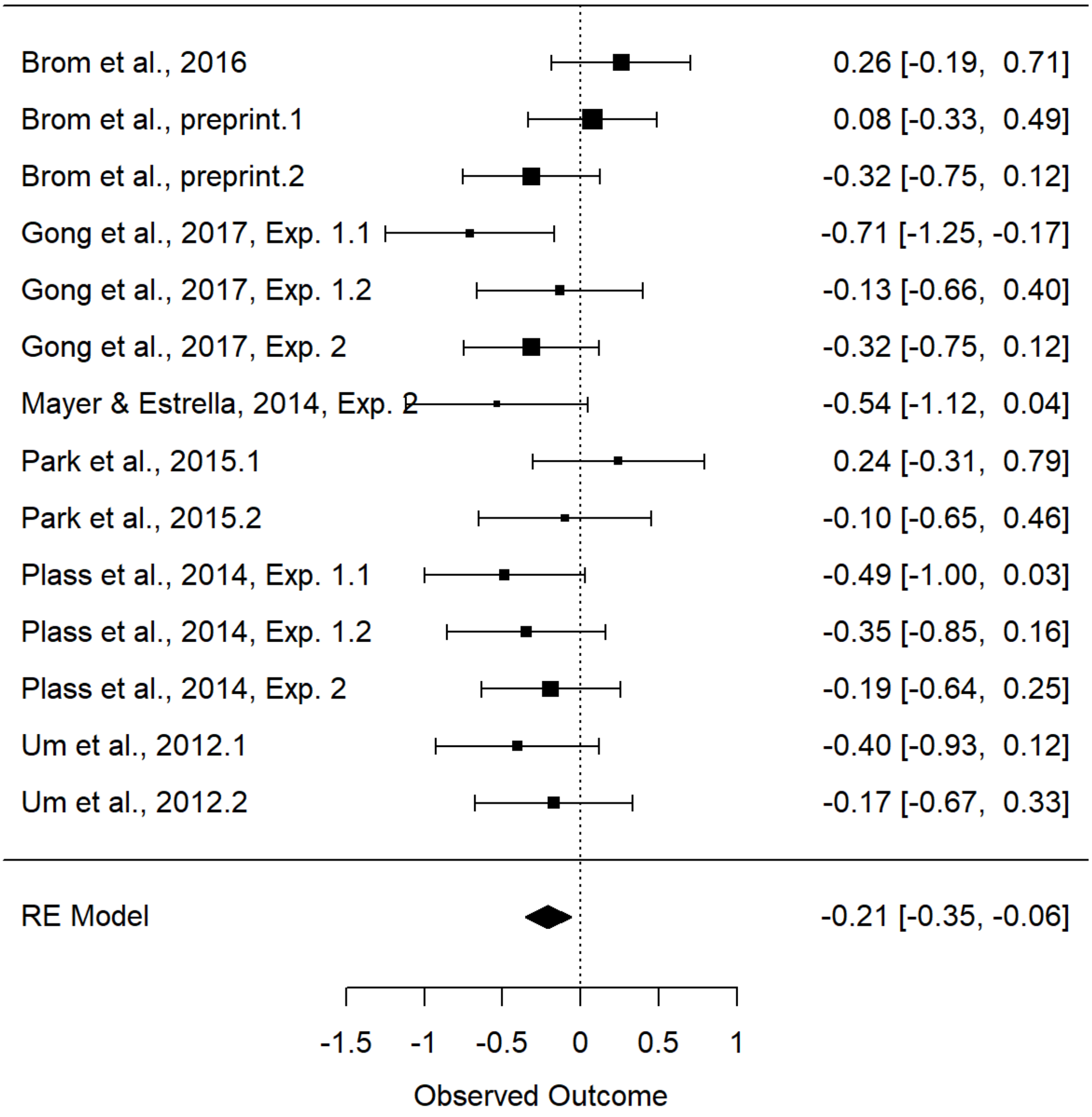

Figure 11. Forest plot for perceived difficulty.

