Title Page -- anything identifying the author should be on this page.

Impact of Graph Technologies in K-12 Science and Mathematics Education

Dermot Francis Donnelly-Hermosillo*, Libby F. Gerard, Marcia C. Linn

Abstract

Graph technologies are now widely available in K-12 science and mathematics classrooms. These technologies have the potential to impact the learning of science and mathematics, especially by supporting student investigations. We use meta-analysis to analyze 42 design and comparison studies involving data from 7699 students spanning over 35 years. In these studies, graphing technologies include computer software such as simulations; online tools such as graph utilities; and sensors such as temperature probes. We characterize the assessments used to measure graphing. We describe the investigative activities that graphing supports including generating hypotheses or predictions, collecting data, analyzing or interpreting data, and reflecting. Studies show that graphing technologies impact learning of mathematics and science topics as well as graphing itself. These technologies are especially advantageous for learning complex topics where students need to conduct investigations to interpret change over time or position such as functions, kinematics, and thermodynamics. Recent studies take advantage of logs of student interactions to study the design of automated guidance for graphing. We discuss the implications of these findings for instruction at the K-12 level.

Keywords: Graph; Learning; Technology; NGSS; CCMS
Declarations of interest: none

cite: Donnelly, D. F., Gerard, L. F., & Linn, M. C. (2020). Impact of Graph Technologies in K-12 Science and Mathematics Education. Computers and Education, 146. https://doi.org/https://doi.org/10.1016/j.compedu.2019.103748

Impact of Graph Technologies in K-12 Science and Mathematics Education

Abstract

Graph technologies are now widely available in K-12 science and mathematics classrooms. These technologies have the potential to impact the learning of science and mathematics, especially by supporting student investigations. We use meta-analysis to analyze 42 design and comparison studies involving data from 7699 students spanning over 35 years. In these studies, graphing technologies include computer software such as simulations; online tools such as graph utilities; and sensors such as temperature probes. We characterize the assessments used to measure graphing. We describe the investigative activities that graphing supports including generating hypotheses or predictions, collecting data, analyzing or interpreting data, and reflecting. Studies show that graphing technologies impact learning of mathematics and science topics as well as graphing itself. These technologies are especially advantageous for learning complex topics where students need to conduct investigations to interpret change over time or position such as functions, kinematics, and thermodynamics. Recent studies take advantage of logs of student interactions to study the design of automated guidance for graphing. We discuss the implications of these findings for instruction at the K-12 level.

Keywords: Graph; Learning; Technology; NGSS; CCMS

1. Introduction

We review research on the impact and value of graph technologies for K-12 science and mathematics learning. We characterize the ways graphing is assessed and the investigative features these technologies support. Graph technologies are widely available in precollege classes where they support a variety of investigative features, such as generating hypotheses or predictions (Mokros & Tinker, 1987; Songer & Linn, 1991), analyzing or interpreting data from multiple sources (Kastberg & Leatham, 2005; Tortosa, 2012), and reflecting on results (McElhaney & Linn, 2011).

Diverse and sophisticated graphing tools allow designers to strengthen graph understanding as part of teaching mathematics or science (Ainsworth, 1999; diSessa, 2004; Greeno & Hall, 1997). Graphing technologies can illustrate relationships between changes in temperature and motion using Microcomputer Based Laboratories (MBL); results from changing variables in simulations of phenomena such as climate change, population growth, or tectonic plate movements; and impacts of changing parameters governing functions in mathematics. Some studies use technology to support graph construction. Others emphasize comprehending features of a graph or labeling graphs (Yeh & McTigue, 2009). Furthermore, students can be challenged to invent graph representations (diSessa, 2004) or create a graph that depicts a narrative such as a hike (Vitale, Lai, & Linn, 2015). Students can simultaneously explore how airbags deploy and how position and motion graphs work (McElhaney & Linn, 2011) or how the parameters of a function impact the graph shape (Berg & Boote, 2017; Berg & Phillips, 1994).

Activities that involve creating and interpreting graphs are central to the U.S. Next Generation Science Standards (NGSS; NGSS Lead States, 2013) and Common Core Mathematics Standards (CCMS; Common Core State Standards Initiative, 2010). These standards focus on integrated understanding and emphasize the use of authentic data across K-12 instruction. The NGSS argue that preparing informed citizens and professionals requires attention to interpretation and design of graphs depicting contemporary dilemmas. For example, several performance e. pectations in the NGSS directly address graphing such as "5-PS1-2. Measure and graph quantities to provide evidence that regardless of the type of change that occurs when heating, cooling, or mixing substances, the total weight of matter is conserved." (NGSS Lead States, 2013, p. 43; See LaDue, Libarkin, & Thomas, 2015, for other NGSS connections to graphs in the high-school context). Likewise, the CCMS addresses graphing with expectations such as "5.G: Graph points on the coordinate plane to solve real-world and mathematical problems." (Common Core State Standards Initiative, p.38). Achieving the NGSS and CCMS requires instruction that incorporates graphs across mathematics and science along with valid and reliable assessments of student graph proficiency (NGSS Lead States, 2013; Wang et al., 2012). Our review investigates how existing research literature is meeting these challenges using meta-analysis techniques for design and comparison studies focused on graph technologies.

Graphs are vital for learning, technical occupations, and public discourse (Arsenault, Smith, & Beauchamp, 2006; Krohn, 1991). They take advantage of the human capacity to visualize large amounts of data in ways that reveal patterns, uncertainty, and critical events (Friel, Curcio, & Bright, 2001). Graph shapes, allow people to infer underlying processes and interactions within systems from individual data points (Shah & Hoeffner, 2002) and to predict future trends (Ellington, 2006; Wang et al., 2012). For example, an analysis of the points on a temperature/time graph can help determine how an object changes temperature over time, and also support predictions that extrapolate the temperature change beyond data within the graph

(Linn, Layman, & Nachmias, 1987). Yet, interpreting graphs is challenging for most people as shown in international comparisons (OECD, 2006) and previous reviews in mathematics education (Cheung & Slavin, 2013; Leinhardt, Zaslavsky, & Stein 1990; Rakes, Valentine, MaGatha, & Ronau, 2010) and science education (Glazer, 2011; Nakhleh, 1994; Shah & Hoeffner, 2002). We build on these reviews to analyze the impact of graphing technologies and the role of investigative features that could add value to graphing technologies. Specifically, we identify and analyze strengths and gaps in use of investigative features that can amplify the impact of graphing technology and improve instruction and understanding in both science and mathematics.

1.1. Characterizing Graphing Instruction

To characterize graphing instruction, we focus on the investigative features described in the research literature to categorize student science and mathematics activities using graphs. We link these investigative features to the broader categorizations of the NGSS science and engineering practices. The NGSS performance expectations include eight science and engineering practices (SEPs): 1. Asking questions and defining problems, 2. Developing and using models, 3. Planning and carrying out investigations, 4. Analyzing and interpreting data, 5. Using mathematics and computational thinking, 6. Constructing explanations and designing solutions, 7. Engaging in argument from evidence, and 8. Obtaining, evaluating, and communicating information (NGSS Lead States, 2013). The practices include investigative features such as “questioning and generating hypotheses, experimenting, designing, and planning, predicting, modeling/visualizing, observing and data collection, analyzing data, interpreting and explaining, developing/evaluating/arguing, reaching conclusions, and communicating findings” (Authors, 2014; See also National Research Council, 2012). These practices and their accompanying investigative features are important for understanding the impact of graph technologies, the ways these technologies support student learning, and the gaps that graph technologies could fill in student learning.

1.2. Measuring Graph Proficiency

To fully capture the value of graph proficiency, we need comprehensive assessments. Our use of the term ‘graph proficiency’ is meant to capture the broad range of roles for graphs articulated in the research literature, e.g., graphicacy, meta-representation, experimentation. Several reviews of the nature of graph-based items in standardized tests reveal that graph proficiency is rarely measured, and that, when it is measured, items often focus on comprehension of graph features such as student ability to locate a point on a graph or to determine whether the graph labels are accurate (Miller & Linn, 2013; Yeh & McTigue, 2009). Choice of assessment may reflect the nature of instruction and could impact the interpretation of learning outcomes. For example, as explained by Leinhardt et al. (1990), “Construction is quite different from interpretation [comprehension]. Whereas interpretation relies on and requires reaction to a given piece of data (e.g. a graph, an equation, or a data set), construction requires generating new parts that are not given.” (p. 12).

Some studies assess *graphicacy*, defined as "proficiency in understanding quantitative phenomena that are presented in a graphical way" (Wainer, 1992, p.16). Graphicacy refers to the ability to read and interpret graphs (Friel & Bright, 1996). Others study *graph sense*, “the ability to recognize components of graphs, speak the language of graphs, understand relationships between tables and graphs, respond to questions about graphs, recognize better graphs, and interpret contextual awareness of graphs” (Delmas, Garfield, & Ooms, 2005, p.2). diSessa (2004) refers to *meta-representational competence* as the ability to choose an appropriate external representation for data or to use novel external representations productively. Many studies assess student experimentation by looking at how they interpret graphs or generate trials in a simulation (Roschelle et al. 2010).

To determine how well outcome measures align with graph proficiency (Pellegrino, Wilson, Koenig, & Beatty, 2014), we analyze the use of three main categories of assessments of graph proficiency: Construction of a graph; critique of a graph, and comprehension of a graph (Lai, Cabrera, Vitale, Madhok, Tinker, & Linn, 2016; Yeh & McTigue, 2009). Measures of graph proficiency vary not only in form (construction, comprehension, and critique), but also in format (multiple-choice, open-response recall, and open-response explanation), and disciplinary focus (mathematics and science). We review all graph technology studies featuring questions about graphs and analyze how form, format, and discipline contribute to our research questions.

1.3. Research Questions

We investigate three research questions. For research questions one and two, we report on the role of graph assessment form (Construction; Comprehension; Critique; Lai et al., 2016) and format (multiple choice and open response) to enrich our analysis in determining how such assessment factors may influence our findings for these two research questions. Our research questions are:

1. *What is the overall impact of instruction supported by graph technology on K-12 students' learning?* We answer this question by conducting a meta-analysis of design studies that analyze graphing instruction using pre/post data measuring graph proficiency.
2. *Does the impact of technology-based graphing instruction differ from the impact of non-technology-based instruction?* We meta-analyze studies that have either pre/post and post-test only results that compare instruction with and without digital technology.

3. *What investigative features characterize the use of K-12 graphing technologies?* We answer this question through a binary scoring of studies based on the presence of a particular investigative feature, such as collecting data, drawing conclusions, reflecting, etc. (See Method for all features).

2. Method

2.1. Identifying Articles on Science and Mathematics Graphing Technology

The first author searched relevant science and mathematics journals to identify articles for this review. We identified 25 journals with a science, mathematics, or a combined science and mathematics focus (See *Figure 1* for journals/Overall logic model). Due to the broad use of the word "graph" across disciplines (demo*graph*y, ethno*graph*ic, mono*graph*, biblio*graph*y, etc.) and consistent with other reviews (Authors, 2014; Kennedy, 2016), a database search would have been unnecessarily time-consuming, without yielding better results than searching relevant journals, and in-text citations within articles.

For each journal, we used the following search parameters: (a) article contains 'graph' AND/OR 'function' AND/OR 'sensor', AND/OR 'microcomputer-based laboratory' AND/OR 'data logging' AND/OR 'probe' (common terms for studies involving graphs and technology), (b) articles published from 1980 to 2018 (1980 is chosen as a cutoff as technology-based approaches to graphing became more common in the late 1980s). Where an online search option was not available for a journal, we searched each issue of the journal for articles using the above terms.

