



OPEN ACCESS

EDITED BY

Wang-Kin Chiu,
The Hong Kong Polytechnic University,
China

REVIEWED BY

Sheila L. Macrine,
University of Massachusetts Dartmouth,
United States
Rafidah Abd Karim,
MARA University of Technology,
Malaysia

*CORRESPONDENCE

Ard W. Lazonder
✉ ard.lazonder@ru.nl

†PRESENT ADDRESS

Sifra E. Muilwijk,
Institute of Movement Studies,
University of Applied Sciences Utrecht,
Utrecht,
Netherlands

SPECIALTY SECTION

This article was submitted to
STEM Education,
a section of the journal
Frontiers in Education

RECEIVED 10 February 2023

ACCEPTED 03 March 2023

PUBLISHED 31 March 2023

CITATION

Muilwijk SE and Lazonder AW (2023) Learning
from physical and virtual investigation: A meta-
analysis of conceptual knowledge acquisition.
Front. Educ. 8:1163024.
doi: 10.3389/educ.2023.1163024

COPYRIGHT

© 2023 Muilwijk and Lazonder. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted which
does not comply with these terms.

Learning from physical and virtual investigation: A meta-analysis of conceptual knowledge acquisition

Sifra E. Muilwijk^{1†} and Ard W. Lazonder^{2*}

¹School of Pedagogical and Educational Sciences, Radboud University, Nijmegen, Netherlands,

²Behavioural Science Institute, Radboud University, Nijmegen, Netherlands

Should students investigate with tangible objects and apparatus or are digitally simulated materials and equipment an adequate or perhaps even preferred alternative? This question remains unanswered because empirical evidence is inconclusive and previous reviews are descriptive and synthesize a limited number of studies with small samples. This meta-analysis, therefore, assessed the relative effectiveness of physical versus virtual investigation in terms of conceptual knowledge acquisition and examined whether and how the aggregate effect size was moderated by substantive and methodological study features. Following a systematic search of Web of Science and ERIC for the period 2000–2021, 35 studies comparing physical and virtual investigations were selected for inclusion. Hedges' g effect sizes for conceptual knowledge acquisition were computed and analyzed using a random effects model. The results showed no overall advantage of either mode of investigation ($g = -0.14$, 95% CI $[-0.33, 0.06]$). However, moderator analysis indicated that virtual investigation is more effective for adults compared with adolescents and children, and when touching objects or equipment does not provide relevant sensory information about the concept under study. These results imply that STEM teachers can decide for themselves whether to opt for physical or virtual investigation except when teaching adult students or when touch sensory feedback is substantively irrelevant; in those cases, virtual investigation is preferable.

KEYWORDS

virtual labs, simulation – computers, physicality, manipulatives and experimentation, touch

1. Introduction

Technological advancements have significantly extended the opportunities to include investigations in courses for students of all ages. The past decade has witnessed several successful initiatives to grant teachers of science, technology, engineering, and mathematics (STEM) free access to computer simulations and online laboratories for teaching and learning in K12 classrooms and beyond. Although designers of these technologies are optimistic about the value of virtual investigation for learning (Perkins et al., 2012; De Jong et al., 2014), empirical evidence is typically mixed. Some studies confirm that virtual investigation is more effective than physical investigation (e.g., Chao et al., 2016), whereas other studies found the opposite effect (e.g., Zacharia et al., 2012) or report no differences (e.g., Renken and Nunez, 2013). As this body of research has, to the best of our knowledge, not been quantitatively reviewed, the true virtue of virtual investigation is yet unknown. This observation sparked the idea for this meta-analysis, which aimed to examine the relative effectiveness of physical versus virtual investigations.

In line with [Klahr et al. \(2007\)](#), we use the term physical investigation to refer to hands-on inquiries where students interact with tangible objects and equipment to acquire a conceptual understanding of the topic being studied. Although such investigations admittedly enable students to strengthen their research skills, learn to collaborate with peers, and build interest in STEM-related careers, these learning outcomes were outside the scope of this meta-analysis. Virtual investigation, then, is the digital analog of physical investigation in that students' examinations involve simulated material and apparatus provided by a computer simulation, virtual laboratory, or virtual–reality application. Both definitions are fleshed out more in the sections below. Following a short overview of instructional approaches that incorporate student investigations, we zoom in on the unique affordances of physical and virtual investigation and summarize the results of previous narrative reviews that contrasted these modes of investigation.

2. Theoretical foundation

2.1. Learning through investigation

Preschoolers learn from play and by exploring the world around them. Schools respond to this investigative drive by engaging children in inquiry projects, for instance, to examine how long it takes for colored ink to dissolve in hot and cold water. High school students spend quite some time in the school science laboratory, and throughout higher education, student research progressively approximates authentic scientific practices. All these instances are rooted in the long-standing belief that the act of investigating is productive to learning because finding things out by oneself leads to more meaningful and sustainable knowledge than being told by a teacher ([Dewey, 1900](#); [Schneider et al., 2022](#); [De Jong et al., in press](#)).

Student investigations are integral to instructional approaches such as experiential learning, problem-based learning, and inquiry-based learning. Although generally embraced by policymakers and field experts, some educational scientists have challenged the effectiveness of these approaches based on a lack of teacher guidance (e.g., [Kirschner et al., 2006](#); [Zhang et al., 2022](#)). A comprehensive meta-analysis confirmed that learning through unguided investigation is less effective than explicit instruction. However, students who were guided during their investigation learned more than students who studied the same material through expository methods ([Alfieri et al., 2011](#)). In other words, instructional approaches that include student investigations are effective if adequate guidance is provided.

Which type of guidance is appropriate for which types of learners is still debated. Contrary to the intuitive belief that young learners need more specific guidance than older learners do, student age does not moderate the influence of guidance on learning activities and learning outcomes ([Lazonder and Harmsen, 2016](#)). A plausible explanation might be that the complexity of students' investigations increases with age: older students not only examine more difficult topics but are also exposed to more open forms of inquiry that necessitate specific forms of guidance ([Bell et al., 2005](#)). As such, creating effective learning arrangements that include student research is a balancing act that becomes even more challenging if teachers can choose between physical and virtual investigations.