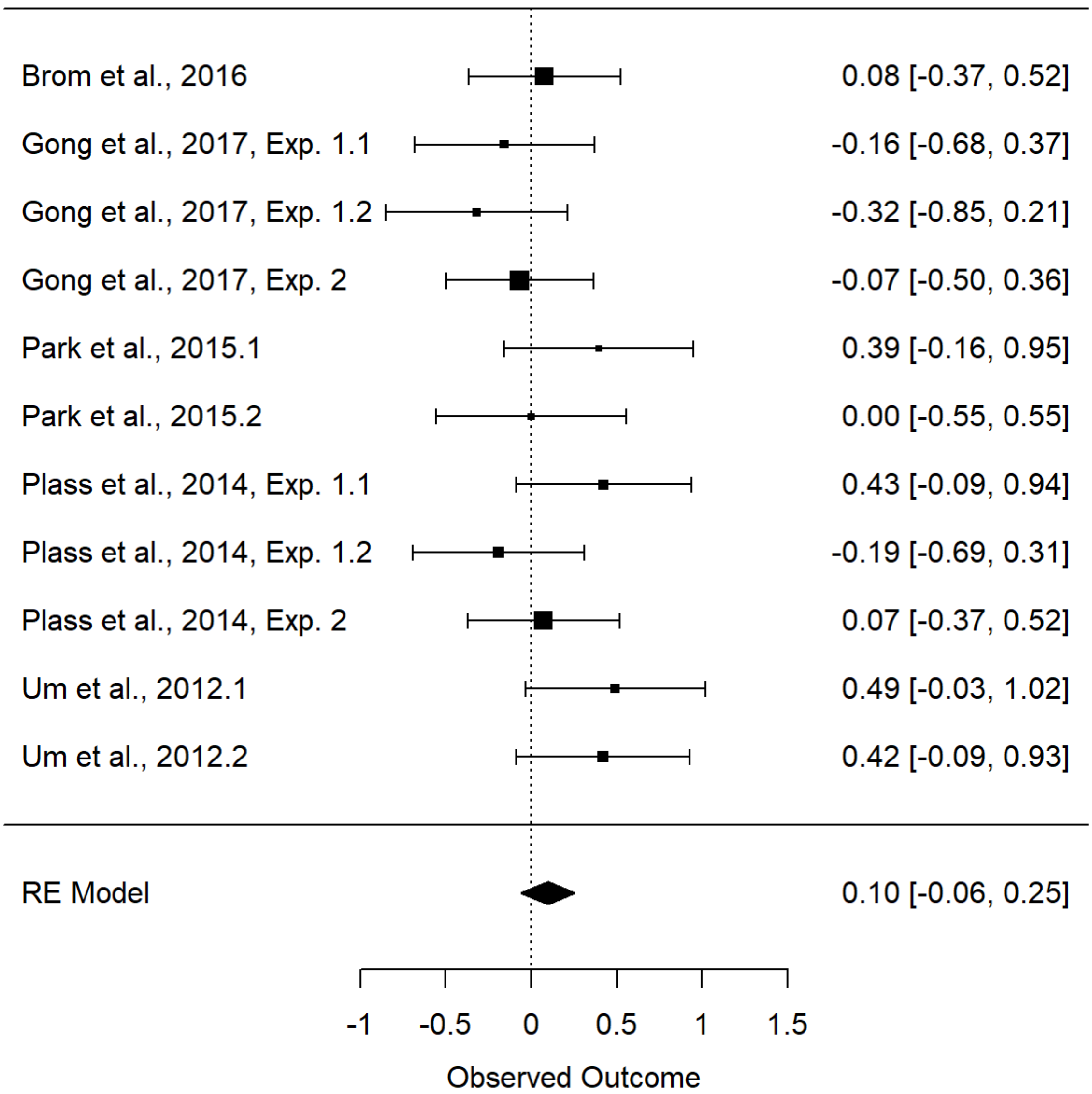

Figure 12. Forest plot for perceived learning

Table 2. Influential case analysis

	Leave-one-out analysis			
Measure	**All studies**	**Minimum**	**Maximum**	**DFBETA**
Learning Outcomes				
Retention	0.387	0.328	0.436	0.623
Comprehension	0.317	0.281	0.347	0.607
Transfer	0.327	0.299	0.349	0.482
Affective-motivational states				
Liking/Enjoyment	0.109	0.084	0.133	0.467
Positive affect	0.113	0.087	0.137	0.455
Intrinsic motivation	0.255	0.211	0.287	0.555
Learning perceptions				
Perceived effort	0.051	-0.008	0.114	0.525
Perceived difficulty	-0.208	-0.248	-0.173	0.597
Perceived learning	0.097	0.061	0.132	0.468

Positive affect = Generalized positive affect gain; Minimum and maximum represent the meta-analytic mean effect obtained after leaving each study out.

3.2. Moderation analysis

The proportion of studies per moderator level is shown in Table 3. The studies were diverse in terms of experimental contrasts, language/cultural origin, and dynamicity, but the majority of studies used college students, tested natural science content, used interventions up to around 30 minutes, and involved self-paced interventions.

Due to lack of variability in the data, we did not conduct moderation analysis for pacing (only 12.1% of samples used system-paced instructional materials). Because all but two studies (6.06%) used random assignment to condition, experimental design assignment (random vs. quasi-experimental) was not included as a moderator. Although some studies considered instructional topic as a moderator (e.g., Schneider, Beege, et al., 2018; Wouters et al., 2013), we did not include it because there was no theoretical justification for this. We did, however, consider publication type (journal versus other) as a methodological moderator.