The inclusion criteria for an article in our analysis are (a) it reported data on K-12 student graph learning (Hence, we searched articles for the terms 'K-12', 'Grade', 'Year', ‘Undergraduate’, ‘Graduate’, ‘University’, ‘Elementary’, ‘Primary’, ‘Middle’, ‘High’, and ‘Secondary’), and (b) reported an experimental design study or an experimental comparison study (Hence, we searched for the terms ‘design’, ‘experimental’, ‘pretest’, and ‘posttest’. For each article included based on these two criteria, the first and third author searched the references of these articles to identify other relevant articles.

The exclusion criteria for an article in our analysis are that it reports: (a) a study that lacks any use of technology, (b) a design or comparison study with insufficient data to calculate effect size, (c) case or survey studies, (d) theoretical aspects of graphs only, (e) experts’ understanding of graphs only, (f) college students’ graph proficiency only, (g) professional development of teachers only, (h) analysis of graphs in textbooks only, (i) the public’s knowledge of graphs only, and (j) classroom graphs anecdotally only.

Overall, we identified 542 articles of potential interest. Through applying the inclusion and exclusion criteria, 500 articles were removed from the article count, leaving 42 articles for our analysis (Asterisked in the References). For the 42 articles, there are 19 experimental design studies (*Table 2*) and 23 experimental comparison studies (*Table 3*).

2.2. Assessment Analysis

We analyzed the form, format, design, and connection to the NGSS practices of the assessments in the identified studies. Since studies often included several types of assessments, we analyzed each assessment described in the article separately. For all 42 articles, we categorized the assessments by form: construction, comprehension, and critique (Lai et al., 2016). In addition, we categorized the format for each item in the assessments: multiple-choice, open-response recall, or open-response explanation. Moreover, we categorized each reported item by the investigative practice it measured. We categorized investigative items by their alignment with the NGSS practices since these are the goal of instruction in places that have adopted the NGSS. Finally, we analyzed the design of the assessment, noting whether the article used researcher-designed or standardized items, as such factors have been shown to influence reported outcomes (Cheung & Slavin, 2016).

To fully assess the impact of instruction featuring graphs, whether these include technology or not, requires measuring progress in aspects of mathematics or science that are captured in graphs. Graphs capture changes in variables over time or position, they represent when phenomena change quickly and when they change slowly, and they may capture patterns such as in graphs of whether objects float or sink based on their mass and volume.

2.2.1. Assessment form

Across the first two research questions, we analyze how the form of the graph assessment impacts interpretation of the results. We chose construction, comprehension, or critique based on research on graphing item formats (Lai et al., 2016).

Graph construction asks students to use and interact with information to represent relationships from data sets in graphical form, consistent with the concept of meta-representational competence (Latour, 1990). An example graph construction item is "Draw a velocity graph which shows the object moving away from the origin at a constant velocity." (Kwon, 2002, p. 61). We also coded items as graph construction if they required interpretation of results from automatic output from a probe or sensor, asking the student to determine the conditions of data collection but not the format of the graph (Beichner, 1990; Friedler & McFarlene, 1997). For graph construction, we looked at assessments that either (a) require students to interpret displays of generated graphs, typical of probe and sensor-based technologies that lack student adjustment of variables or (b) require students to construct or manipulate graphs themselves from data provided by the instructor, technology, or generated by students themselves. Search terms for graph construction-based assessments included ‘construct’, ‘draw’, ‘plot’, ‘sketch’, ‘graph’, ‘manipulate’, and ‘variable’.

Graph comprehension has several aspects (*Table 1*). It can involve (a) *graph features*, interpreting scales and data points, (b) *graph patterns or trends*, recognizing the significance of the shape of data and graph characteristics such as breakpoints, maxima, and noise, and (c) disciplinary context, understanding the underlying scientific ideas in a graph (Lai et al., 2016). An example graph interpretation item focused on graph features is “Based on the graph above [Population/time graph included for students], about how many Black-capped Chickadees are there in Cambridge in December?” (Kamarainen et al., 2013, p. 551).

Graph comprehension assessments require explanations of (a) specific data points on a graph, (b) overall trends across data points and (c) a science context. Search terms across articles included ‘point(s)’, ‘locate’, ‘find’, ‘data’, ‘trend’, ‘noise’, ‘features’, ‘interpret’, ‘overall’, ‘context’, and ‘concept’.

[Insert Table 1 here]

Graph critique requires the student to detect flaws or inaccurate implications of graphs. Critique involves arguing from evidence in the graph or evidence from other sources. Critique is important for scientists and mathematicians, and aligns with the NGSS and CCMS (Lai et al., 2016). Critique is also essential for citizens who might be misled by persuasive messages featuring graphs. The aspects of graph comprehension in *Table 1*, alongside graph construction and critique require both overlapping and unique capabilities (Ainley, Nardi, & Pratt, 2000). An example graph critique item is “Jon took a trip on his bicycle. Identify which of [these] three graphs could possibly represent a bicycle trip [One of the graphs illustrates backward motion in time]. Explain your reasoning." (Vitale et al., 2015, p. 1432-1433).

Graph critique assessments require critiques of one or multiple graphs. Search terms included 'argue', 'evidence', 'justify', 'compare', 'contrast', 'distinguish', and 'claim'.

2.2.2. Assessment Format.

Assessment format and design contributes to the validity of measures of graphing (Berg & Boote, 2017; Berg & Phillips, 1994; Lee, Liu, & Linn, 2011; Liu, Lee, & Linn, 2011). For example, multiple-choice instruments may have features that reinforce the intuition to view graphs as pictures (Berg & Boote, 2017, p.13). Open response can provide more valid indicators of student learning than multiple-choice (Berg & Boote, 2017; Berg & Phillips, 1994; Lee, et al., 2011).

We categorized items into three formats: multiple-choice, open-response recall, or open-response explanation. An example of a multiple-choice item is "Which one of the following equations belongs to the graph above [Function graph provided to students]? A) $x^2 + 1$, B) $x^2 - 1$, C) $-x^2 - 1$, D) $-x^2 + 1$, E) $2x^2 - 1$" (Erbas, Ince, & Kaya, 2015, p. 306). An example of an open-response recall item is "Based on the graph above, about how many Black-capped Chickadees are there in Cambridge in December?" (Kamarainen et al., 2013, p. 551). An example of an open-response explanation is "Describe what happened between the driver and airbag in this crash [Velocity/time graph provided to students]. Was the driver injured by the airbag?" (McElhaney & Linn, 2011, p. 753).

We categorized assessment format for all design and comparison studies. If a study had a two-tier assessment (for example, multiple-choice and open-response explanation items), it was coded for both assessment formats. Similarly, if a study included components of graph construction and graph comprehension, it was coded for both assessment types.

2.2.3. Assessment design.

We categorized assessments as standardized or researcher designed. Most of the studies identified for this meta-analysis include researcher-generated assessments and are focused on the classroom/student level (Lipsey et al., 2012). This finding is unsurprising since reviews of standardized assessments show that they include few graphing items and that those featuring graphs often ask only about graph features (Yeh & McTigue, 2009). Thus, most standardized items are likely to be poorly aligned with the advantages of technology-enhanced instruction. A few studies used or customized previously reported assessments such as TOGS - Test of Graphing in Science (Adams & Shrum, 1990).

2.3. Data Sources and Analysis for Research Questions

2.3.1. RQ1: Impact of graph instruction with technology on student learning.

To assess the impact of graph instruction with technology on student learning (RQ1), we compute and average the effect sizes of all (significant and nonsignificant) measures of graph proficiency reported in pre/post-test design studies. Thus, we capture the impact of varied technology and instructional approaches on graph proficiency (*Table 2* (See also *Appendix A* for greater detail); n = 19 studies).

2.3.2. RQ2: Comparison of graph instruction with or without technology.

For RQ2, we investigate the pooled effect size (for significant and non-significant results) of studies comparing technology to non-technology approaches using pre/post-test and post-test only comparison studies (*Table 3* (See also *Appendix B* for greater detail); n = 23 studies). We use the term 'non-technology' to describe conditions where digital technology is absent and students follow a standard curriculum that could include tools such as stopwatches, thermometers, data tables, paper and pencil, etc.

2.3.3. RQ3: Investigative features characterizing graphing instruction.

We analyzed instruction for the investigative practice it emphasized (RQ3). We searched for investigative features identified from the literature in a previous review (Authors, 2014). Further, we grouped these investigative features based on the eight NGSS science and engineering practices (SEPs). These features include hypothesis, questions, or predictions (SEP1), embodied learning and modeling (SEP2), planning an experiment, choosing/manipulating variables, collecting data, selecting resources (SEP3), analyze or interpret, draw or annotate (SEP4), construct graphs (SEP5), explain content (SEP6), make an argument (SEP7), and draw conclusions, reflect, and present (SEP8). We reviewed and scored all studies for these investigative features by using various search terms and also through carefully reading the articles (See *Table 5* for relevant search terms). For example, when reviewing an article for SEP7 (Engaging in argument from evidence), we used search terms including: argu* [e, ing, ment], refut*, claim*, debat*, consensus, etc.

2.3.4. Article analysis.

In consultation with the other authors, the first author conducted article analysis to investigate the research questions. Meta-analysis was conducted using Comprehensive Meta-Analysis™ software. Effect sizes (Hedges' g) were computed using pre/post means, standard deviations, sample size, and pre/post correlations.

We used Hedges' (Hedges & Olkin, 1985) categories of low (0-0.29), medium (0.3-0.59), and large effect sizes (0.6 or higher). It is important to note that the relative magnitude of such categorizations should be considered tentatively, as "appropriate norms are those based on distributions of effect sizes for comparable outcome measures from comparable interventions targeted on comparable samples" (Lipsey et al., 2012, p. 4). For example, some studies find larger effect sizes for researcher-developed assessments across all grade levels (>0.5) compared to effect sizes for standardized assessments that are rarely larger than 0.3 (Lipsey et al., 2012). Additionally, effect sizes are commonly higher for quasi-experimental versus randomized studies, and small sample sizes versus large sample sizes, and are also influenced by grade level (Cheung & Slavin, 2016).

We coded studies as quasi-experimental when students were intentionally placed in a treatment based on particular attributes (gender, socio-economic status, teacher reported academic performance, etc.). We coded studies as randomized when the treatment was randomly assigned at the class level or at the student level (Both levels for each study are specified in *Appendix B*). Most classroom studies for K-12 are randomized at the class level in order to minimize the influence of one treatment on the other treatment, and hence our decision to report it as randomized. For sample sizes, we assigned studies with less than 250 students as small samples and as large for studies with more than 250 students, similar to Cheung and Slavin (2016). For grade level analysis, we grouped studies by elementary (Grades K-5) and secondary (Grades 6-12). The reasoning for such grouping is that science/graphing are not taught consistently until middle school so the instructional context is different for middle and high schools compared to elementary school. Studies that do report psychometric properties have acceptable levels of internal consistency.

Based on these characteristics of the assessments, we predict pooled effect sizes of 0.3-0.6. Effect sizes were determined for independent samples (Wilson, 2009). Thus, if a study reported two independent experiments, we included the two effect sizes in the analysis. Pre/post correlations are commonly not reported and thus we used an estimated Pearson correlation between different outcomes of 0.36 that is consistent with typical correlations among multiple-choice, open-response, and a mix of multiple-choice and open-response items (Lee et al., 2011). A fail-safe N was calculated across all analysis to address publication bias, in particular for file drawer studies (Orwin, 1983).