2.2. The case for the physical investigation

Imagine giving a child a set of cubes and spheres to investigate what determines how fast objects sink in water. While experimenting, the child receives sensory feedback through the eyes (e.g., spheres sink in a straight line whereas cubes whirl down) and ears (sound indicates when an object hits the bottom of the water cylinder). A unique additional affordance of physical investigations is that handling the spheres and cubes produces touch sensory feedback about their mass and surface not normally available in virtual investigations. [It can be mimicked by a haptic device, but these tools are still rare in educational settings ([Luo et al., 2021](#)) and were, therefore, not included in this meta-analysis].

The educational importance of touch sensory feedback is articulated in theories of embodied cognition, which assert that a person interacting with the material world creates 'embodied' knowledge of physical objects and phenomena (e.g., [Gallese and Lakoff, 2005](#); [Barsalou, 2008](#)). Neuroimaging studies have provided evidence that conceptual understanding is stored in the sensory–motor circuits of the brain, meaning that the brain regions involved in seeing, hearing, and touching objects are also activated during recall ([Kiefer and Trumpp, 2012](#)). If students are deprived of touch, conceptual knowledge becomes less rich as it is exclusively based on verbal and auditory stimuli ([Zacharia, 2015](#)). The child in our sinking objects example learns about the mass of the objects by picking them up and the experience of feeling the difference between, for instance, a 10 and 100 g cube complements mass-related information from the other senses.

While embodied cognition theories emphasize the storage and retrieval of information, the additional sensory channel theory addresses the encoding process. Rooted in theoretical conceptions of working memory ([Baddeley, 2012](#)) and cognitive load ([Sweller et al., 2019](#)), this theory postulates that the brain has separate processing channels for visual, auditory, and tactile information. If multiple channels are used for learning a particular piece of information, the effective working memory capacity expands and, hence, the chance of better learning outcomes increases. Note that this theory merely applies to sensory information relevant to the concept students are investigating. Suppose the child in our example senses that metal objects feel colder than Teflon objects, then this tactile feedback will not lower her cognitive load when examining how the shape of an object influences its sink time.

2.3. The case for the virtual investigation

Proponents of virtual investigation point to the practical advantages of simulations and virtual laboratories. These digital environments require little preparation from the teacher and enable students to design and conduct many investigations in a short amount of time ([De Jong et al., 2013](#)). Virtual investigations also offer a viable alternative for physical investigation if material or apparatus is expensive or when the research site is geographically remote ([Hannel and Cuevas, 2018](#))—think, for example, of a field trip to the Falkland telescope to have astronomy students observe distant galaxies. Similar advantages apply when the topic of investigation is rare (e.g., lunar eclipses) or dangerous (e.g., radioactivity).

The learning benefits of virtual investigation are essentially twofold. Contentwise, designers of digital investigation environments

can impose productive constraints on students—for instance, by simplifying a phenomenon or restricting the values that can be set in an experiment—or provide visualizations that allow students to perceive what is not directly observable in the material world (De Jong et al., 2013; Sullivan et al., 2017). These options aim to reduce intrinsic cognitive load. Extraneous cognitive load can be decreased by embedding instructional support features in the virtual environment, which is the second learning advantage. Software designers can, for instance, use virtual–reality technology to direct students' attention to important parts of the screen at key moments during an inquiry or augment digital objects and processes with additional explanations (De Jong et al., 2013).

2.4. Previous narrative reviews

An early research overview by Ma and Nickerson (2006) found no significant and consistent difference between physical and virtual investigation. De Jong et al. (2013) reached a similar conclusion and speculated about possible differential effects by suggesting that young learners might benefit more from physical investigation because they tend to lack tactile experience with the objects or processes under study. Virtual investigation, according to De Jong et al., might be more advantageous in situations that align with the learning advantages described in the previous paragraph. Zacharia (2015) also concluded that touch sensory feedback from physical investigation is not a requirement for the acquisition of conceptual knowledge.

Related literature overviews challenge the latter conclusion by showing that haptic augmentation of virtual environments often improves the development of conceptual knowledge (Minogue and Jones, 2006; Zacharia, 2015) and procedural skills (Rangarajan et al., 2020). This positive trend seems due to the fact that all haptic devices provided 'force feedback' directly relevant to the topic being studied (e.g., gears and lever principles). Future research could test this presumption by comparing physical and virtual investigation, the latter without haptics, in situations where touch sensory information helps students build an understanding of the concepts or processes they are investigating. Furthermore, a meta-analysis in the field of mathematics education strengthened the tentative conclusion regarding the moderating influence of learners' age (Carbonneau et al., 2013). Using concrete manipulatives in math classes was found to be more effective for children in the concrete operational stage compared with children in the formal operational stage, allegedly because younger children rely more on physical interaction with the material world when constructing meaning than older children who are capable of formal operational reasoning.

In summary, previous research integrations converge on the equivalent effectiveness of physical and virtual investigation but differ regarding the existence of possible age-related differences as well as the educational affordances of touch. However, as most of these works descriptively synthesized a selective number of studies with small sample sizes, more rigorous and quantitative research integrations are needed to draw any definitive conclusion.

3. Research questions

This meta-analysis aimed to answer three research questions:

1. What is the relative effectiveness of physical versus virtual investigation in terms of conceptual knowledge acquisition?
2. How does this relative effectiveness depend on the substantive contribution of touch sensory information to the concept under study?
3. How does this relative effectiveness depend on the students' age?