We used random effects models to examine the influence of seven moderators (experimental contrast, age, language/culture, time-on-task, dynamicity, prior mood, and publication type) on the four dependent variables with sufficient heterogeneity (retention, transfer, intrinsic motivation, and perceived effort). Due to the small number of samples (*ks* from 18 to 27), we considered each moderator individually rather than jointly testing all moderators using meta-regression. To prevent a Type I error (false positive error), we adjusted the *p* values of the moderator coefficients (Q_m) for each dependent variable (i.e., *p* values were adjusted for seven comparisons, one for each moderator, per dependent variable) using the false-discovery rate (FDR) adjustment (Benjamini & Hochberg, 1995).

With one exception, none of the moderators were significant (*ps* > .063) predictors of the outcomes. The only significant effect was that age was a significant predictor of intrinsic motivation $Q_m(2) = 14.8$, $p < 0.01$, such that the effects of anthropomorphisms/colors on intrinsic motivation were stronger for younger children ($d_+ = .855$, $SE = .174$) than college students (d_+ = .231, SE = .137 for college – psychology/educational and d_+ = .110, SE = .085 for other college students). However, this finding should be taken with considerable caution since only three (out of 23 or 13%) studies measuring intrinsic motivation included a younger sample and we did not conduct meta-regression.

Table 3. The proportion of samples per moderator level

Study feature	Levels	*k*	%
Experimental contrast	color only	5	15.2%
	anthropomorphisms only	11	33.3%
	both	13	39.4%
	combined	4	12.1%
Age	college - psychology/educational	6	18.2%
	college - other	22	66.7%
	younger	5	15.2%
Language/culture	U.S.	7	21.2%
	German	14	42.4%
	Chinese	4	12.1%
	Czech	3	9.10%
	other	5	15.2%
Topic	natural sciences	24	72.7%
	technical content	8	24.2%
	other	1	3.03%
Time-on-task	<=15 min	23	69.7%
	> 15 min	10	30.3%
Pacing	system-paced	4	12.1%
	self-paced	28	84.8%
	other	1	3.03%
Dynamicity	static	14	42.4%
	dynamic	19	57.6%
Prior mood induction	none	22	66.7%
	positive	5	15.2%
	neutral	6	18.2%
Publication type	journal	23	69.7%
	other	10	30.3%

4. Discussion

Is emotional design effective at improving learning and affective-motivational states? We investigated two emotional design principles – anthropomorphizing graphics and/or adding pleasant colors to multimedia learning materials – and found strong evidence that they improved learning, but evidence for improving affective-motivational outcomes was less conclusive. Specifically, whereas the effects on retention, comprehension, and transfer were small to medium in size (d_+ ranging from 0.317 – 0.387), the effects on affective-motivational variables were mixed, with a more reliable effect for intrinsic motivation (d_+ = 0.255) compared to liking/enjoyment (d_+ = .109) and generalized positive affect (d_+ = .113). The effect on intrinsic motivation was moderated by age (i.e., larger for children), but this finding should be interpreted with caution due to only three child samples. Despite the somewhat small sample size, there were limited concerns about publication bias, outliers, and influential cases, but some of the effect distributions were insufficiently heterogeneous. Thus, anthropomorphisms and/or pleasant colors enhance learning. Increased intrinsic motivation can be one of the reasons, though further research is needed to elucidate precise causal mechanisms.

We also did not find any effects for perceived effort and perceived learning. However, instructional materials enhanced with anthropomorphisms and/or pleasant colors were perceived to be less difficult (d_+ = -0.208), similar to what was found for the personalization principle (Ginns et al., 2013). It is not clear, however, whether these augmentations reduced cognitive load (see de Jong, 2010; DeLeeuw & Mayer, 2008 for a discussion on how perceived difficulty can relate to cognitive load) or if learners simply thought that learning from “nicer” materials was easier. To this latter point, a similar effect has been observed in human-computer interaction research where users perceive aesthetically pleasing interfaces as being easier to use (Tractinsky, Katz, & Ikar, 2000).

With one exception, there was a noted lack of moderation effects, especially with respect to experimental contrasts (anthropomorphisms vs. colors vs. combinations). Perhaps the different manipulations may influence cognitive processes in the same way, but the evidence is only provisional. In general, the moderation analysis was limited by the small sample size, which prevented meta-regression, and lack of heterogeneity for some of the outcomes.

4.1. Future research directions

There are several prudent steps for future work. For one, there is not enough data on how emotional design augmentations influence primary/elementary school children. It is tempting to speculate that age could have a moderating role on the effectiveness of the manipulations. The moderation analyses yielded some preliminary evidence to this effect in that the manipulations were more effective in increasing intrinsic motivation for younger children, a finding that awaits replication when more studies are conducted.

The degree of anthropomorphizing has also rarely been examined. Results of Park et al. (2015) indicated that expressive eyes and mouth may attract more attention compared to simpler, geometric eyes and mouth. Three experiments by Schneider and colleagues (2018b) pointed to a possible interaction between degree of anthropomorphizing and age of the participants. These experiments showed that more complex anthropomorphisms not necessarily improve learning, especially if complex anthropomorphisms induce a higher cognitive load for a specific group of learners (i.e., younger ones, but not older ones).