We report significance levels for random effects since the studies are heterogeneous with regard to population characteristics, grade level, teacher experience, and features of the instruction (Borenstein, Hedges, Higgins, & Rothstein, 2010). Random effects analyses assume each study is an estimate of the population mean and therefore accord equal status to each study (rather than weighting results by sample size, as is the case for fixed effects). With random effects, when sample sizes vary greatly, there is likely to be a moderator effect for sample size as we report. For fixed effects, sample size is weighted and studies with large sample sizes contribute much more to the final computation than those with small samples. We report both random and fixed effects significance levels in *Table 4* for completeness. Since it could be argued that the assessments are homogeneous, we pay attention to fixed effects in moderator analyses for assessments (*Table 4*).

3. Results

We summarize the nature of the assessments in the corpus of studies to provide context for the meta-analysis. Then we report on the meta-analysis findings.

3.1. Assessment Format

We categorized the format of the assessments as construction, comprehension, and critique for all design and comparison studies. The majority of studies (*Appendices A and B*; 42 studies) measure components of graph comprehension (Graph Features: 38 studies, 90%; Graph Trends: 38 studies, 90%; and Science Context: 25 studies, 60%). The science assessments measured a range of disciplinary contexts. In mathematics, most assessments measured aspects of functions while a few included applying functions to a specific example.

Graph construction assessments mostly involved students manipulating or constructing a graph with or without technology (31 studies; 74%) rather than having students passively observe a graph constructed through video or using probes/sensors; 20 studies, 48%).

Only three out of 42 studies (7%) assessed graph critique. This finding represents a gap in the focus of the assessments that is echoed in the instruction. Critique of graphs using content knowledge is an important aspect of graph understanding. Graphs are commonly used in persuasive messages and students need the ability to view these messages critically.

The form of questions for most studies consisted of multiple-choice assessments (27 studies, 64%) and open-response explanation assessments (23 studies, 58%). Many articles included both of these assessment types. There are fewer open-response recall assessments than other formats (13 studies, 31%). Overall, open-response measures generally require more explanation than recall items. Although open response may be crucial for measuring deep understanding, these items can be difficult to score. Emerging technologies are adding automatic scoring of graph assessments (e.g., Roschelle, et al, 2012; Vitale, et al., 2018).

Most studies included researcher-designed assessments of graphing, attesting to the recent rapid growth of graphing technologies and need for assessments aligned with these opportunities. Researcher-designed assessments were also necessary since standardized tests have few graph items (Yeh & McTigue, 2009). Researcher-designed assessments of graphing aligned with the instruction in the units, tapping into the advantages of graphing in the units addressed. They often required students to combine their disciplinary and graph understanding. Disciplinary items, in contrast, were often multiple choice or recall items.

3.2. Impact of Graphing Technologies on Learning (RQ1; Design Studies)

Overall, we found that in design studies, instruction using technology impacted graph proficiency. We found an effect size of 0.59 (Medium effect size, 95% CI [Random: 0.57, 0.82]; *Table 2*/See *Appendix A* for greater detail) for instruction across discipline, form, and format of the assessments. We identified 31 effect sizes from 19 design studies of 2293 students. We calculated a classic fail-safe N (Orwin, 1983) of 8696 ($p < .001$); thus, it would take 8696 additional studies with effect sizes of zero to reduce the z-value (32.88) of the observed studies to reach statistical non-significance. Power analysis resulted in $1–\beta$ error probability of 1.0 for low, moderate, and high heterogeneity, confirming the power of the meta-analysis to detect low through high effects.

[Insert Table 2 here]

The 19 studies mostly investigated middle school science (10 middle school studies, four high school, three elementary, and two studies across elementary, middle, or high school) and focused on kinematics (seven studies of kinematics; three of thermodynamics, and nine from other topic areas including function, ecosystems, water quality, climate change, plant growth, and others with multiple topics).

The technology-based approaches included graph generation (either during experimentation or afterward) using various tools such as simulations or probes, data collection with probes and sensors (Kamarainen et al., 2013; Kwon, 2002), graphing software (Kramarski, 1999), virtual laboratories (Chao, Chiu, DeJaegher, & Pan, 2016), and a tablet applet (Purba &

Hwang, 2017). Studies featured various scaffolds including automated scoring of graphs (McElhaney & Linn, 2011; Vitale et al., 2015). One study featured an online school-community partnership with scientists (Peterman, Cranston, Pryor & Kermish-Allen, 2015).

Seventeen out of the 19 design studies (89%) used researcher-generated assessment items while 16 of the studies (84%) reported either all or some of their assessment items. The internal consistency of the items was reported across measures in two of the 19 studies, ranging from 0.70 to 0.87.

Thus, design studies aimed to improve student graph proficiency using technology are effective across mathematics and science and for multiple topics. Researchers generally conduct design research to refine the impact of new graphing technologies. As a result, it is worthwhile to delve deeper into the impacts of graphing technologies by examining comparison studies that are likely to use instruction that has been improved using design research (RQ2).

3.3. Impact of Graphing Technologies on Learning (RQ2; Comparison Studies)

Overall, in comparison studies, we found that instruction featuring digital technologies was more effective than instruction without digital technologies (See *Table 3* and *Appendix B* for more detail; 23 studies). We found an effect size of 0.43 (Medium effect size; 95% CI [Random: 0.33, 0.66]; based on 44 effect sizes from 23 studies for 5406 students[1]). We calculated a classic fail-

---

[1] We initially ran an analysis for studies with pre/post-test designs and an analysis for studies with post-test comparisons. As we found consistent results across the two design types (Pre/Post-test: ES = 0.48 (23 effect sizes; 15 studies); Post-test only: ES = 0.34 (21 effect sizes; 10 studies); Medium Effect Sizes), we combined the results of all studies into one post-test analysis. Note: Some studies included both pre/post and post-test results so the number of studies (15 and 10) do not add to 23 studies.

safe N of 2942 ($p < .001$); thus, it would take 2942 extra studies with effect sizes of zero to reduce the z-value (16.14) to statistical non-significance. Power analysis resulted in $1–\beta$ error probability of 1.0 for low, moderate, and high heterogeneity, confirming the power of the meta-analysis to detect low through high effects.

[Insert Table 3 here]

The digital technologies in these studies included computer software/online tools such as simulations (11 studies; e.g. Ardac & Sezan, 2002; Cavanaugh et al., 2008; Chiu & Linn, 2013); probes and/or sensors (eight studies; e.g. Adams & Shrum, 1990; Ates & Stevens, 2003; Deng, Chen, Chai, & Qian, 2011); and other tools involving robotics, video analysis, and graph calculators (e.g. Beichner, 1990; Huntley et al., 2000; Park, 2015). Fourteen of the 23 studies (61%) reported either all or some of their assessment items. Nine of the 23 studies (39%) included the internal consistency of their assessments (range is from 0.71 to 0.92; See *Appendix B*).

We found a trend for discipline moderating the effect of technology, suggesting that graphs are more effective in helping students learn some topics than others. Specifically, 15 of the 23 studies focused on science (mostly kinematics and thermodynamics) with a combined effect size of 0.36. The eight mathematics studies focused on functions (seven studies) and probability (one study) with a combined effect size of 0.47. Moderator analysis of discipline split by science or mathematics produced total between heterogeneity Q-Value of 3.57 (random, df = 1, $p = 0.059$) in favor of mathematics. This finding is consistent with the similarity of disciplinary focus in mathematics studies compared to science studies. See *Table 4* for details.

[Insert Table 4 here]

There is no significant moderating effect of grade level for technology although there is a trend for larger effects in elementary grade studies (Elementary (Grades K-6): ES = 0.62; Secondary (Grades 7-12): ES = 0.43; Q-Value = 1.36 (Random effect; $p = 0.243$)). This is consistent with the likely novelty of graphing technologies in elementary grades.

We found a moderating effect for assessment design (researcher-generated versus standardized). Researcher-designed assessments were more likely to detect an impact of technology (effect size = 0.56) than standardized measures (effect size = -0.05), possibly because the researcher designed measures were well aligned with the instruction. Researcher designed assessments often asked students to distinguish among graphs or to interpret graphs to explain science or mathematics phenomena. Although there were no overall effects for standardized measures, some standardized measures did detect effects. For example, Yang and Heh (2007) show pre/post gains using the Process Skills Test for the experimental treatment (an online virtual physics laboratory), but show no gains for the conventional treatment. Similarly, Ates and Stevens (2003) show pre/post gains on the I-TOGS (Test of Graphing Skills; an open-response format of the multiple-choice TOGS) and Lee and Thomas (2011) show gains with written (open-response) state assessments.

We found differences in effect size for assessment format, as studies with open-response explanations showed a larger effect size (ES = 0.52) than studies with only multiple choice or open-response recall (ES = 0.35). However, a moderator analysis revealed no significant differences. Overall these results illustrate the need for research on methods for assessing instruction featuring graphing.

To investigate relationships between the separate moderators for RQ2, we conducted a meta-regression analysis (*Table 4*). This revealed a significant impact of assessment design for technology and graphing (RQ2). For RQ2, comparing technology versus non-technology approaches to graphing, researcher-generated assessments results in a 0.60 SD larger effect size on graphing instruction compared to standardized assessments.

In summary, treatments using graphing technologies have a greater effect on student outcomes than treatments without graphing technologies. The primary moderating effect was for type of assessment. Researcher designed, constructed response assessments detected more substantial effects, possibly due to better alignment with treatments using technology. The researcher-designed tests could also, themselves, serve as learning events if they engaged students in activities similar to those in the instruction.

3.4. Investigative Features Characterizing Graphing Technology Use (RQ3; All Studies)

All 42 studies were scored using binary coding on investigative features (See *Table 5*). Findings indicate that graphing technologies target particular investigative features more than others. Graphing technologies are primarily used to target investigative features of analyzing or interpreting graphs (100%), drawing conclusions about graphs (83%; 35/42), constructing graph data through data collection display or interaction (98%), and modeling graphs (95%; 40/42). Graphing technologies also commonly address investigative features of choosing/manipulating graph variables (71%; 30/42), collecting data for graphs (71%; 30/42), and explaining content relevant to graphs (79%; 33/42).

Despite the presence of these investigative features, many important investigative features are less prevalent in graphing instruction. Such features include planning an experiment using graphs (52%; 22/42), selecting resources (types and size of equipment, amount of volumes, etc.) for an experiment with graphs (21%; 9/42), drawing or annotating on graphs (2%, 1/42), hypotheses, questions, or predictions (64%; 27/42), intentional reflection steps on graph data (52%; 22/42), presenting on graph data through reports, letters, posters, etc. (26%; 11/42), intentional argument steps on graph data (21%; 9/42), and analyzing graphs through embodied learning (14%; 6/42) (see Figure 2).

[Insert Table 5 here]

These analyses show the potential of instruction featuring graph technologies to support students to conduct investigations. The studies address investigative features unevenly (see *Figure 2*). A few gaps in instruction are noteworthy. Specifically, engaging in argument from evidence is rarely part of graphing instruction in spite of its importance. In addition, communication of results could be strengthened in the context of graphing instruction.