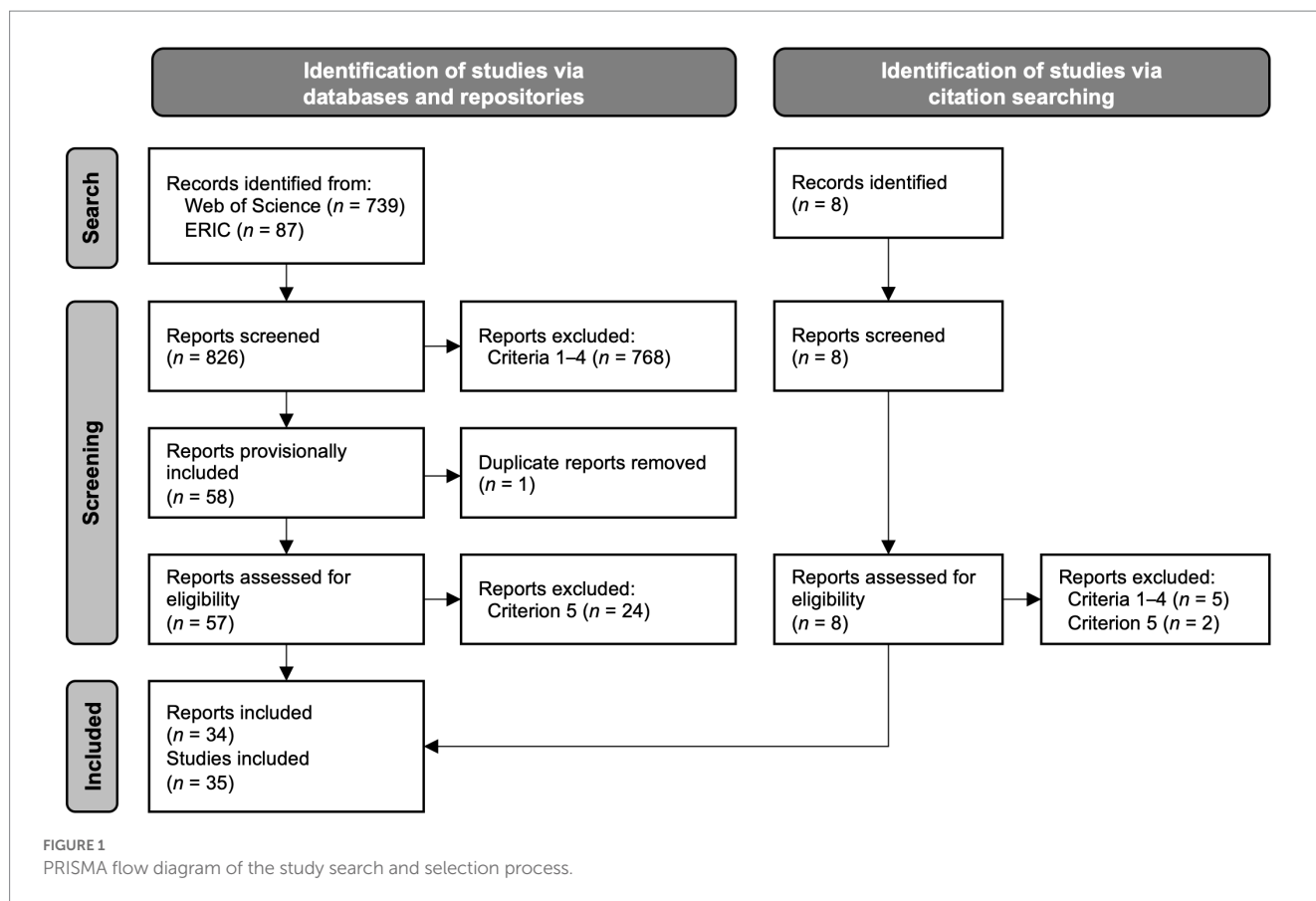
4. Method

4.1. Search and selection of studies

The literature was searched for studies that satisfied the following inclusions criteria:

1. The study examined students investigating STEM-related topics for learning purposes.
2. The study compared the conceptual knowledge acquisition of students who did their research with physical materials to that of students who performed the same investigation with virtual materials.
3. The study was set up to ensure that similar instructional regimes were implemented in both conditions.
4. The study controlled for possible differences in prior domain knowledge either by randomization or analyzing pre- and post-assessment scores.
5. The study administered a between-subject design and reported data from which effect sizes can be calculated.
6. The study was published between 2000 and 2021 and is available online in full text.

The search and selection processes are visualized in Figure 1. All searches were performed in the Web of Science Core Collection and the ERIC repository using the following query: [(physical experiment* OR physical lab* OR hands-on) AND (virtual experiment* OR virtual lab* OR hands-off OR simulation*)]. The Web of Science search, restricted to the SSCI and SCI Expanded citation indexes and further limited to the categories 'Education Educational Research,' 'Education Scientific Disciplines,' 'Psychology Experimental,' and 'Psychology Multidisciplinary,' uncovered 739 reports. The ERIC database was searched similarly except that all search terms were wrapped in quotes to perform a literal search, and the results were limited to publications available in full text. This search returned 87 hits. Next, all 826 reports were retrieved and subjected to a title and abstract screening. A total of 58 reports passed this initial test, and after one duplicate was removed, 57 reports were read in full to assess their eligibility for inclusion. In addition, the perusal of the citations in previous reviews (De Jong et al., 2013; Zacharia, 2015) yielded eight reports that were not identified through the online search. These reports were retrieved and screened similarly, which led to the inclusion of one additional report. This brought the total number of included reports to 34. As one of these investigations presented data from two experiments with separate samples, the total number of studies in this meta-analysis was 35. Their main characteristics are summarized in Supplementary Appendix 1.



4.2. Coding of moderator and outcome variables

Students' *knowledge acquisition* served as the main outcome measure. It was defined as the conceptual knowledge participants developed through either physical or virtual investigation, as indicated by assessments administered during or shortly after the learning process. The first moderator, *assessment type*, classified the measurement used as a multiple-choice test, constructed-response test, performance-based assessment, or a combined format.