Relatedly, not much is known about the influence of prior knowledge. Could there be an expertise-reversal effect (Kalyuga, 2007) where anthropomorphisms/colors are more effective for low vs. high prior knowledge learners due to processing differences between the two? This question remains to be answered.

There is also limited information on how these anthropomorphisms/colors work in longer treatments. Would they be boring or, contrarily, would they help maintain interest across extended timeframes? Similarly, only one study used delayed knowledge assessments

(Brom et al., unpublished manuscript), so further research should investigate whether the beneficial immediate learning effects observed here are robust after a delay.

Beyond their hypothesized influence on affective-motivational states, we highlighted the possibility that anthropomorphisms/colors might have signaling functions (Schneider, Beege, et al., 2018; van Gog, 2014). Or they might just be more salient and thus more memorable. As far as we know, only two studies have used process data, specifically eye tracking (Brom et al., unpublished manuscript; Park et al., 2015), to test for signaling, saliency, and other more basic cognitive effects. More research is needed in this regard.

The studies measured affective-motivational variables using self-report questionnaires which learners typically completed after the learning session. This is a limitation for longer study sessions as well as in terms of studying mediation by these variables. The use of process measures in future research would help expand our understanding of the interplay between affective-motivational states and cognitive processes during learning (see Uzun & Yıldırım, 2018 for a related example on the use of process measures to gauge positive emotions).

4.2. Limitations

Like all studies, ours had limitations. First, this meta-analysis utilized a relatively small number of samples, though it was comprehensive given our search and inclusion criteria and within the range of other influential meta-analyses (e.g., Gegenfurtner et al., 2014; Ginns et al., 2013; Ginns, 2005; Takacs, Swart, & Bus, 2014). Nevertheless, as the field emerges, it would be beneficial to update this meta-analysis when more studies are available, especially when it comes to assessing moderation effects.

Second, some included studies manipulated other “minor” elements (e.g., font type, roundness edges; see Sec. 2.1) in addition to anthropomorphisms and/or pleasant colors. Perhaps the most salient of these was the alteration of facial expressions of human characters (neutral versus positive) in the study by Uzun and Yıldırım (2018) and replacing some gaze or visible movement directions by arrows in the non-anthropomorphic versions (Mayer & Estrella, 2014; Brom et al., unpublished manuscript). The influence of these "minor" elements could not be ascertained in the present meta-analysis, but it would be useful to do so when more studies become available.

Third, we excluded a set of "older" (i.e., before 1990) instructional film and TV studies, somewhat forgotten in the context of modern emotional design. These studies examined effects of switching from monochrome to colorful instructional formats (e.g., Johnson & Robertson, 1979; Kanner & Rosenstein, 1960; Kanner, 1968; Dwyer & Lamberski, 1982), but did not manipulate color systematically or for a specific purpose, such as using color as a cognitive cue or to trigger affective-motivational states (see also Sec. 1.3). Not surprisingly, this "random" usage of color produced null/mixed effects (see Kanner, 1968; Dwyer & Lamberski, 1982). Nevertheless, it might have still been useful to re-examine this body of research. Unfortunately, this posed a challenge because several of these studies were difficult to locate, others did not report key statistics (e.g., missing SDs), and most did not include detailed descriptions of manipulations, let alone screenshots, to satisfy our inclusion criterion (see Johnson & Robertson, 1979; Kanner & Rosenstein, 1960 for examples).

Finally, studies generally intended anthropomorphisms to increase positive affective-motivational states. However, some of the anthropomorphisms used might not be positive (e.g., the faces in Figure 1 are frowning or fearful). We could not quantify the extent of positive versus negative emotions depicted in the anthropomorphic elements due to a lack of access of the primary instructional materials, though this could play a moderating role. Future studies should consider quantifying depicted emotions in detail.

4.3. Alternative emotional design principles

It is worth considering other potential emotional design principles in addition to anthropomorphisms and colors. In particular, various gamification approaches (Deterding, Dixon, Khaled, & Nacke, 2011) have been applied in educational settings (e.g., additions of individual game elements, such as points, badges or leaderboards), but with mixed evidence of effectiveness (see Ortiz-Rojas, Chiluiza, & Valcke, 2017[1]). For instance, badges and/or points with leaderboards can increase participation in various activities (e.g., Barata, Gama, Jorge, & Gonçalves, 2013; Denny, 2013; Halan, Rossen, Cendan, & Lok, 2010), but may not improve learning outcomes (e.g., de-Marcos, Garcia-Lopez, & Garcia-Cabot, 2016; Hanus & Fox, 2015). Therefore, gamifying educational experiences is not (thus far) a promising emotional design principle. The same conclusion holds for game elements tested within games (rather than outside of games as is done in gamification research) as studies examining motivational effects of game elements, such as narrative framing, customization, or interactivity, have been so far few and yielded inconclusive results (Wouters & Oostendorp, 2017; see also Clark et al., 2016; Mayer, 2014c, Ch. 5). One of the reasons for mixed findings is that these game elements may increase extraneous load too much (unlike "minimalistic" affective-motivational manipulations).