4. Discussion

This meta-analysis indicates that graphing technologies improve learning in general (ES = 0.59; design studies) and improve learning when compared to non-graph technology approaches (ES = 0.43; comparison studies). Graphing technologies provide immediate, visual feedback about complex phenomena and support autonomous investigations that are difficult to achieve without technology. Technology has benefits over non-technology approaches in helping students connect physical phenomena with the representations displayed on graphs by directly linking sensors measuring temperature, motion, or chemical concentrations to scientific phenomena (Beichner, 1990; Linn et al., 1987; Roschelle et al., 2010). These technologies can help students distinguish between a picture of the situation such as a biker riding up and down a hill and a graph of the same situation (Mokros & Tinker, 1987). As a result, graph technologies can deepen understanding of scientific phenomena.

To realize the potential of advances in graphing technologies assessments used in the studies, the meta-analysis shows the advantage of technological features that allow students to conduct their own investigations. Technology can support rapid construction and modification of representations, thus revealing the trends and patterns in data (Vitale et al., 2015). Simulations can connect graphs to complex concepts such as climate change or car collisions (Adams & Shrum, 1990; McElhaney & Linn, 2011). Technologies can support students to autonomously conduct investigations (Beichner, 1990; Vitale et al., 2015).

In mathematics for example, Roschelle et al. (2010) analyzed a SimCalc simulation that links a position versus time line graph with an animation of characters jogging. Students were guided to make predictions about what they expect to happen, they then observed and compared how a given feature in one representation (e.g., a fast, forward jogging speed) is depicted in the alternative representation (i.e., a steep positive slope). They use this evidence to develop sophisticated explanations of graphs. The authors argue that when programs like SimCalc are combined with scaffolds to encourage students to make sense of the visual feedback they help students link the graph to the concrete situation (Roschelle et al., 2010). Both teachers and software supports can guide students to take advantage of this technology.

Similarly, in science, design studies (*Table 2*: Applebaum et al., 2011; McElhaney & Linn, 2011; Vitale et al., 2015) use simulations that provide visual feedback that is not available in typical instruction. For example, a design study by Vitale et al. (2015) embedded a simulation in a learning environment that scaffolded the students to help them understand position and time graphs. In the Vitale et al. study, students predicted the position and time graph for a story about a hike where the participants turn back when they encounter a bear and then complete their hike. The digital technology compared an accurate animation of the hike to the animation of the hike drawn by the students. Vitale et al (2015) argued that the animation allowed students to distinguish between their own graph and a graph that captured the bear story. Thus, the animation helped students recognize that the graph was not a picture by providing hints when students attempted to draw a graph that went back in time. The animation also helped students recognize that time is continuous and that a line without a slope means the person hiking is standing still. Both Vitale et al. (2015) and Roschelle et al. (2010) illustrate the value of visualizations for developing understanding of position and time graphs and highlight the key contributions of graph technology in supporting students to plan and conduct investigations, particularly in making predictions.

Several studies show that even when digital technologies do not improve graph proficiency, they still support investigations by increasing efficiency of data collection. Studies using LoggerPro software in chemistry (Ates & Stevens, 2003), data collection with probes or sensors (Adams & Shrum, 1990; Deniz & Dulger, 2012), and VideoGraph for motion graphs (Beichner, 1990) do not improve outcomes compared to no technology yet can accelerate student learning through more efficient pathways. Students gathering data with digital technologies finish activities faster than students using typical approaches (Beichner, 1990). These technologies reduce the physical demands of typical graph construction (Adams & Shrum, 1990) and teachers can invest the additional time in other student tasks (Beichner, 1990). Thus, using technology for K-12 graph instruction can simplify data collection (Ates & Stevens, 2003), help students visualize graph relationships (Roschelle et al., 2010), and allow students to move quickly from experimentation to data interpretation (Adams & Shrum, 1990).

An analysis of investigative features of graphing technologies illustrate important gaps in the existing literature. Research on instruction using graphs indicates that a broad challenge in teaching graphing is to be mindful of and tackle the perception of a graph as an end-product of the scientific process rather than as a means to inform the entire scientific process (Nicoloau et al., 2007; Rodrigues, 1994). Our review specifically illustrates this critique for graphing technologies by revealing gaps in the use of graphing technologies in both the initial stages of the scientific process (planning an experiment, selecting resources, etc.) and in the latter stages of the scientific process (reflecting on graph data, presenting on graph data, arguing on graph data, etc). In light of the NGSS SEPs, our review of investigative features reveals a need for future K-12 graphing technology research to focus more on supporting students to plan and carry out investigations (SEP3), to engage in argument from evidence (SEP7), and to obtain, evaluate, and communicate information (SEP8).

4.1. Limitations

We limited this investigation to peer reviewed articles to ensure the consistency and quality of the studies. Including conference papers and unpublished research reports, could modify the findings. Even with peer-reviewed reports, however, we were able to show that many non-significant studies would be needed to change the main conclusions. We used an article search approach to locate articles rather than searching large databases. A database search may have yielded different articles. Even with an article search, however, we had to eliminate many studies (92%; 500/542 articles). A database search is problematic given the frequent use of “graph” as a word or component of a word in most journal articles.

5. Conclusion

Graph proficiency is a critical requirement for all students and all citizens in the 21st century, both within mathematics and science education, and beyond. The NGSS (NGSS Lead States, 2013) and CCMS (Common Core State Standards Initiative, 2010) show how proficiency with graphs in science and mathematics can support sustained autonomous work. This review illustrates the value of technological supports to support graph proficiency. It reveals the need for continued research on the use of technology to support students' understanding of graphs, both as a process and product of the scientific endeavor.

References

*Adams, D., & Shrum, J.W. (1990). The effects of microcomputer-based laboratory exercises on the acquisition of line graph construction and interpretation skills by high school biology students. *Journal of Research in Science Teaching*, *27*(8), 777–787. doi:10.1002/tea.3660270807

Ainley, J., Nardi, E. & Pratt, D. (2000). The construction of meanings for trend in active graphing. *International Journal of Computers for Mathematical Learning*, *5*(2), 85–114. doi:10.1023/A:1009854103737

Ainsworth, S. (1999). The functions of multiple representations. *Computers and Education*, *33*(2-3), 131–152. doi:10.1016/S0360-1315(99)00029-9

*Applebaum, L. R., Vitale, J. M., Gerard, E., & Linn, M. C. (2017). Comparing design constraints to support learning in technology-guided inquiry projects. *Educational Technology & Society*, *20*(4), 179–190. https://www.jstor.org/stable/26229216

*Ardac, D., & Sezen, A. H. (2002). Effectiveness of computer-based chemistry instruction in enhancing the learning of content and variable control under guided versus unguided conditions. *Journal of Science Education and Technology*, *11*(1), 39–48. doi:10.1023/A:1013995314094

Arsenault, D. J., Smith, L. D., & Beauchamp, E. A. (2006). Visual inscriptions in the scientific hierarchy: Mapping the "Treasures of Science". *Science Communication*, *27*(3), 376–428. doi:10.1177/1075547005285030

*Ates, S., & Stevens, J. T. (2003). Teaching line graphs to tenth grade students having different cognitive developmental levels by using two different instructional modules. *Research in Science & Technological Education*, *21*(1), 55–66. doi:10.1080/02635140308339

Attalim, Y., & Goldschmidt, C. (1996). The effects of component variables on performance in graph comprehension tests. *Journal of Educational Measurement*, *33*(1), 93–105. doi:10.1111/j.1745-3984.1996.tb00481.x

Authors (2014)

*Barab, S. A., Sadler, T. D., Heiselt, C., Hickey, D., & Zuiker, S. (2007). Relating narrative, inquiry, and inscriptions: Supporting consequential play. *Journal of Science Education and Technology*, *16*(1), 59–82. doi:10.1007/s10956-006-9033-3

*Basu, S., Sengupta, P., & Biswas, G. (2015). A scaffolding framework to support learning of emergent phenomena using multi-agent-based simulation environments. *Research in Science Education*, *45*(2), 293–324. doi:10.1007/s11165-014-9424-z

*Beichner, R. J. (1990). The effect of simultaneous motion presentation and graph generation in a kinematics lab. *Journal of Research in Science Teaching, 27*(8), 803–815. doi:10.1002/tea.3660270809

Boote, S. K., & Boote, D. N. (2017). Leaping from discrete to continuous independent variables: Sixth graders' science line graph interpretations. *The Elementary School Journal*, *117*(3), 455–484. doi:10.1086/690204

Berg, C.A., & Phillips, D.G. (1994). An investigation of the relationship between logical thinking structures and the ability to construct and interpret line graphs. *Journal of Research in Science Teaching*, *31*(4), 323–344. doi:10.1002/tea.3660310404

Bertin, J. (1973). *Semiologie Graphique* (2nd ed.). The Hague: Mouton-Gautier. English translation by William Berg & Howard Wainer (1983) and published as *Semiology of graphics,* Madison, WI: University of Wisconsin Press.

Borenstein, M., Hedges, L. V, Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111. doi:10.1002/jrsm.12

*Cavanaugh, C., Gillan, K. J., Bosnick, J., Hess, M., & Scott, H. (2008). Effectiveness of interactive online algebra learning tools. *Journal of Educational Computing Research*, *38*(1), 67–95. doi:10.2190/EC.38.1.d

*Chao, J., Chiu, J. L., DeJaegher, C. J., & Pan, E. A. (2016). Sensor-augmented virtual labs: Using physical interactions with science simulations to promote understanding of gas behavior. *Journal of Science Education and Technology*, *25*(1), 16–33. doi:10.1007/s10956-015-9574-4

Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, *45*(5), 283–292. doi:10.3102/0013189X16656615

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, *9*, 88–113. doi:10.1016/j.edurev.2013.01.001

*Chiu, J., & Linn, M. (2014). Supporting knowledge integration in chemistry with a visualization-enhanced inquiry unit. *Journal of Science Education and Technology*, *23*(1), 37–58. doi:10.1007/s10956-013-9449-5

Common Core State Standards Initiative. (2010). Common Core State Standards for mathematics. Retrieved from www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, *18*(5), 382–393. http://doi.org/10.2307/749086

Delmas, R., Garfield, J., & Ooms, A. (2005). Using assessment items to study students' difficulty reading and interpreting graphical representations of distributions. In K. Makar (Ed.), *Proceedings of the Fourth International Research Forum on Statistical Reasoning, Literacy, and Reasoning*. Auckland, New Zealand: University of Auckland.

*Deng, F., Chai, C., Chen, W., & Qian, Y. (2011). Constructivist-oriented data-logging activities in Chinese chemistry classroom: Enhancing students' conceptual understanding and their metacognition. *Asia-Pacific Education Researcher, 20*(2), 207–221.