The next two moderators served to answer this study's research questions. The first one, *tactile feedback*, indicated whether physical manipulation provided touch sensory information that helps students build an understanding of the concepts they are investigating. The moderator *student age* gave a broad indication of the sample's mean age. In keeping with Piaget's and Erikson's stages of cognitive development (Thomas, 2005), a distinction was made between school-age children (6–11 years), adolescents (12–18 years), and young adults (19–27 years). In case participants' age was not provided, the age category was inferred from the students' grade levels, considering the differences in educational systems across countries.

The remaining moderators provided some descriptive details of the included studies. *Publication year* was used as an approximation of the time when the study was conducted. As computer technology becomes increasingly more sophisticated, older studies using computer simulations might yield different results than recent research with highly advanced virtual investigation facilities. To investigate whether such a differential effect exists, studies were

classified according to the decade of publication (2000–2010 or 2011–2021). *The research setting* concerned the site where the study took place. Two broad categories were distinguished: research laboratory and regular classroom. The former indicated that data were collected in a researcher-controlled environment, for example, a genuine university research laboratory or a separate room in the school building. Studies that were carried out in an authentic learning environment (e.g., a lecture room, the school's science laboratory, or a computer laboratory) were placed in the category 'regular classroom'.

Interrater agreement was determined in case moderator coding required subjective interpretation by the raters. Agreement on 'assessment type' (78%, Fleiss' $\kappa = 0.71$) and 'tactile feedback' (86%, Fleiss' $\kappa = 0.72$) was substantial according to the benchmarks proposed by Landis and Koch (1977). All disagreements were resolved through discussion. The remaining moderators were coded by the first author, who conferred with the second author when in doubt. Fisher's exact tests were run to determine whether the five moderators were related. Using a Bonferroni-corrected alpha level of 0.005, none of the comparisons turned out to be statistically significant, the p -values were > 0.133 , which means that all moderators were mutually independent.

4.3. Computation of effect sizes

Standardized mean differences were computed and corrected for upward small-sample bias. This effect size metric, known as Hedges' g , was calculated as follows:

$$g = \frac{M_1 - M_2}{SD_{pooled}} \times \left(\frac{N-3}{N-2.25} \right) \times \sqrt{\frac{N-2}{N}}$$

where N is the total sample size, M_1 is the mean knowledge gain score of the students in the physical investigation condition, M_2 is the mean gain of the students in the virtual investigation condition, and SD is the weighted standard deviation of both groups combined. If gain scores were not reported, the study's effect size was calculated from pre- and post-assessment scores, test statistics (F , t , and χ^2), or frequency distributions, using the conversion formulas by Lipsey and Wilson (2001). Note that, as per computation, a positive effect size indicates that students learned more from physical investigation, whereas a negative effect size denotes higher learning from virtual investigation.

Studies reporting data for multiple subgroups or multiple outcome measures were handled according to the guidelines proposed by Borenstein et al. (2009). Specifically, one study presented separate scores for the high and low achievers in both the physical and virtual investigation conditions. To reduce bias, scores of the two cohorts within each condition were combined to yield a summary effect. Other studies assessed students' knowledge acquisition by multiple post-tests. As these tests were equally relevant in determining which concepts students had learned through investigation, their scores were combined to compute the study's effect size.

4.4. Data analysis

Main analyses were conducted with Meta Essentials (Suurmond et al., 2017). The random effects model was used because studies examining different-aged students engaged in different inquiry tasks with different objects and equipment are unlikely to share the same true effect size. Following a descriptive analysis of the studies' effect sizes, the summary effect was tested for significance by a z -test. Egger's regression test (Egger et al., 1997) and Orwin's (1983) Fail-safe N were used to determine whether and to what extent the observed overall effect was subject to publication bias. Next, Q -tests based on analysis of variance were used to determine whether the between-study variation in effect sizes was attributable to the moderator variables. If so and where appropriate, planned comparisons (Hedges and Pigott, 2004) were made to unveil which moderator categories differed significantly from one another.

5. Results

Data for this meta-analysis were extracted from 35 studies with 3,303 participants. The effect sizes of two studies were significantly greater than zero ($g=1.23$ and 1.63), which denotes a benefit of physical investigation over virtual investigation. Seven studies had a significant negative effect size in the range of -1.51 to -0.45 , which indicates in favor of virtual investigation, and in 26 studies, the physical-virtual comparison was a tie ($-0.37 < g < 0.52$).

The studies' overall mean effect size (g) was -0.14 , $SE=0.09$, 95% CI $[-0.33, 0.06]$. The I^2 statistic indicated that 82.44% of the effect size variance reflects true score variation; the variance of the true effect size (τ^2) was 0.21. As can be inferred from the confidence interval, the

investigation mode had no significant overall effect on students' knowledge acquisition, $z=-1.43$, $p=0.152$, meaning that experimenting with physical and virtual materials is equally beneficial to concept learning. Egger's regression test showed no sign of publication bias as the estimated intercept (-1.81) did not differ significantly from zero, $t(34)=0.73$, $p=0.472$. Orwin's Fail-safe N indicated that 476 studies with a nil effect would be needed to turn the Hedges' g to zero.

The results of the moderator analyses showed that the variation in effect sizes was independent of how students' conceptual knowledge was assessed and in which year a study was published (see Table 1). However, a significant moderating effect was found for tactile feedback. This result indicates that physical and virtual investigations yield comparable knowledge gains if touching materials provide relevant information about the concepts to be learned, but that virtual investigation is more effective when the touch experience is extraneous.

The participants' age also moderated the findings. The mean effect size of studies conducted with children was higher than that of studies with adolescents and adults combined, $z=3.58$, $p<0.001$, and the difference between the latter two age groups was also significant, $z=2.36$, $p=0.009$. The confidence intervals in Table 1 further show that adults benefit more from virtual investigations than physical investigations, whereas children and adolescents benefit as much from either mode of investigation.