The inclusion of pedagogical agents appears to have more promise. The average effect size of agents on learning is small (see Schroeder, Adesope, & Gilbert, 2013, but see also Heidig & Clarebout, 2011), but it increases when the level of agent embodiment is increased, for example, via facial expressions or gesturing (Mayer, 2014d, p. 361; see also Plass & Kaplan, 2015). That said, it is problematic to consider the inclusion of pedagogical agents as

[1] This review mixes studies with and without a control group. We were informed (email dating from 10 October 2017) that among the studies with a control group, three had positive results, three had null/mixed results, and two had negative results.

one monolithic design principle. In particular, when the agent does not present additional instructionally relevant information (compared to the no-agent condition), it basically acts as a seductive detail (e.g., Moreno, Mayer, Spires, & Lester, 2001). When the agent does provide such information, the information may enhance learning in and of itself. For example, Lusk and Atkinson (2007) found that the agent enhanced learning, but it gestured and used gaze signals to direct learner attention to relevant parts of the screen, which likely aided in selecting relevant information. Thus, it is still unclear which agent features and behaviors enhance learning. Whereas some target attention-cognitive processes (e.g., deictic gestures), others target affective-motivational ones (e.g., expressions of enthusiasm: Liew, Zin, Sahari, 2017). This research avenue may eventually produce multiple agent-based design principles, but some may be encapsulated in already existing principles, such as the signaling principle (van Gog, 2014), or not-yet-established principles, such as "enthusiasm" principle. As concerns the latter, teacher enthusiasm, be it in class or instructional videos, has positive effects on various student outcomes (see Keller, Hoy, Goetz, & Frenzel, 2016), but "the effect on students' achievement … [still] needs further clarification" (p. 763). The review by Keller and colleagues mentioned five studies with positive effects, five with no significant difference, and one with a negative finding (p. 761).

Further, research on other pleasant graphical manipulations has been emerging, such as changes to font and roundness/sharpness of object edges (see Section 1.2). Equally important, albeit nascent, is research into effects of other-than-pleasant color combinations purposefully targeting affective-motivational states (e.g., negative color: Kumar et al., 2016; see also Section 1.3), and research on how to productively utilize negative affective states (e.g., confusion, D'Mello, Lehman, Pekrun, Graesser, 2014). Researchers have also started to reexamine whether seductive details can enhance learning under specific conditions (e.g., Park, Flowerday, Brünken, 2015; Schneider, Nebel, & Rey, 2016).

Altogether, more research is needed to examine effects of these alternate emotional design principles.

4.4. Conclusion

Using meta-analytic techniques, we found that augmenting multimedia learning materials with anthropomorphisms and/or pleasant colors improved learning and intrinsic motivation. To a small extent, these augmentations also enhanced liking/enjoyment and generalized positive affect and reduced perceptions of difficulty. Thus, these augmentations appear to be a useful design principle, though it is still unclear if they reflect an *emotional* design principle due to a lack of causal analyses. Suggested future research directions thus include examining possible moderators as more studies accrue (different age groups, long exposures, different levels of prior knowledge) and investigating causal mechanisms using more sensitive measures (e.g., attention-capturing effects using eye trackers; physiological arousal by tracking electrodermal activity). As it currently stands, instructional designers might consider incorporating anthropomorphisms and/or pleasant colors in their instructional materials.

References

Note. Studies with * are included in the meta-analysis.

Abrami, P. C., Leventhal, L., & Perry, R. P. (1982). Educational seduction. *Review of Educational Research, 52*(3), 446-464.

Adams, D. M., Mayer, R. E., MacNamara, A., Koenig, A., & Wainess, R. (2012). Narrative games for learning: Testing the discovery and narrative hypotheses. *Journal of Educational Psychology, 104*(1), 235-249.

Andrey, M.-N., Brunisholz, V., Dos Reis, B., & Molinari, G. (2016). How to improve learners' emotional experiences in multimedia learning environments: Effects of emotional design and induced achievement goals. *Paper presented at EARLI SIG 2 meeting:* Geneva, Switzerland.

Barata, G., Gama, S., Jorge, J., & Gonçalves, D. (2013). Improving participation and learning with gamification. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications* (pp. 10-17): ACM.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*, 1088–1101.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological, 57*(1), 289-300.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Hoboken, NJ: John Wiley & Sons Inc.

Brom, C., Děchtěrenko, F., Frollová, N., Stárková, T., Bromová, E., & D'Mello, S. K. (2017). Enjoyment or involvement? Affective-motivational mediation during learning from a complex computerized simulation. *Computers & Education, 114*, 236-254.