*Deniz, H., & Dulger, M. F. (2012). Supporting fourth graders' ability to interpret graphs through real-time graphing technology: A preliminary study. *Journal of Science Education and Technology*, *21*(6), 652–660. doi:10.1007/s10956-011-9354-8

*Dickes, A. C., & Sengupta, P. (2013). Learning natural selection in 4th grade with multi-agent-based computational models. *Research in Science Education*, *43*(3), 921–953. doi:10.1007/s11165-012-9293-2

diSessa, A. A. (2004). Metarepresentation: Native competence and targets for instruction. *Cognition and Instruction*, *22*(3), 293–331. doi:10.1207/s1532690xci2203_2

*Dorji, U., Panjaburee, P., & Srisawasdi, N. (2015). A learning cycle approach to developing educational computer game for improving students' learning and awareness in electric energy consumption and conservation. *Educational Technology & Society*, *18*(1), 91–105. https://www.jstor.org/stable/jeductechsoci.18.1.91

Ellington, A. J. (2006). The effects of non- CAS graphing calculators on student achievement and attitude levels in mathematics: A meta- analysis. *School Science and Mathematics*, *106*(1), 16–26. doi:10.1111/j.1949-8594.2006.tb18067.x

*Erbas, A. K., Ince, M., & Kaya, S. (2015). Learning mathematics with interactive whiteboards and computer-based graphing utility. *Journal of Educational Technology & Society*, *18*(2), 299–312. www.jstor.org/stable/jeductechsoci.18.2.299

*Freidler, Y. & McFarlane, A. (1997). Data logging with portable computers: A study of the impact on graphing skills in secondary pupils. *Journal of Computers in Mathematics and Science Teaching*, *16*(4), 527–550. https://eric.ed.gov/?id=EJ557478

Friel, S.N., Curcio, F.R., & Bright, G.W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, *32*(2), 124–158. www.jstor.org/stable/749671

Friel, S., & Bright, G. (1996). *Building a theory of graphicacy: How do students read graphs?* Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Glazer, N. (2011). Challenges with graph interpretation: a review of the literature. *Studies in Science Education*, *47*(2), 183–210. doi:10.1080/03057267.2011.605307

Greeno, J.G. & Hall, R.P. (1997). Practicing representation: Learning with and about representational forms. *Phi Delta Kappan*, *78*(5), 361–367.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

*Hsu, Y.-S., Fang, S.-C., Zhang, W.-X., Wu, H.-K., Wu, P.-H., & Hwang, F.-K. (2016). Identifying effective design features of technology-infused inquiry learning modules: A two-year study of students' inquiry abilities. *Educational Technology & Society*, 19 (2), 228–244. https://www.jstor.org/stable/jeductechsoci.19.2.228

*Huntley, M. A., Rasmussen, C. L., Villarubi, R. S., Sangtong, J., & Fey, J. T. (2000). Effects of standards-based mathematics education: A study of the Core-plus Mathematics Project Algebra and Functions Strand. *Journal for Research in Mathematics Education*, *31*(3), 328–361. doi:10.2307/749810

*Kamarainen, A. M., Metcalf, S., Grotzer, T., Browne, A., Mazzuca, D., Tutwiler, M. S., & Dede, C. (2013). EcoMOBILE: Integrating augmented reality and probeware with environmental education field trips. *Computers & Education*, *68*, 545–556. doi:10.1016/j.compedu.2013.02.018

Kastberg, S., & Leatham, K. (2005). Research on graphing calculators at the secondary level: Implications for mathematics teacher education. *Contemporary Issues in Technology and Teacher Education*, *5*(1), 25–37. www.citejournal.org/vol5/iss1/mathematics/article1.cfm

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, *86*(4), 945–980. doi:10.3102/0034654315626800

*Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, 8, 30–43. Doi:10.1.1.330.2572

*Koklu, O., & Topcu, A. (2012). Effect of Cabri-assisted instruction on secondary school students' misconceptions about graphs of quadratic functions. *International Journal of Mathematical Education in Science and Technology*, *43*(8), 999–1011. doi:10.1080/0020739X.2012.678892

*Kramarski, B. (1999). The study of graphs by computers: Is easier better? *Educational Media International*, *36*(3), 203–209. doi:10.1080/0952398990360306

Krohn, R. (1991). Why are graphs so central in science. *Biology and Philosophy*, *6*(2), 181–203. doi:10.1007/BF02426837

*Kwon, O. N. (2002). The effect of calculator- based ranger activities on students' graphing ability. *School Science and Mathematics*, *102*(2), 57–67. doi:10.1111/j.1949-8594.2002.tb17895.x

LaDue, N. D., Libarkin, J. C., & Thomas, S. R. (2015). Visual representations on high school biology, chemistry, earth science, and physics assessments. *Journal of Science Education and Technology*, *24*(6), 818–834. doi:10.1007/s10956-015-9566-4

Lai, K., Cabrera, J., Vitale, J. M., Madhok, J., Tinker, R., & Linn, M. C. (2016). Measuring graph comprehension, critique, and construction in science. *Journal of Science Education and Technology*, *25*(4), 665–681. doi:10.1007/s10956-016-9621-9

Latour, B. (1990). Drawing things together. In M. Lynch & S. Woolgar (Eds.), *Representation in scientific practice.* Cambridge, MA: MIT Press, pp. 19–68.

Lee, H. -S., Liu, O. L., & Linn, M. C. (2011). Validating measurement of knowledge integration science using multiple-choice and explanation items. *Applied Measurement in Education*, *24*(2), 115–136. doi:10.1080/08957347.2011.554604

*Lee, V. R., & Thomas, J. M. (2011). Integrating physical activity data technologies into elementary school classrooms. *Educational Technology Research and Development*, *59*(6), 865–884. doi:10.1007/s11423-011-9210-9

Leinhardt, G., Zaslavsky, O., & Stein, M.K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research*, *60*(1), 1–64. doi:10.3102/00346543060001001

*Linn, M. C., Layman, J. W. and Nachmias, R. (1987). Cognitive consequences of microcomputer-based laboratories: graphing skills development. *Contemporary Educational Psychology*, *12*, 244-253.

Lipsey, M.W., Puzio, K., Yun, C., Hebert, M.A., Steinka-Fry, K., Cole, M.W., Roberts, M., Anthony, K.S., & Busick, M.D. (2012). Translating the statistical representation of the effects of education interventions into more readily interpretable forms. National Center for Special Education Research (NSCER 2013-3000).

Liu, O. L., Lee, H.S. & Linn, M.C. (2011). A comparison among multiple-choice, constructed-response and explanation multiple-choice items. *Educational Assessment*, 16, 164–184.

*Malone, K. L., Schunn, C. D., & Schuchardt, A. M. (2018). Improving conceptual understanding and representation skills through excel-based modeling. *Journal of Science Education and Technology*, *27*(1), 30–44. doi:10.1007/s10956-017-9706-0

*McElhaney, K. W., & Linn, M. C. (2011). Investigations of a complex, realistic task: Intentional, unsystematic, and exhaustive experimenters. *Journal of Research in Science Teaching*, *48*(7), 745–770. doi:10.1002/tea.20423

Miller, D., & Linn, M.C. (2013). How does traditional science education assess visual and spatial thinking? Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco. CA.

*Mokros, J. R., & Tinker, R. F. (1987) The impact of microcomputer-based labs on children's ability to interpret graphs. *Journal of Research in Science Teaching*, *24*(4), 369–383. doi:10.1002/tea.3660240408

Nakhleh, M. B. (1994). A review of microcomputer-based labs: How have they affected science learning? *Journal of Computers in Mathematics and Science Teaching*, *13*(4), 368–381. https://eric.ed.gov/?id=EJ288758

National Research Council. (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board of Science Education, Division of Behavioral and Social Sciences and Education. Washington, D.C.: The National Academies Press

NGSS Lead States (2013). Next generation science standards: For states, by states. Washington, DC: The National Academy Press.

*Nicolaou, C. T., Nicolaidou, I. A., Zacharia, Z. C., & Constantinou, C. P. (2007). Enhancing fourth graders' ability to interpret graphical representations through the use of microcomputer-based labs implemented within an inquiry-based activity sequence. *Journal of Computers in Mathematics & Science Teaching*, *26*(1), 75–99. https://eric.ed.gov/?id=EJ752321

Organization for Economic Cooperation and Development. (2006). *PISA 2006 Science Competencies for Tomorrow's World*. Accessed May 7th, 2018, from www.oecd.org/fr/education/scolaire/programmeinternationalpourlesuividesacquisdeselevespisa/pisa2006results.htm

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*(2), 157–159. doi:10.2307/1164923

*Park, J. (2015). Effect of robotics enhanced inquiry based learning in elementary science education in South Korea. *Journal of Computers in Mathematics and Science Teaching*, *34*(1), 71-95. https://www.learntechlib.org/primary/p/130555/

Pellegrino, J. W., Wilson, M. R., Koenig, J. A., & Beatty, A. S. (2014). Developing Assessments for the Next Generation Science Standards. National Academies Press. 500 Fifth Street NW, Washington, DC 20001.

*Peterman, K., Cranston, K. A., Pryor, M., & Kermish-Allen, R. (2015). Measuring primary students' graph interpretation skills via a performance assessment: A case study in instrument development. *International Journal of Science Education*, *37*(17), 2787–2808. doi:10.1080/09500693.2015.1105399

*Purba, S. W. D., & Hwang, W.-Y. (2017). Investigation of learning behaviors and achievement of vocational high school students using an ubiquitous physics tablet PC app. *Journal of Science Education and Technology*, *26*(3), 322–331. doi:10.1007/s10956-016-9681-x

Rakes, C. R., Valentine, J. C., McGatha, M. B., & Ronau, R. N. (2010). Methods of instructional improvement in algebra: A systematic review and meta-analysis. *Review of Educational Research*, *80*(3), 372–400. doi:10.3102/0034654310374880

Rodrigues, S. (1994). Data handling in the primary classroom: Children's perception of the purpose of graphs. Research in Science Education, 24(1), 280–286. https://doi.org/10.1007/BF02356354

*Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S., … Gallagher, L. P. (2010). Integration of technology, curriculum, and professional development for advancing middle school mathematics: Three large-scale studies. *American Educational Research Journal*, *47*(4), 833-878. doi:10.3102/0002831210367426

*Ryoo, K., Bedell, K., & Swearingen, A. (2018). Promoting linguistically diverse students' short-term and long-term understanding of chemical phenomena using visualizations. *Journal of Science Education and Technology*, n/a, n/a. doi:10.1007/s10956-018-9739-z

Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Education Psychology Review, 14*(1), 47–69. doi:10.1023/A:1013180410169

*Songer, N. B., & Linn, M. C. (1991). How do students' views of science influence knowledge integration? *Journal of Research in Science Teaching*, *28*(9), 761–784. doi:10.1002/tea.3660280905

*Tan, C.-K. (2012). Effects of the application of graphing calculator on students' probability achievement. *Computers & Education*, *58*(4), 1117–1126. doi:10.1016/j.compedu.2011.11.023

*Tan, S.C., Yeo, A.C.J. & Lim, W.Y. (2005). Changing epistemology of science learning through inquiry with computer-supported collaborative learning. *Journal of Computers in Mathematics and Science Teaching*. 24(4), 367–386. https://eric.ed.gov/?id=EJ724784

Tortosa, M. (2012). The use of microcomputer based laboratories in chemistry secondary education: Present state of the art and ideas for research-based practice. *Chemistry Education Research and Practice*, *13*(3), 161–171. doi:10.1039/C2RP00019A

*Varma, K., & Linn, M. C. (2012). Using interactive technology to support students' understanding of the greenhouse effect and global warming. *Journal of Science Education and Technology*, *21*(4), 453–464. doi:10.1007/s10956-011-9337-9

*Vitale, J. M., Lai, K., & Linn, M. C. (2015). Taking advantage of automated assessment of student-constructed graphs in science. *Journal of Research in Science Teaching*, *52*(10), 1426–1450. doi:10.1002/tea.21241

Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher*, *21*(1), 14–23. doi:10.3102/0013189X021001014

Wang, Z. H., Wei, S., Ding, W., Chen, X., Wang, X., & Hu, K. (2012). Students' cognitive reasoning of graphs: Characteristics and progression. *International Journal of Science Education*, *34*(13), 2015–2041. doi:10.1080/09500693.2012.709333

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching*, *46*(6), 716–730. doi:10.1002/tea.20318

*Yang, K.-Y., & Heh, J.-S. (2007). The impact of internet virtual physics laboratory instruction on the achievement in physics, science process skills and computer attitudes of 10thgrade students. *Journal of Science Education and Technology*, *16*(5), 451–461. doi:10.1007/s10956-007-9062-6

Yeh, Y., & McTigue, E. (2009). The frequency, variation, and function of graphical representations within standardized state science tests. *School Science and Mathematics*, *109*(8), 435–449. doi:10.1111/j.1949-8594.2009.tb18291.x

*Zehavi, N. (1988). Evaluation of the effectiveness of mathematics software in shaping students' intuitions. *Journal of Educational Computing Research*, *4*(4), 391–401. doi:10.2190/HGKE-GTRE-Q92T-7N20

*Zengin, Y., & Tatar, E. (2017). Integrating dynamic mathematics software into cooperative learning environments in mathematics. *Educational Technology & Society*, 20 (2), 74–88. https://eric.ed.gov/?id=EJ1137656

Zucker, A., Staudt, C., & Tinker, R. (2015). Teaching graph literacy across the curriculum. *Science Scope*, *38*(6), 19–24.