The summary effect also depended on the site where a study was carried out. Studies conducted under researcher-controlled circumstances had a significantly higher mean effect size than studies performed in more authentic settings such as a regular classroom. The direction of these effect sizes implies that students benefit more from virtual investigation if their research is situated in authentic settings guided by regular classroom teachers. But when students' inquiry takes place in a quiet space under the surveillance of a proctor, physical investigation is more effective than virtual investigation. It should be noted that the distribution of studies among these two categories was rather skewed and may have impacted the findings.

6. Discussion

The summary effect of the 35 primary studies included in this meta-analysis indicates that physical and virtual investigation are generally equally effective in promoting students' conceptual knowledge of STEM-related topics. This outcome confirms the tentative conclusion from descriptive reviews (Ma and Nickerson, 2006; De Jong et al., 2013; Zacharia, 2015) and implies that the true effect, although slightly in favor of virtual investigation, is close to zero. The fact that this result was independent of the year in which a study was published further suggests that technological advancements have no impact on how much knowledge students acquire from virtual investigation relative to physical investigation. In other words, computer simulations from the early 2000s are as productive to concept learning as contemporary virtual laboratories with highly realistic 3D rendering.

However, the equivalence of investigation modes does not apply to all learning situations. Adults, for example, benefit more from virtual investigation than physical investigation, while no such

TABLE 1 Results of the moderator analyses.

	<i>k</i>	<i>N</i>	<i>g</i>	95% CI	<i>Q</i>	<i>p</i>	<i>I</i> ²
Assessment type					2.48	0.479	0.00
Multiple choice	12	996	-0.20	[-0.42, 0.02]			
Constructed response	16	1,119	-0.02	[-0.37, 0.32]			
Combined	3	336	-0.04	[-0.26, 0.17]			
Performance based	4	852	-0.42	[-0.99, 0.15]			
Tactile feedback					4.51	0.034	77.83
Relevant	15	1,334	0.09	[-0.24, 0.42]			
Irrelevant	20	1969	-0.30	[-0.49, -0.11]			
Student age					8.79	0.012	88.62
Children	10	801	0.22	[-0.25, 0.69]			
Adolescents	8	639	-0.04	[-0.17, 0.08]			
Adults	17	1863	-0.38	[-0.59, -0.17]			
Publication year					0.01	0.967	0.00
2000–2010	8	556	-0.13	[-0.36, 0.10]			
2011–2021	27	2,747	-0.14	[-0.37, 0.10]			
Research setting					6.18	0.013	83.82
Research lab	4	220	0.81	[0.01, 1.61]			
Classroom	31	3,083	-0.24	[-0.39, -0.09]			

benefit was found in adolescents and children. Whether the comparable effectiveness of physical and virtual investigation for younger learners is attributable to their developmental stage or a lack of experience with the objects being investigated (De Jong et al., 2013) cannot be concluded from our meta-analysis. Theoretical evidence supports the former option, but the observed superiority of virtual investigation in adults supports the latter. Future research could resolve this discord by comparing the knowledge gains of children and adults in an investigation with familiar and unfamiliar tangible objects.

The relative effectiveness of investigation modes also depends on the substantive contribution of tactile feedback. Virtual experimentation is more effective when the focal variables in an investigation cannot be experienced by touch. But when tactile cues do provide relevant information, physical investigation is just as effective. This differential effect is in line with the additional channel theory, which assumes that information from touch is processed in a distinct part of the human brain and, hence, reduces cognitive load during learning. A direct assessment of the latter claim could unfortunately not be made here because none of the included studies measured students' cognitive load—which is remarkable because quite many studies mentioned cognitive load reduction as one of the advantages of either physical or virtual investigation.

Implications for theories of embedded cognition are less straightforward. On the one hand, our results lend no direct support to the notion that physical manipulation promotes conceptual understanding since manipulating virtual objects on a computer screen was generally equally effective. On the other hand, the data do not disqualify the embodiment idea either because the physical experience of touch could have compensated for the absence of the affordances of virtual investigation, such as simplifying, annotating,

and visualizing concepts and processes. The value of such features was demonstrated by Lee et al. (2006), who found that separated screen displays and optimized visual representations enhance middle school students' conceptual understanding.

Virtual experimentation environments can also be augmented by haptic feedback. Previous reviews have shown that incorporating haptics can produce significant gains in students' conceptual knowledge (Minogue and Jones, 2006; Zacharia, 2015), but its implementation in educational research and practice is still in its infancy (Luo et al., 2021). Of the few studies we found, none satisfied the inclusion criteria so the comparison between physical investigation with haptic-augmented virtual investigation is yet to be made. Once the body of research has grown, it would be interesting to replicate this meta-analysis and focus specifically on the facilitative role of haptic feedback, in particular when touch conveys information relevant to conceptual understanding.

Beyond contrasting physical and virtual investigations, scholars have started to consider how the two are best combined. The conclusions are still indecisive as some studies favored the physical–virtual sequence (e.g., Winn et al., 2006), other studies the reverse order (e.g., Toth et al., 2014) or reported no difference between both sequences (Flegr et al., 2023). The results of our meta-analysis suggest that starting with physical investigation is preferred when students have an insufficient tactile experience with the concepts or materials being studied, which is often the case with children. When virtual investigation precedes physical investigation, students can benefit from the unique affordances of virtual investigation to efficiently acquire basic knowledge and then deepen and broaden this understanding by investigating the same concepts in more authentic (i.e., 'messy') physical contexts. Although our results provide no direct implications for this option, it seems appropriate for use with adolescents and adults.