* Brom, C., Hannemann, T., Stárková, T., Bromova, E., & Dechterenko, F. (2016). Anthropomorfic faces and funny graphics in an instructional animation may improve superficial rather than deep learning: A quasi-experimental study. *Proceedings of European Conference on e-Learning* (pp. 89 - 97): Academic Conferences International Limited.

Brom, C., Hannemann, T., Stárková, T., Bromová, E., & Děchtěrenko, F. (2017). The role of cultural background in the personalization principle: Five experiments with Czech learners. *Computers & Education, 112*, 37-68.

* Brom, C., Stárková, T., Lukavský, J., & Javora, O. (unpublished manuscript). *Anthropomorphisms in multimedia learning: Are they universally effective?*

Brünken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load. In *Cognitive Load Theory* (pp. 181-202): Cambridge University Press.

Crouzet, S., & Thorpe, S. (2011). Low-level cues and ultra-fast face detection. *Frontiers in Psychology, 2*(342). doi:10.3389/fpsyg.2011.00342

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*(2), 145-157.

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction, 29*, 153-170.

de-Marcos, L., Garcia-Lopez, E., & Garcia-Cabot, A. (2016). On the effectiveness of game-like and social approaches in learning: Comparing educational gaming, gamification & social networking. *Computers & Education, 95*, 99-113.

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science, 38*(2), 105-134.

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology, 100*(1), 223-234.

Denny, P. (2013). The effect of virtual achievements on student engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 763-772): ACM.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9-15): ACM.

Doolittle, P. E., & Altstaedter, L. L. (2009). The effect of working memory capacity on multimedia learning: Does attentional control result in improved performance? *Journal of Research in Innovative Teaching, 2*(1), 7-25.

Domagk, S. (2010). Do pedagogical agents facilitate learner motivation and learning outcomes? The role of the appeal of agent's appearance and voice. *Journal of Media Psychology, 22*, 84-97.

Duval, S. J. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.) *Publication Bias in Meta-analysis: Prevention, Assessment, and Adjustments* (pp. 127–144). Chichester, England: Wiley.

Dwyer, F. M. (1971). Color as an instructional variable. *AV Communication Review, 19*(4), 399-416.

Dwyer, F. M., & Lamberski, R. J. (1982). A review of the research on the effects of the use of color in the teaching-learning process. *International Journal of Instructional Media, 10*(4), 303-327.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629–634.

Ekman, P., & Rosenberg, E. L. (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*: Oxford University Press, USA.

Fiedler, K., & Beier, S. (2014). Affect and cognitive processes in educational contexts. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International Handbook of Emotions in Education* (pp. 36-56). New York, NY: Routledge.

Fulmer, S., D'Mello, S., Strain, A., & Graesser, A. (2015). Interest-based text preference moderates the effect of text difficulty on engagement and learning. *Contemporary Educational Psychology, 41*, 98-110.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological Bulletin, 133*(4), 694-724.

Garner, R., Brown, R., Sanders, S., & Menke, D. J. (1992). "Seductive details" and learning from text. In *The Role of Interest in Learning and Development* (pp. 239-254): Lawrence Erlbaum Associates.

Gegenfurtner, A., Quesada-Pallarès, C., & Knogler, M. (2014). Digital simulation-based training: A meta-analysis. *British Journal of Educational Technology, 45*(6), 1097-1114.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*(4), 313-331.

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.

Guo, Y. R., & Goh, D. H. L. (2016). Evaluation of affective embodied agents in an information literacy game. *Computers & Education, 103*, 59-75.

* Gong, S., Shangguan, C., Zhai, K., & Guo, Y. (2017). The effects of emotional design on multimedia learning. *Acta Psychologica Sinica, 49*(6), 771-782.

* Haaranen, L., Ihantola, P., Sorva, J., & Vihavainen, A. (2015). In search of the emotional design effect in programming. In *IEEE/ACM 37th IEEE International Conference on Software Engineering* (Vol. 2, pp. 428-434): IEEE.

Halan, S., Rossen, B., Cendan, J., & Lok, B. (2010). High score!-motivation strategies for user participation in virtual human development. In *International Conference on Intelligent Virtual Agents* (pp. 482-488): Springer.

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education, 80*, 152-161.

Hawlitschek, A., & Joeckel, S. (2017). Increasing the effectiveness of digital educational games: the effects of a learning instruction on students' learning, motivation and cognitive load. *Computers in Human Behavior, 72*, 79-86.

Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review, 6*(1), 27-54.

* Heidig, S., Müller, J., & Reichelt, M. (2015). Emotional design in multimedia learning: Differentiation on relevant design features and their effects on emotions and learning. *Computers in Human Behavior, 44*, 81-95.

Isen, A. M., & Reeve, J. (2005). The influence of positive affect on intrinsic and extrinsic motivation: Facilitating enjoyment of play, responsible work behavior, and self-control. *Motivation and Emotion, 29*(4), 295-323.