Table 1 - Graph Comprehension Categories

Author(s)	**Individual Graph Features**	**Trends Within or Across Graphs**	**Relationships between Graphs and Disciplinary Ideas**
Ainley, Nardi, & Pratt, 2000	Feature-spotting	Shape-spotting	Normalizing
Attalim & Goldschmidt (1996)	Interpretation of points on the graph	Interpretation of trends, shapes, intervals	Quantitative and qualitative interpretation
Bertin (1973)	Data extraction - determine components represented and their connection to variables	Interpretation of trends	Comparing trends and recognizing groupings
Curcio (1987)	Reading with the data	Reading between the data	Reading beyond the data
Friel et al. (2001)	Reading information	Manipulating information	Generalize, predict, and identify
Shah & Hoeffner (2002)	Encoding and identifying visual features	Relating visual features (points) to conceptual relations (function: linear, exponential, etc.)	Determine the concept variables being quantified and associate those variables to the encoded functions
Wainer (1992)	Basic data points	Trends	Integration
Wang et al. (2012)	Explicit information	Tacit information (deduction)	Conclusive information (Summary of analyses of graphs)
Zucker, Staudt, & Tinker (2015)	Identify and encode prominent visual graph features	Link visual graph features to quantitative facts, trends, or other relationships	Integrate the features and relationships with the context of the graph

Table 2 - Graph Technology Design Studies (Shortened Table)
*See Appendix B for Extended Table

Reference	Nature of Study	Topic	Technology	No. of Students	Grade	Effect size(s) - Cohen's d and Hedges' g (Bold - significant/Not Significant (NS)
1. Applebaum et al., 2017	Use of a web-supported curriculum involving student designs of self-propelled vehicles	Kinematics	Simulations	228	8th (Mid.)	d = 0.35/**g = 0.33** (p <.001) d = 0.41/**g = 0.41** (p <.001)
2. Barab et al., 2007	Use of a multi-user virtual environment, Quest Atlantis, to engage students in an inquiry investigation	Water quality	Multi-User Virtual Environment	28	4th (Ele.)	d = 1.55/ **g = 1.48** (p <.001)
3. Basu et al., 2015	Use of simulations to learn about a desert (Hawks/Doves) ecosystem	Ecosystem	Simulations	20	8th (Mid.)	d = 6.44/**g = 5.86** (p <.001)
4. Chao et al., 2016	Use of an augmented virtual lab to learn the gas laws	Gas Laws	Augmented virtual lab.	16	10th & 11th (High)	d = 1.50/**g = 1.45** (p <.001)
5. Dickes & Sengupta, 2013	Use of simulations to learn about a bird-butterfly ecosystem	Ecosystem	Simulations	10	4th (Ele.)	d = 1.97/**g = 1.68** (p <.001)
6. Kamarainen et al., 2013	Use of probeware alongside an augmented reality experience	Water quality	Probeware for field trips	71	6th (Mid.)	d = 0.86/**g = 0.85** (p <.001)
7.Kramarski, 1999	Use of a computer graphics plotter to illustrate student difficulties in graph	Kinematics	Computer Graphics Plotter	82	8th (Mid.)	d = 0.03/g = 0.02 (NS)

	construction for everyday situations					
8. Kwon, 2002	Use of Calculator-Based Ranger (CBR) activities in improving graphing abilities	Kinematics	Calculator Based Ranger (CBR) activities	428	7th & 8th (Mid.)	d = 0.82/**g = 0.81** (p <.001) d = 0.94/**g = 0.94** (p <.001) d = 0.16/**g = 0.16** (p <.014) d = 0.86/**g = 0.86** (p <.001)
9. Linn, Layman, & Nachmias, 1987	Use of probeware to learn about the functional relationships of physical phenomena	Multiple Topics	MBL graph templates with probeware	117	8th (Mid.)	d = 0.89/**g = 0.88** (p <.001) d = 1.11/**g = 1.10** (p <.001) d = 0.49/**g = 0.48** (p <.001)
10. McElhaney & Linn, 2011	Use of simulations with an airbag context to learn kinematics concepts	Kinematics	Simulations	148	11th & 12th (High)	d = 0.59/**g = 0.58** (p <.001)
11. Mokros & Tinker, 1987	Use of MBL to support student understanding of graphs	Kinematics	Use of MBL	125	7th & 8th (Mid.)	d = 0.86/**g = 0.85** (p <.001)
12. Peterman et al., 2015	Use of an online community of students, teachers, fishermen & scientists to support graph interpretation	Thermo-dynamics	Weather-Blur	55	1st- 4th (Ele.)	d = 0.59/**g = 0.58** (p <.001) d = 0.11/g = 0.11 (p

	skills					<0.37) d = 0.78/**g = 0.77** (p <.001)
13. Purba & Hwang, 2017	Use of a Physics Tablet PC App to learn about pendulum motion	Kinematics	Tablet App	36	9th - 12th (High)	d = 0.58/**g = 0.55** (p <.001)
14. Ryoo et al., 2018	Use of chemical visualizations to support student understanding of heat transfer and chemical reactions	Thermo-dyamics	Simulations	152	8th (Mid.)	d = 1.54/**g = 1.52** (p <.001) d = 1.50/**g = 1.47** (p <.001)
15. Songer & Linn, 1991	Use of simulated and real time experiments to make graph predictions	Thermo-dynamics	Simulated & real time experiments	153	8th (Mid.)	d = 0.36/**g = 0.35** (p <.001)
16. Tan, Yeo, & Lim, 2005	Use of an online discussion tool to support student understanding of graphs	Multiple Topics	Knowledge Forum (Online Forum)	13	7th to 10th (Mid./ High)	d = 0.28/g = 0.25 (NS)
17. Varma & Linn, 2012	Use of virtual experiments to understand the greenhouse effect	Climate Change	Simulations	190	6th (Mid.)	d = 1.01/**g = 1.12** (p <.001)
18. Vitale et al., 2015	Use of automated assessments and feedback to support student understanding of graphs	Kinematics	Automated scoring within an inquiry learning platform involving simulations	397	8th- 12th (Mid./ High)	d = 0.62/**g = 0.62** (p <.001) d = 0.67/**g = 0.67** (p <.001) d = 0.28/**g = 0.28** (p <.001)

						d = 0.41/**g = 0.41** (p <.001)
19. Zengin & Tatar, 2017	Use of mathematics software alongside a cooperative learning model to support understanding of functions	Function	Mathematics software GeoGebra	24	10th/ 11th (High)	d = 0.28/**g = 0.27** (p <.05)
General Features of Pre/Post Design Studies (Topic, Tech., No. of Students, Grade, and Effect Size (31 effect sizes)		**Kinematics: 7; Thermo-dynamics: 3; Others: 9**	**Simulation: 7; Probeware: 5; Others: 7**	**2293**	**Elem: 3; Mid: 10; High: 4; Mix: 2**	**g = 0.59 (Fixed: 0.55 - 0.62; Random: 0.57 - 0.82)**

Table 3 - Graph Technology Comparison Studies (Shortened Table)
*See Appendix B for Extended Table

Reference (Study Design Type)	**Nature of the Study (Comparison Type)**	**Topic**	**No. of Stu- dents**	**Grade**	**Condition Differences (n - Sample Size, M - Mean, SD - Standard Deviation)**	**Effect size(s) - Cohen's d and Hedges' g (Bold - significant/NS - not significant)**
1. Adams & Shrum, 1990 (Post-test)	MBL (computer used to collect, display, store, and print the data from exercises) vs. Conventional (used traditional equipment (i.e., stopwatches, thermometers, data tables, pencils, and paper) to teach graph construction and graph interpretation.	Thermo- dynamics	20	10th (High)	MBL (n = 10), M = 14.1, SD = 4.1; Conven. (n = 10), M = 14.6, SD = 5.08 MBL (n = 10), M = 17.2, SD = 3.83; Conventional (n = 10), M = 17.6, SD = 3.30 MBL (n = 10), M = 6.2, SD = 1.96 Conventional (n = 10), M = 7.6, SD = 1.39 MBL (n = 10), M = 11, SD = 2.3 Conventional (n = 10), M = 10, SD = 2.09 MBL (n = 10), M = 4.9, SD = 3.51 Conventional (n = 10), M = 4.9, SD = 3.76	TOGS d = -0.11/g = -0.10 (NS) I-TOGS d = -0.11/g = -0.10 (NS) I-TOGS Construction d = -0.86/g = -0.78 (NS) I-TOGS Interpretation d = 0.48/g = 0.43 (NS) GALT d = 0.00/g = 0.00 (NS)
2. Ardac & Sezen, 2002 (Pre/Post-test)	Guided Technology (GT; computer software) Use vs. Conventional Textbook (T) Approach Unguided Technology	Thermo- dynamics	63	9th (High)	Guided Tech. (n = 18) Pre: M = 3.38, SD = 1.85/Post: M = 6.55, SD = 2.30 Unguided Tech. (n = 22) Pre: M = 3.86, SD = 1.08/Post: M = 3.86,	*Content Knowledge* GT vs. T d = 0.14/g = 0.13 (NS) UT vs. T

	(UT) Use vs. Conventional Textbook Approach				SD = 1.85 Textbook (n = 21) Pre: M = 2.33, SD = 1.42/Post: M = 5.23, SD = 1.67 *Process Skills* Guided Tech. (n = 23) Pre: M = 3.83, SD = 3.21/Post: M = 6.52, SD = 3.01 Unguided Tech. (n = 20) Pre: M = 4.25, SD = 2.65/Post: M = 5.90, SD = 2.69 Textbook (n = 20) Pre: M = 5.05, SD = 2.87/Post: M = 6.10, SD = 3.57	d = -2.30/**g = -1.61** (p <.001) *Process* GT vs. T d = 0.49/**g = 0.49** (p <.001) UT vs. T d = 0.19/g = 0.18 (NS)
3. Ates & Stevens, 2003 (Pre/Post-test and Post-test)	Digital tech. (Universal Lab Interface, Logger Pro software, sensors) vs. Conven. (Line graphing unit without technology) Teaching line graphs	Kine-matics	43	10th (High)	*Pre/Posttest (I-TOGS)* Technology (n = 22): Pre: M = 12.2, SD = 3.2/Post: M = 16.46, SD = 3.64 Conventional (n = 21) Pre: M = 12.0, SD = 2.7/Post: M = 16.00, SD = 3.65 *Posttest (PAT)* Technology(n = 22): M = 11.27, SD = 3.10 Conventional (n = 21): M = 11.24, SD = 3.49	d = 0.07/g = 0.07 (NS) d = 0.00/g = 0.00 (NS)
4. Beichner, 1990 (Pre/Post-test)	VideoGraph With or Without Motion View vs. Conventional (With or	Kine-matics	222	12th (High)	Videograph *(n = 58)*: Pre M = 12.3, SD = 3.4/Post M = 12.4, SD = 4.0	*With Motion View* d = -0.16/g = -0.16 (NS)