Several limitations should be considered when interpreting the results of our meta-analysis. One constraining factor is that the number of included studies was quite small and disproportional across STEM domains. In total, 23 of the 35 studies (66%) were carried out in physics classes, so the current findings do not necessarily apply to other domains. In a similar vein, very few included studies assessed learning outcomes beyond conceptual knowledge, such as students' inquiry skills or their understanding of the nature of science. With a larger set of studies, these outcome measures could have been analyzed to paint a more complete picture of the relative effectiveness of physical and virtual investigations. On a related matter, non-cognitive learning outcomes such as student motivation could be examined to establish whether students are equally interested in doing physical and virtual investigations. Finally, our meta-analysis did not attend to the role of the teacher. This leaves questions regarding whether teachers guide their students equally well during physical and virtual investigations. Research answering questions like these could provide valuable explanations for the relative effectiveness of both modes of investigation.

To conclude, although physical and virtual investigation are generally equally beneficial to promote students' conceptual understanding, the virtual variant is preferred when students are over 18 and have to investigate concepts for which tactile feedback is substantively irrelevant. We, therefore recommend university teachers and adult educators to let students investigate with virtual material and equipment, in particular when the research is conducted in regular classrooms; a switch to physical investigations can be considered if tactile feedback provides relevant sensory information about the concepts being studied or when students conduct their investigation under well-controlled circumstances. Elementary school and high school teachers can decide on a case-by-case basis whether to opt for physical or virtual investigation. They can base their choice on personal preferences and pragmatic considerations while bearing in mind that virtual investigation is more effective when it is not possible to work one-on-one with individual or small groups of students.

References

- References marked with an asterisk (*) indicate studies included in the meta-analysis.
- * Abdel-Maksoud, N. F. (2018). When virtual becomes better than real: Investigating the impact of a networking simulation on learning and motivation. *Int. J. Educ. Pract.* 6, 253–270. doi: 10.18488/journal.61.2018.64.253.270
 - * Ajredini, F., Izairi, N., and Zajkov, O. (2013). Real experiments versus PhET simulations for better high-school students' understanding of electrostatic charging. *Eur J Phys Educ* 5, 59–70. doi: 10.20308/ejpe.v5i1.63
 - Alfieri, L., Brooks, P. J., Aldrich, N. J., and Tenenbaum, H. R. (2011). Does discovery based instruction enhance learning? *J. Educ. Psychol.* 103, 1–18. doi: 10.1037/a0021017
 - Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annu. Rev. Psychol.* 63, 1–29. doi: 10.1146/annurev-psych-120710-100422
 - * Baki, A., Kosa, T., and Guven, B. (2011). A comparative study of the effects of using dynamic geometry software and physical manipulatives on the spatial visualisation skills of pre-service mathematics teachers. *Br. J. Educ. Technol.* 42, 291–310. doi: 10.1111/j.1467-8535.2009.01012.x
 - * Barrett, R., Gandhi, H. A., Naganathan, A., Daniels, D., Zhang, Y., Onwunaka, C., et al. (2018). Social and tactile mixed reality increases student engagement in undergraduate lab activities. *J. Chem. Educ.* 95, 1755–1762. doi: 10.1021/acs.jchemed.8b00212
 - Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639
 - Bell, R. L., Smetana, L., and Binns, I. (2005). Simplifying inquiry instruction. *Sci. Teach.* 72, 30–33.
 - Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2009). *Introduction to meta-analysis*. Hoboken, NJ: Wiley.
 - * Brown, S. E. (2007). *Counting blocks or keyboards? A comparative analysis of concrete versus virtual manipulatives in elementary school mathematics concepts*. (Unpublished Master's Thesis). Detroit (MI): Marygrove College.
 - Carbonneau, K. J., Marley, S. C., and Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *J. Educ. Psychol.* 105, 380–400. doi: 10.1037/a0031084
 - * Chang, W. L., Yuan, Y., Lee, C. Y., Chen, M. H., and Huang, W. G. (2013). Using magic board as a teaching aid in third grader learning of area concepts. *Educ. Technol. Soc.* 16, 163–173.
 - Chao, J., Chiu, J. L., DeJaegher, C. J., and Pan, E. A. (2016). Sensor-augmented virtual labs: Using physical interactions with science simulations to promote understanding of gas behavior. *J. Sci. Educ. Technol.* 25, 16–33. doi: 10.1007/s10956-015-9574-4
 - * Chen, S., Chang, W. H., Lai, C. H., and Tsai, C. Y. (2014). A comparison of students' approaches to inquiry, conceptual learning, and attitudes in simulation-based and microcomputer-based laboratories. *Sci. Educ.* 98, 905–935. doi: 10.1002/sce.21126
 - * Chien, K. P., Tsai, C. Y., Chen, H. L., Chang, W. H., and Chen, S. (2015). Learning differences and eye fixation patterns in virtual and physical science laboratories. *Comput. Educ.* 82, 191–201. doi: 10.1016/j.compedu.2014.11.023
 - * Chini, J. J., Madsen, A., Gire, E., Rebello, N. S., and Puntambekar, S. (2012). Exploration of factors that affect the comparative effectiveness of physical and virtual

Data availability statement

The dataset analyzed for this study is available upon request to the corresponding author.

Author contributions

SM: conceptualization, methodology, systematic search and coding of studies, and writing—reviewing and editing. AL: conceptualization, methodology, validation, formal analysis, writing—original draft, visualization, and supervision. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2023.1163024/full#supplementary-material>