Johnson, B., & Roberson, J. (1979). Color in instructional films: Does it still make no difference in learning achievement? *Educational Technology, 19*(1), 32-35.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*(4), 509-539.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review, 23*(1), 1-19.

Kanner, J. H. (1968). The instructional effectiveness of color in television: A review of the evidence. Stanford University, ERIC ID ED015675.

Kanner, J. H., & Rosenstein, A. J. (1960). Television in army training: Color vs black and white. *Audiovisual Communication Review, 8*(6), 243-252.

Kaplan, R. M., & Pascoe, G. C. (1977). Humorous lectures and humorous examples: Some effects upon comprehension and retention. *Journal of Educational Psychology, 69*(1), 61-65.

Ke, F. (2008). Computer games application within alternative classroom goal structures: cognitive, metacognitive, and affective evaluation. *Educational Technology Research and Development, 56*(5-6), 539-556.

Keller, M. M., Hoy, A. W., Goetz, T., & Frenzel, A. C. (2016). Teacher enthusiasm: reviewing and redefining a complex construct. *Educational Psychology Review, 28*(4), 743-769.

Knörzer, L., Brünken, R., & Park, B. (2016). Facilitators or suppressors: Effects of experimentally induced emotions on multimedia learning. *Learning and Instruction, 44*, 97-107.

* Kumar, J. A. (2016). *Effects Of Multimedia Induced Emotions on Achievement, Intrinsic Motivation And Satisfaction Among Polytechnic Students With Different Levels of Emotional Intelligence.* (PhD thesis) Universiti Sains Malaysia, Malaysia

* Kumar, J. A., Muniandy, B., & Yahaya, W. A. J. W. (2016). Emotional design in multimedia learning: How emotional intelligence moderates learning outcomes. *International Journal of Modern Education and Computer Science, 8*(5), 54-63.

Leppink, J., Paas, F., Van Gog, T., van Der Vleuten, C. P., & Van Merrienboer, J. J. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30*, 32-42.

Leutner, D. (2014). Motivation and emotion as mediators in multimedia learning. *Learning and Instruction, 29*, 174-175.

Liew, T. W., Zin, N. A. M., & Sahari, N. (2017). Exploring the affective, motivational and cognitive effects of pedagogical agent enthusiasm in a multimedia learning environment. *Human-centric Computing and Information Sciences, 7*(1), doi: 10.1186/s13673-017-0089-2

Lorenz, K. (1970). *Studies in Animal and Human Behaviour*: Harvard University Press.

Lusk, M. M., & Atkinson, R. K. (2007). Animated pedagogical agents: Does their degree of embodiment impact learning from static or animated worked examples? *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 21*(6), 747-764.

Mayer, R. E. (2009). *Multimedia Learning* (2nd ed.): Cambridge University Press.

Mayer, R. E. (Ed.) (2014a). *The Cambridge handbook of multimedia learning* (2nd ed.): Cambridge University Press.

Mayer, R. E. (2014b). Incorporating motivation into multimedia learning. *Learning and Instruction, 29*, 171-173.

Mayer, R. E. (2014c). *Computer Games for Learning: An Evidence-Based Approach*: The MIT Press.

Mayer, R. E. (2014d). Principles based on social cues in multimedia learning: personalization, voice, image, and embodiment principles. In *The Cambridge handbook of multimedia learning* (pp. 345-368): Cambridge University Press.

* Mayer, R. E., & Estrella, G. (2014). Benefits of emotional design in multimedia instruction. *Learning and Instruction, 33*, 12-18.

* Miller, C. (2011). Aesthetics and e-assessment: the interplay of emotional design and learner performance. *Distance Education, 32*(3), 307-337.

Moreno, R. (2005). Instructional technology: Promise and pitfalls. *Technology-based Education: Bringing Researchers and Practitioners Together* (pp. 1-19): Information Age Publishing.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction, 19*(2), 177-213.

* Münchow, H. (2017). *I feel, therefore I learn–Effectiveness of affect induction interventions and possible covariates on learning outcomes.* (PhD thesis), Julius-Maximilians-Universität Würzburg, Würzburg.

* Münchow, H., Mengelkamp, C., & Bannert, M. (2017). The better you feel the better you learn: Do warm colours and rounded shapes enhance learning outcome in multimedia learning? *Education Research International, 2017*, Art. No. 2148139.

* Ng, K. H., & Chiu, T. K. (2017). Emotional multimedia design for developing mathematical problem-solving skills. In *New Ecology for Education—Communication X Learning* (pp. 131-141): Springer.

Norman, A. D. (2004). *Emotional Design: Why We Love (or Hate) Everyday Things*: Basic Books.

* Nurminen, M., Leppänen, L., Väätäjä, H., & Ihantola, P. (2017). Search of the emotional design effect in programming revised. In *European Conference on Technology Enhanced Learning* (pp. 434-440): Springer.