	Without Motion View (Made use of photographs))				Conventional (*n = 51):* Pre M = 11.5, SD = 3.7/Post M = 12.3, SD = 4.3 Videograph (n = 55): Pre M = 12.5, SD = 3.5/Post M = 13.5, SD = 4.0 Conventional (n = 58): Pre M = 12.2, SD = 4.4/Post M = 13.4, SD = 4.4	*No Motion View* d = -0.04/g = -0.04 (NS)
5. Cavanaugh et al., 2008 (Pre/Post-test)	Online algebra graphing tool vs. conventional textbook approach	Function	47	6th - 12th (Mid./ High)	Graphing tool (n = 33): Pre M = 15.02, SD = 15.02/Post M = 18.08, SD = 5.69 Conventional (n = 14): Pre M = 17.5, SD = 6.43/Post M = 19.21, SD = 7.45	d = 0.21/g = 0.21 (NS)
6. Chiu & Linn, 2014 (Pre/Post-test)	Crime Scene Investigators (CSI) online unit vs. conventional textbook approach	Chemical Reactions	49	10th/ 11th (High)	CSI (n = 24): Pre M = 2.59, SD = 0.85/Post M = 3.03, SD = 0.82 Conventional (n = 25): Pre M = 2.02, SD = 0.70/Post M = 2.17, SD = 0.67	d = 0.38/**g = 0.38** (p <.05)
7. Deng, Chen, Chai, & Qian, 2011 (Post-test)	Data-logging Based Learning Environment (DBLE) vs. Conventional (Lecture/Textbook problems)	Multiple Topics - Science	96	11th (High)	DBLE (n = 51) M = 11.57, SD = 2.85 Conventional (n = 45) M = 6.82, SD = 2.47	d = 1.79/**g = 1.75** (p <.001)
8. Deniz & Dulger, 2012 (Pre/Post-test)	MBL (probes and graphing software) vs. Conventional (tapes, meter sticks, and	Multiple Topics - Science	39	4th (Ele.)	MBL (n = 19) Pre M = 2.39, SD = 0.78/Post M = 2.89, SD = 1.13	d = 1.12/g = 1.12 (NS)

	thermometers)				Conventional (n = 20): Pre M = 2.85, SD = 1.04/Post M = 2.00, SD = 1.21	
9. Dorji et al., 2015 (Pre/Post-test)	Computer game vs. textbook approach	Electric Energy Use	111	10th (High)	Comp. Game (n = 69): Post M = 11.97, SD = 3.84 Conventional (n = 52): Post M = 10.08, SD = 3.33	d = 0.52/**g = 0.51** (p <.05)
10. Erbas et al., 2015 (Pre/Post-test)	Interactive whiteboard and NuCalc graphing software vs. Con. (normal instruction, no computers)	Function	65	12th (High)	Technology (n = 31): Pre M = 10.68, SD = 9.24; Post M = 45.58, SD = 5.75 Conventional (n = 34): Pre M = 10, SD = 8.59; Post M = 35.21, SD = 8.64	d = 1.29/**g = 1.29** (p <.001)
11. Friedler & McFarlene, 1997 (Pre/Post-test)	MBL (Datalogging) vs. Conventional with 9th and 11th graders (Conventional; normal instruction, no data loggers)	Thermo-dynamics	178	9th & 11th (High)	MBL (n = 40): Pre M = 45.2, SD = 18.0; Post M = 59.2, SD = 21.0 Conventional (n = 46) Pre M = 41.8, SD = 23.2; Post M = 50.6, SD = 20.9 MBL (n = 46): Pre M = 60.1, SD = 20.7; Post M = 69.2, SD = 18.2 Conventional (n = 46) Pre M = 61.6, SD = 20.7; Post M = 73.1, SD = 14.2	d = 0.24/**g = 0.24** (p = .03) d = -0.14/g = -0.14 (p = .12)
12. Hsu, Fang, Zhang, Wu, Wu, & Hwang, 2016 (Pre/Post-test)	Technology-based Inquiry Units with Sensors vs. Conventional Textbook Approach	Multiple Topics - Science	51	7th (Mid.)	A2: Analyzing Data Tech. Inq. Unit (n = 24): Pre M = 1.38, SD = 0.77/Post M = 1.75, SD = 0.53 Conventional (n = 27) Pre M = 1.26, SD = 0.81/Post M = 1.48,	d = 0.18/g = 0.20 (NS)

					SD = 0.85	
13. Huntley, Rasmussen, Villarubi, Sangtong, & Fey, 2000 (Post-test)	Graphics calculator vs textbook approach Experimental group used graphics calculators for Test 1 (T1) and Test 2 (T2), but not for Test 3 (T3)	Function	593	8th/ 9th (Mid/ High)	Calculator (n = 320): M = 42.6, SD = 21.3 Conven. (n = 273): M = 34.1, SD = 14.8 Calculator (n = 184): M = 1.43, SD = 1.35 Conven. (n = 191): M = 1.07, SD = 1.2 Calculator (n = 312): M = 29.0, SD = 18.4 Conven. (n = 265): M = 38.4, SD = 16.2	T1: d = 0.45/**g = 0.45** (p <.001) T2: d = 0.28/**g = 0.28** (p <.01) T3: d = -0.54/**g = -0.53** (p <.001)
14.Koedinger, Anderson, Hadley, & Mark, 1997 (Post-test)	PUMP Algebra Curriculum + Intelligent Tutoring vs. Conventional Textbook Approach	Function	168	9th (High)	PUMP Algebra Curriculum + Intell. Tutor (n = 124) M = 0.37, SD = 0.32 Conven. (n = 44) M = 0.15, SD = 0.18	d = 0.76/**g = 0.75** (p <.01)
15. Koklu & Topcu, 2012 (Post-test)	Cabri-assisted instruction vs. Conventional (normal instruction, no software)	Function	44	10th (High)	*Achievement Scores (Posttest)* Experimental (n = 20):M = 64.57, SD = 14.24 Conventional (n = 24): M = 52.41, SD = 21.78	d = 0.52/**g = 0.50** (p <.05)
16. Lee & Thomas, 2011 (Pre/Post-test)	Physical Activity Data Sensors vs. Conventional (Normal instruction, pencil & paper)	Kine-matics	46	5th (Ele.)	Experimental (n = 25): Pre M = 8.11, SD = 2.52/Post M = 14.14, SD = 4.84 Conventional (n = 21) Pre M = 6.78, SD = 3.35/Post M = 14.00, SD = 4.11 Experimental (n = 4) Pre M = 0.875, SD = 0.64/Post M = 2.625, SD = 1.19	*Written Assessment* d = 0.28/g = 0.28 (NS) *Struct. Interview* d = 1.20/**g = 1.19** (p

					Conventional (n = 4) Pre M = 1.25, SD = 0.75/Post M = 1.81, SD = 0.60	<.05)
17. Malone, Schunn, & Schuchardt, 2018 (Post-test)	Excel-based Modelling vs. Conventional Textbook Approach	Eco-system	424	9th-11th (High)	Modelling (n = 255) M = 54, SD = 18 Conven. (n = 169) M = 48, SD = 19	d = 0.32/**g = 0.32**
18. Nicolaou, Nicolaidou, Zacharia, & Constantinou, 2007 (Pre/Post-test)	MBL+Inquiry vs. Inquiry vs. Conven. (traditional laboratory investigation) for phase changes	Thermo-dynamics	65	4th (Ele.)	*Construct* (F(2, 65) = 13.99, p < .001) MBL + Inq. (n = 23): Pre M = .41, SD = .283; Post M = .74, SD = .255 Inquiry (n = 22): Pre M = .34, SD = .283; Post M = .45, SD = .283 Con. (n = 20): Pre M = .38, SD = .197; Post M = .48, SD = .197 *Interpret* (F(2, 65) = 17.659, p < .001) MBL + Inq.: Pre M = .30, SD = .170; Post M = .79, SD = .170 Inquiry: Pre M = .36, SD = .264/Post M = .52, SD = .264 Con.: Pre M = .38, SD = .201/Post M = .43, SD = .201	(MBL+ Inq. Vs Con.) d = 0.92 **g = 0.91** (p <.001) (MBL+ Inq. Vs Con.) d = 2.43 **g = 2.33** (p <.001)
19. Park, 2015 (Pre/Post-test)	Robotics-Enhanced Inquiry-Based Learning vs. Conv. Textbook Approach	Multiple Topics - Science	123	4th/ 5th (Ele.)	Robotics (n = 63): Pre M = 74.55, SD = 9.04/Post M = 84.82, SD = 7.23 Conventional (n = 60):	d = 0.32/**g = 0.40** (p =.005)

					Pre M = 75.00, SD = 8.79/Post M = 82.37, SD = 6.85	
20. Roschelle et al., 2010 (Pre/Post-test)	Use of SimCalc curriculum (computer simulations) vs. Conventional (business as usual curriculum)	Function	2451	7th/ 8th (Mid.)	7th grade students using Simcalc (n = 796) Pre M = 13.2, SD = 5.7; Post M = 19.0, SD = 6.0 Conventional (n = 825) Pre M = 12.7, SD = 5.7; Post M = 15.0, SD = 5.7 8th grade students using Simcalc (n = 522) Pre M = 11.9, SD = 7.3; Post M = 18.9, SD = 8.7 Conventional (n = 308) Pre M = 12.5, SD = 7.6; Post M = 15.4, SD = 8.4	d = 0.59/**g = 0.59** (p< .001) d = 0.47/**g = 0.47** (p< .001)
21. Tan, 2012 (Pre/Post-test)	Graphing Calculator vs. Conventional Textbook Approach	Probab-ility	65	12th (High)	Experimental (n = 32): Pre M = 1.99, SD = 1.954/Post M = 75.71, SD = 5.03 Conventional (n = 33): Pre M = 2.95, SD = 2.630/Post M = 42.19, SD = 23.1	d = 2.07/**g = 2.01** (p < .05)
22. Yang & Heh, 2007 (Pre/Post-test and Post-test)	Online Virtual Physics Laboratory vs. Conventional Laboratory Setting	Multiple Topics - Science	150	10th (High)	*Process Pre/Post-tests* Experimental (n = 75) Pre M = 23.48, SD = 5.15/Post M = 26.43, SD = 5.15 Conventional (n = 75) Pre M = 23.61, SD = 5.09/Post M = 23.69, SD = 5.09 *Conceptual Post-test* Experimental (n = 75): Post M = 61.01, SD = 11.31 Conventional (n = 75): Post M = 53.89, SD	d = 0.56/**g = 0.55** (p< .01) d = 0.61/**g = 0.60** (p< .01)