- manipulatives in an undergraduate laboratory. *Phys. Rev. Spec. Top. Phys. Educ. Res.* 8:010113. doi: 10.1103/PhysRevSTPER.8.010113
- * Darrach, M., Humbert, R., Finstein, J., Simon, M., and Hopkins, J. (2014). Are virtual labs as effective as hands-on labs for undergraduate physics? A comparative study at two major universities. *J. Sci. Educ. Technol.* 23, 803–814. doi: 10.1007/s10956-014-9513-9
- De Jong, T., Lazonder, A. W., Chinn, C. A., Fischer, F., Gobert, J., Hmelo-Silver, C., et al. (in press). *Let's talk evidence – The case for combining inquiry-based and direct instruction*. Educational Research Review.
- De Jong, T., Linn, M. C., and Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science* 340, 305–308. doi: 10.1126/science.1230579
- De Jong, T., Sotiriou, S., and Gillet, D. (2014). Innovations in STEM education: The go-lab federation of online labs. *Smart Learn Environ.* 1:3. doi: 10.1186/s40561-014-0003-6
- Dewey, J. (1900). *The school and society*. Chicago, IL: The University of Chicago Press.
- Egger, M., Smith, G. D., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Br. Med. J.* 315, 629–634. doi: 10.1136/bmj.315.7109.629
- * Ekmekci, A., and Gulacar, O. (2015). A case study for comparing the effectiveness of a computer simulation and a hands-on activity on learning electric circuits. *EURASIA J. Math. Sci. Technol. Educ.* 11, 765–775. doi: 10.12973/eurasia.2015.14384
- Flegr, S., Kuhn, J., and Scheiter, K. (2023). When the whole is greater than the sum of its parts: Combining real and virtual experiments in science education. *Comput. Educ.* 197:104745. doi: 10.1016/j.compedu.2023.104745
- Gallese, V., and Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cogn. Neuropsychol.* 22, 455–479. doi: 10.1080/0264329042000310
- * Gecu-Parmaksiz, Z., and Delialioğlu, Ö. (2019). Augmented reality-based virtual manipulatives versus physical manipulatives for teaching geometric shapes to preschool children. *Br. J. Educ. Technol.* 50, 3376–3390. doi: 10.1111/bjet.12740
- * Gibbard, L. L., and Salajan, F. (2009). A novel interactive online module in a traditional curriculum through a blended learning approach. *Electr. J. E-Learning* 7, 301–308.
- * Hannel, S. L., and Cuevas, J. (2018). A study on science achievement and motivation using computer-based simulations compared to traditional hands-on manipulation. *Georgia Educ. Res.* 15, 40–55. doi: 10.20429/ger.2018.15103
- * Hawkins, I., and Phelps, A. J. (2013). Virtual laboratory vs. traditional laboratory: Which is more effective for teaching electrochemistry? *Chem. Educ. Res. Pract.* 14, 516–523. doi: 10.1039/C3RP00070B
- Hedges, L. V., and Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychol. Methods* 9, 426–445. doi: 10.1037/1082-989X.9.4.426
- * Hensen, C., Glinowiecka-Cox, G., and Barbera, J. (2020). Assessing differences between three virtual general chemistry experiments and similar hands-on experiments. *J. Chem. Educ.* 97, 616–625. doi: 10.1021/acs.jchemed.9b00748
- * Husnaini, S. J., and Chen, S. (2019). Effects of guided inquiry virtual and physical laboratories on conceptual understanding, inquiry performance, scientific inquiry self-efficacy, and enjoyment. *Phys. Rev. Phys. Educ. Res.* 15:010119. doi: 10.1103/PhysRevPhysEducRes.15.010119
- * Jaakkola, T., and Nurmi, S. (2008). Fostering elementary school students' understanding of simple electricity by combining simulation and laboratory activities. *J. Comput. Assist. Learn.* 24, 271–283. doi: 10.1111/j.1365-2729.2007.00259.x
- * Kapici, H. O., Akcay, H., and De Jong, T. (2019). Using hands-on and virtual laboratories alone or together—which works better for acquiring knowledge and skills? *J. Sci. Educ. Technol.* 28, 231–250. doi: 10.1007/s10956-018-9762-0
- Kiefer, M., and Trumpp, N. M. (2012). Embodiment theory and education: The foundations of cognition in perception and action. *Trends Neurosci. Educ.* 1, 15–20. doi: 10.1016/j.tine.2012.07.002
- Kirschner, P. A., Sweller, J., and Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educ. Psychol.* 41, 75–86. doi: 10.1207/s15326985sep4102_1
- Klahr, D., Triona, L. M., and Williams, C. (2007). Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. *J. Res. Sci. Teach.* 44, 183–203. doi: 10.1002/tea.20152
- Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. doi: 10.2307/2529310
- * Lazonder, A. W., and Ehrenhard, S. (2014). Relative effectiveness of physical and virtual manipulatives for conceptual change in science: How falling objects fall. *J. Comput. Assist. Learn.* 30, 110–120. doi: 10.1111/jcal.12024
- Lazonder, A. W., and Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Rev. Educ. Res.* 86, 681–718. doi: 10.3102/0034654315627366
- Lee, H., Plass, J. L., and Homer, B. D. (2006). Optimizing cognitive load for learning from computer-based science simulations. *J. Educ. Psychol.* 98, 902–913. doi: 10.1037/0022-0663.98.4.902
- Lipsey, M. W., and Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Luo, H., Li, G., Feng, Q., Yang, Y., and Zuo, M. (2021). Virtual reality in K-12 and higher education: A systematic review of the literature from 2000 to 2019. *J. Comput. Assist. Learn.* 37, 887–901. doi: 10.1111/jcal.12538
- Ma, J., and Nickerson, J. V. (2006). Hands-on, simulated, and remote laboratories: A comparative literature review. *ACM Comput. Surv.* 38, 1–24. doi: 10.1145/1132960.1132961
- * Martinez, G., Naranjo, F. L., Perez, A. L., Suero, M. I., and Pardo, P. J. (2011). Comparative study of the effectiveness of three learning environments: Hyper-realistic virtual simulations, traditional schematic simulations and traditional laboratory. *Phys. Rev. Spec. Top. Phys. Educ. Res.* 7:020111. doi: 10.1103/PhysRevSTPER.7.020111
- * Merkouris, A., Chorianopoulou, B., Chorianopoulos, K., and Chrissikopoulos, V. (2019). Understanding the notion of friction through gestural interaction with a remotely controlled robot. *J. Sci. Educ. Technol.* 28, 209–221. doi: 10.1007/s10956-018-9760-2
- Minogue, J., and Jones, M. G. (2006). Haptics in education: Exploring an untapped sensory modality. *Rev. Educ. Res.* 76, 317–348. doi: 10.3102/00346543076003317
- * Olympiou, G., and Zacharia, Z. C. (2012). Blending physical and virtual manipulatives: An effort to improve students' conceptual understanding through science laboratory experimentation. *Sci. Educ.* 96, 21–47. doi: 10.1002/sce.20463
- Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *J. Educ. Stat.* 8, 157–159. doi: 10.2307/1164923
- Perkins, K., Podolefsky, N., Lancaster, K., and Moore, E. (2012). Creating effective interactive tools for learning: Insights from the PhET interactive simulations project. In T. Amiel and B. Wilson (Eds.), *Proceedings of EdMedia 2012—World Conference on Educational Media and Technology* (pp. 436–441). Waynesville, NC: Association for the Advancement of Computing in Education. Available at: <https://www.learntechlib.org/primary/p/40781/>
- * Pyatt, K., and Sims, R. (2012). Virtual and physical experimentation in inquiry-based science labs: Attitudes, performance and access. *J. Sci. Educ. Technol.* 21, 133–147. doi: 10.1007/s10956-011-9291-6
- Rangarajan, K., Davis, H., and Pucher, P. H. (2020). Systematic review of virtual haptics in surgical simulation: A valid educational tool? *J. Surg. Educ.* 77, 337–347. doi: 10.1016/j.jsurg.2019.09.006
- Renken, M. D., and Nunez, N. (2013). Computer simulations and clear observations do not guarantee conceptual understanding. *Learn. Instr.* 23, 10–23. doi: 10.1016/j.learninstruc.2012.08.006
- Schneider, B., Krajcik, J., Lavonen, J., Salmela-Aro, K., Klager, C., Bradford, L., et al. (2022). Improving science achievement—is it possible? Evaluating the efficacy of a high school chemistry and physics project-based learning intervention. *Educ. Res.* 51, 109–121. doi: 10.3102/0013189X211067742
- * Sullivan, S., Gnesdilow, D., Puntambekar, S., and Kim, J. S. (2017). Middle school students' learning of mechanics concepts through engagement in different sequences of physical and virtual experiments. *Int. J. Sci. Educ.* 39, 1573–1600. doi: 10.1080/09500693.2017.1341668
- Suurmond, R., Van Rhee, H., and Hak, T. (2017). Introduction, comparison and validation of meta-essentials: A free and simple tool for meta-analysis. *Res. Synth. Methods* 8, 537–553. doi: 10.1002/jrsm.1260
- Sweller, J., Van Merriënboer, J. J. G., and Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* 31, 261–292. doi: 10.1007/s10648-019-09465-5
- * Tarnq, W., Lee, C. Y., Lin, C. M., and Chen, W. H. (2018). Applications of virtual reality in learning the photoelectric effect of liquid crystal display. *Comput. Appl. Eng. Educ.* 26, 1956–1967. doi: 10.1002/cae.21957
- Thomas, R. M. (2005). *Comparing theories of child development, 6th Edn*. Belmont, CA: Thomson/Wadsworth.
- Toth, E. E., Ludvico, L. R., and Morrow, B. L. (2014). Blended inquiry with hands-on and virtual laboratories: The role of perceptual features during knowledge construction. *Interact. Learn. Environ.* 22, 614–630. doi: 10.1080/10494820.2012.693102
- * Triona, L. M., and Klahr, D. (2003). Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments. *Cogn. Instr.* 21, 149–173. doi: 10.1207/S1532690XCI2102_02
- * Wang, T. L., and Tseng, Y. K. (2018). The comparative effectiveness of physical, virtual, and virtual-physical manipulatives on third-grade students' science achievement and conceptual understanding of evaporation and condensation. *Int. J. Sci. Math. Educ.* 16, 203–219. doi: 10.1007/s10763-016-9774-2
- * Winn, W., Stahr, F., Sarason, C., Fruland, R., Oppenheimer, P., and Lee, Y.-L. (2006). Learning oceanography from a computer simulation compared with direct experience at sea. *J. Res. Sci. Teach.* 43, 25–42. doi: 10.1002/tea.20097
- * Yuan, Y., Lee, C. Y., and Wang, C. H. (2010). A comparison study of polyominoes explorations in a physical and virtual manipulative environment. *J. Comput. Assist. Learn.* 26, 307–316. doi: 10.1111/j.1365-2729.2010.00352.x