Ozcelik, E., Karakus, T., Kursun, E., & Cagiltay, K. (2009). An eye-tracking study of how color coding affects multimedia learning. *Computers & Education, 53*(2), 445-453.

Ortiz-Rojas, M., Chiluiza, K., & Valcke, M. (2017). Gamification and learning performance: a systematic review of the literature. In *European Conference on Games Based Learning* (pp. 515-522): Academic Conferences International Limited.

Park, B., Flowerday, T., & Brünken, R. (2015). Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behavior, 44*, 267-278.

* Park, B., Knörzer, L., Plass, J. L., & Brünken, R. (2015). Emotional design and positive emotions in multimedia learning: An eyetracking study on the use of anthropomorphisms. *Computers & Education, 86*, 30-42.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315-341.

Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In *Handbook of Research on Student Engagement* (pp. 259-282): Springer Science+Business Media.

Pett, D., Wilson, T. (1996). Color research and its application to the design of instructional materials. *Educational Technology Research and Development, 44*(3), 19-35.

* Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D., & Um, E. (2014). Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction, 29*, 128-140.

Plass, J. L., & Kaplan, U. (2015). Emotional design in digital media for learning. In *Emotions, Technology, Design, and Learning* (pp. 131-162): Academic Press.

Renkl, A., & Scheiter, K. (2017). Studying visual displays: How to instructionally support learning. *Educational Psychology Review, 29*(3), 599-621.

Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review, 7*(3), 216-237.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54-67.

Schallberger, U. (2005). Kurzskalen zur Erfassung der Positiven Aktivierung, Negativen Aktivierung und Valenz in Experience Sampling Studien (PANAVA-KS).

Theoretische und methodische Grundlagen, Konstruktvalidität und psychometrische Eigenschaften bei der Beschreibung intra-und interindividueller Unterschiede.

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.

* Schneider, S., Nebel, S., Beege, M., & Rey, G. D. (2018a). Anthropomorphism in decorative pictures: benefit or harm for learning? *Journal of Educational Psychology 110*(2), 218-232.

* Schneider, S., Häßler, A., Habermeyer, T., Beege, M., & Rey, G. D. (2018b, March 29). The more human, the higher the performance? Examining the effects of anthropomorphism on learning with media. *Journal of Educational Psychology*. Advance online publication. http://dx.doi.org/10.1037/edu0000273

Schneider, S., Nebel, S., & Rey, G. D. (2016). Decorative pictures and emotional design in multimedia learning. *Learning and Instruction, 44*, 65-73.

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, 49*(1), 1-39.

Stark, L., Brünken, R., & Park, B. (2018). Emotional text design in multimedia learning: A mixed-methods study using eye tracking. *Computers & Education, 120*, 185-196.

Stone, N. J., & English, A. J. (1998). Task type, posters, and workspace color on mood, satisfaction, and performance. *Journal of Environmental Psychology, 18*(2), 175-185.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. New York: Springer.

Takacs, Z. K., Swart, E. K., & Bus, A. G. (2014). Can the computer replace the adult for storybook reading? A meta-analysis on the effects of multimedia stories as compared to sharing print stories with an adult. *Frontiers in Psychology, 5*(1366), doi:10.3389/fpsyg.2014.01366

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers, 13*(2), 127-145.

*Um, E. R., Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Educational Psychology, 104*(2), 485-498.

* Um, E. R., Song, H., & Plass, J. (2007). The effect of positive emotions on multimedia learning. *EdMedia: World Conference on Educational Media and Technology* (pp. 4176-4185): Association for the Advancement of Computing in Education (AACE).

* Uzun, A. M., & Yıldırım, Z. (2018). Exploring the effect of using different levels of emotional design features in multimedia science learning. *Computers & Education, 119*, 112-128.

Valdez, P., & Mehrabian, A. (1994). Effects of color on emotions. *Journal of Experimental Psychology: General, 123*(4), 394-409.

van Buskirk, W. L. (1932). An experimental study of vividness in learning and retention. *Journal of Experimental Psychology, 15*(5), 563-573.

van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. *The Cambridge Handbook of Multimedia Learning* (pp. 263-278): Cambridge University Press.

van Wermeskerken, M., & van Gog, T. (2017). Seeing the instructor's face and gaze in demonstration video examples affects attention allocation but not learning. *Computers & Education, 113*, 98-107.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*(2), 219-235.

Weller, L., & Livingston, R. (1988). Effect of color of questionnaire on emotional responses. *The Journal of General Psychology, 115*(4), 433-440.

Wolfson, S., & Case, G. (2000). The effects of sound and colour on responses to a computer game. *Interacting with Computers, 13*(2), 183-192.

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105*(2), 249-265. doi:10.1037/a0031311

Wouters, P., & van Oostendorp, H. (2017). Overview of instructional techniques to facilitate learning and motivation of serious games. In *Instructional Techniques to Facilitate Learning and Motivation of Serious Games* (pp. 1-16): Springer.

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research, 81*(2), 267-301.