Zacharia, Z. C. (2015). Examining whether touch sensory feedback is necessary for science learning through experimentation: A literature review of two different lines of research across K-16. *Educ. Res. Rev.* 16, 116–137. doi: 10.1016/j.edurev.2015.10.001

* Zacharia, Z. C., and Constantinou, C. P. (2008). Comparing the influence of physical and virtual manipulatives in the context of the physics by inquiry curriculum: The case of undergraduate students' conceptual understanding of heat and temperature. *Am. J. Phys.* 76, 425–430. doi: 10.1119/1.2885059

* Zacharia, Z. C., and De Jong, T. (2014). The effects on students' conceptual understanding of electric circuits of introducing virtual manipulatives within a physical manipulatives-oriented curriculum. *Cogn. Instr.* 32, 101–158. doi: 10.1080/07370008.2014.887083

* Zacharia, Z. C., Loizou, E., and Papaevripidou, M. (2012). Is physicality an important aspect of learning through science experimentation among kindergarten students? *Early Child. Res. Q.* 27, 447–457. doi: 10.1016/j.ecresq.2012.02.004

* Zacharia, Z. C., and Olympiou, G. (2011). Physical versus virtual manipulative experimentation in physics learning. *Learn. Instr.* 21, 317–331. doi: 10.1016/j.learninstruc.2010.03.001

Zhang, L., Kirschner, P. A., Cobern, W. W., and Sweller, J. (2022). There is an evidence crisis in science educational policy. *Educ. Psychol. Rev.* 34, 1157–1176. doi: 10.1007/s10648-021-09646-1