					= 12.15	
23. Zehavi, 1988 (Post-test)	*Study 1 (S1, Grade 8)* Mathematics software Vs. Conven. (textbook approach) *Study 2 (S2, Grade 7)* Mathematics software (*n = 78*) Vs. Conv. (*n = 77*)	Function	293	7th & 8th (Mid.)	S1, Section 1: Experimental (n = 84): M = 86, SD = 21, Conventional (n = 54): M = 63, SD = 30; S1, Section 2: Experimental (n = 84): M = 85, SD = 18, Conventional (n = 54): M = 48, SD = 29; S1, Section 3: Experimental (n = 84): M = 77, SD = 24, Conventional (n = 54): M = 33, SD = 18. S2, Section 1: Experimental (n = 78): M = 82, SD = 20, Conventional (n = 77): M = 85, SD = 19; S2, Section 2: Experimental (n = 78): M = 83, SD = 19, Conventional (n = 77): M = 67, SD = 22; S2, Section 3: Experimental (n = 78): M = 71, SD = 22, Conventional (n = 77): M = 66, SD = 24	d = 0.93/**g = 0.91** (p< .001) d = 1.62/**g = 1.60** (p< .001) d = 2.02/**g = 2.00** (p< .001) d = -0.15/g = -0.15 (NS) d = 0.78/**g = 0.77** (p< .001) d = 0.21/g = 0.21 (NS)
Combined Effect Size in favor of Technology (44 Effect Sizes from 23 studies)		**7 func.; 4 thermo.; 3 kinematics; 9 others**	**5406**	**4 Ele. 3 Mid. 14 High 2 Mix**	**Technology: Computer Software/Online Tools (Simulations) = 11 studies; Probes/Sensors = 8 studies; Others = 4 studies (Graphing Calculator, Robotics, Videograph)**	**g = 0.43 (Fixed: 0.39 - 0.48; Random: 0.33 - 0.66)**

Table(s)

Table 4 - Moderator Analysis and Meta-Regression Estimates of Study Characteristics and Effect Sizes for Graphing Technology

Study Features	(1) Discipline	(2) Grade Level	(3) Assess. Type	(4) Sample Size	(5) Assess. Design	(6) Design Quality
Technology Versus Non-Technology						
Effect Sizes (ESs) - Moderator Analysis						
Pooled ES by category and [Number of Studies/Number of ESs] (Random = R; Fixed = F)	Math: 0.47 [8/16] Sci: 0.36 [15/28] Q-Val. (R) = 2.57 Q-Val. (F) = 4.11*	Elem.: 0.62 [4/6] Sec.: 0.43 [19/38] Q-Val. (R) = 1.36 Q-Val. (F) = 2.67	ORE: 0.52 [12/21] No-ORE: 0.35 [11/23] Q-Val. (R) = 0.39 Q-Val. (F) = 11.72***	Small: 0.54 [20/38] Large: 0.37 [3/6] Q-Val. (R) = 1.68 Q-Val. (F) = 12.40***	Researcher-gen.: 0.56 [21/32] Stan.: -0.05 [7/12] Q-Val. (R) = 11.05*** Q-Val. (F) = 106.2***	Ran.: 0.51 [17/27] Quasi: 0.31 [6/17] Q-Val. (R) = 0.016 Q-Val. (F) = 16.56***
Meta-Regression Estimates						
Discipline	-0.325 (0.175)					
Grade level		-0.313 (0.259)				
Assessment Type			-0.106 (0.174)			
Sample Size				0.271 (0.229)		
Assess. Design					0.607*** (0.177)	
Design Quality						0.022 (0.175)
Intercept	0.691*** (0.135)	0.772** (0.243)	0.543*** (0.115)	0.270 (0.209)	0.037 (0.154)	0.483*** (0.137)
k (n)	44 (23)	44 (23)	44 (23)	44 (23)	44 (23)	44 (23)

Table(s)

Table 5 – Investigation Features and NGSS SEPs

Investigation Features	Hypotheses, Questions, or Predictions	Embodied Learning	Model	Plan an Experiment	Choosing/ Manipulate Variables	Collect data	Selecting resources	Analyze or Interpret	Draw or Annotate	Constructing graphs	Explain Content	Make an Argument	Draw conclusions	Reflect	Present
NGSS SEPs	SEP1: Asking questions & defining problems		SEP2: Developing & using models				SEP3: Planning & carrying out investigations		SEP4: Analyzing & interpreting data	SEP5: Using mathematics & computational thinking	SEP 6: Constructing Explanations	SEP7: Engaging in argument from evidence	SEP8: Obtaining, evaluating, & communicating information		
DESIGN STUDIES															
1. Applebaum et al., 2017	*		*	*	*	*	*	*		*	*	*	*	*	
2. Barab et al., 2007	*	*	*	*		*		*		*	*	*	*	*	*
3. Basu et al., 2015	*		*	*	*	*		*		*	*	*	*	*	
4. Chao et al., 2016	*		*		*	*		*		*	*		*	*	
5. Dickes & Sengupta, 2013	*		*		*	*		*		*	*		*		
6.		*	*			*		*		*	*		*	*	

Kamarainen et al., 2013															
7.Kramarski, 1999			*		*			*		*			*		
8. Kwon, 2002	*	*	*		*	*		*		*	*		*		
9. Linn et al., 1987			*	*	*	*		*		*			*		
10. McElhaney & Linn, 2011	*		*	*	*	*		*	*	*	*	*	*	*	
11. Mokros & Tinker, 1987	*	*	*	*	*	*		*		*	*	*	*	*	
12. Peterman et al., 2015						*		*		*			*		*
13. Purba & Hwang, 2017			*			*		*		*	*		*		
14. Ryoo et al., 2018	*		*		*	*		*		*	*		*	*	
15. Songer & Linn, 1991	*		*	*	*	*	*	*		*	*		*	*	

16. Tan et al., 2005	*			*	*	*	*	*			*		*	*	*
17. Varma & Linn, 2012	*		*	*	*	*		*		*	*		*	*	*
18. Vitale et al., 2015	*		*	*	*	*	*	*		*	*		*	*	
19. Zengin & Tatar, 2017			*		*			*		*	*		*		
COMPARISON STUDIES															
1. Adams & Shrum, 1990			*			*		*		*			*		
2. Ardac & Sezen, 2002	*		*	*	*	*		*		*	*		*		
3. Ates & Stevens, 2003	*		*	*	*	*		*		*					
4. Beichner, 1990			*					*		*					
5. Cavanaugh et al., 2008			*		*			*		*					
6. Chiu & Linn, 2014	*		*	*	*	*		*		*	*	*	*	*	*

7. Deng et al., 2011	*		*	*	*	*	*	*		*	*		*	*	
8. Deniz & Dulger, 2012	*	*	*			*		*		*	*	*			*
9. Dorji et al., 2015			*	*	*	*	*	*		*	*		*		
10. Erbas et al., 2015			*					*		*					
11. Friedler & McFarlene, 1997	*		*	*	*	*	*	*		*	*		*		
12. Hsu et al., 2016	*		*	*	*	*		*		*	*		*	*	*
13. Huntley et al., 2000			*					*		*	*		*		
14.Koedinger et al., 1997	*		*		*			*		*	*		*	*	*
15. Koklu & Topcu, 2012			*					*		*			*	*	
16. Lee & Thomas, 2011	*	*	*	*	*	*		*		*	*		*		*

17. Malone et al., 2018	*		*	*	*			*		*	*	*	*	*	
18. Nicolaou et al., 2007	*		*	*	*	*		*		*	*		*	*	
19. Park, 2015	*		*	*	*	*	*	*		*	*		*	*	*
20. Roschelle et al., 2010	*		*		*			*		*	*	*	*	*	*
21. Tan, 2012			*					*		*	*				
22. Yang & Heh, 2007	*		*	*	*	*	*	*		*	*		*	*	
23. Zehavi, 1988			*					*		*	*				
TOTAL	27	6	40	22	30	30	9	42	1	41	33	9	35	22	11
Percent	64	14	95	52	71	71	21	100	2	98	79	21	83	52	26
Search Terms	predict	avatar	model	plan	mani-pulate	data	select	analyze	draw	graph	explain	argue	conclu-sion	reflect	pres-ent
	hypo-thesis	role	robot	exper-iment	pick	evidence	resource	inter-pret	ann-otate	chart	explan-ation	argument	draw	con-sider	write
	generate	play	sim-ulation	develop	choose	collect	pick		label	plot	detail	arguing	drew	eval-uate	report
	spec-ulate	char-acter	toy cars	design	select	gather	choose			draw		debate	con-clude		letter

		robot	tablet				equip-ment			con-struct		refute	solution		work-sheet
							tools					claim	make		table
												advise			
												mentor			
												consensus			

Figure(s)

Figure 1: Logic Model for K-12 Graph Technology Meta-Analysis

Situation: Graph proficiency plays a critical role in K-12 students' science and mathematics learning. There is no meta-analysis that has attempted to review graph technology studies across science and mathematics to inform new directions for science and mathematics instruction and assessment.

Input	**Outputs** — *Activities for Research Questions*	**Outputs** — *Participants Influenced*	**Outcomes – Specific and Broad**
Graph-based articles from 25 peer-reviewed journals: *American Educational Research Journal*; *Cognition and Instruction*; *Computers and Education*; *Educational Studies in Mathematics*; *Environmental Education Research*; *International Journal of Science and Mathematics Education*; *International Journal of Science Education*; *International Journal of Mathematical Education in Science and Technology*; *Journal for Research in Mathematics Education*; *Journal of Computer-Assisted Learning*; *Journal of Computers in Mathematics and Science Teaching*; *Journal of Educational Psychology*; *Journal of Educational Technology & Society*; *Journal of Environmental Education*; *Journal of Mathematical Behavior*; *Journal of Research in Science Teaching*; *Journal of Science Education and Technology*; *Journal of the Learning Sciences*; *Mathematical Thinking and Learning*; *Mathematics Education Research Journal*; *Research in Science Education*; *School Science and Mathematics*; *Science Education*; *Studies in Science Education*; and *Technology, Knowledge, and Learning*.	*Identified and analyzed:* -***RQ1***: 19 design studies (pre/post-tests) that report the impact of K-12 graph learning (Table 2) -***RQ2***: 23 pre/post-tests and post-test only K-12 comparison studies that compare technology versus non-technology for graph learning. (Table 3) **-RQ1 and RQ2:** Form (Construction, Comprehension, and Critique) and Format (Multiple-Choice, Open Response Recall, and Open Response Explanation) **-RQ3:** Investigation features aligned with NGSS Science and Engineering Practices (SEPs)	K-12 science and mathematics students K-12 science and mathematics teachers Science and mathematics education researchers Instructional designers Assessment designers (state, national, and international) Policy makers (state, national, and international)	**Specific Outcomes** • Evidence of the overall impact of pre/post design study interventions to graph proficiency (RQ1) • Evidence of the overall impact of technology-based approaches over non-technology approaches to graph proficiency (RQ2) • Evidence of the nature and type of assessments for graph-based interventions (RQ1-RQ2) • Evidence of investigation features targeted by graphing technologies and identifying gaps in NGSS SEPs (RQ3) **Broad Outcomes** • Documented areas for improvement in current K-12 science and mathematics graph technology instruction and assessment • Documented gaps in the existing research literature • Better communication of assessment items in research articles • Greater high-quality studies in K-12 graph technology • Improved alignment of K-12 graph-based studies with expectations of the Next Generation Science Standards (NGSS) and the Common Core Mathematics Standards (CCMS)

Assumptions

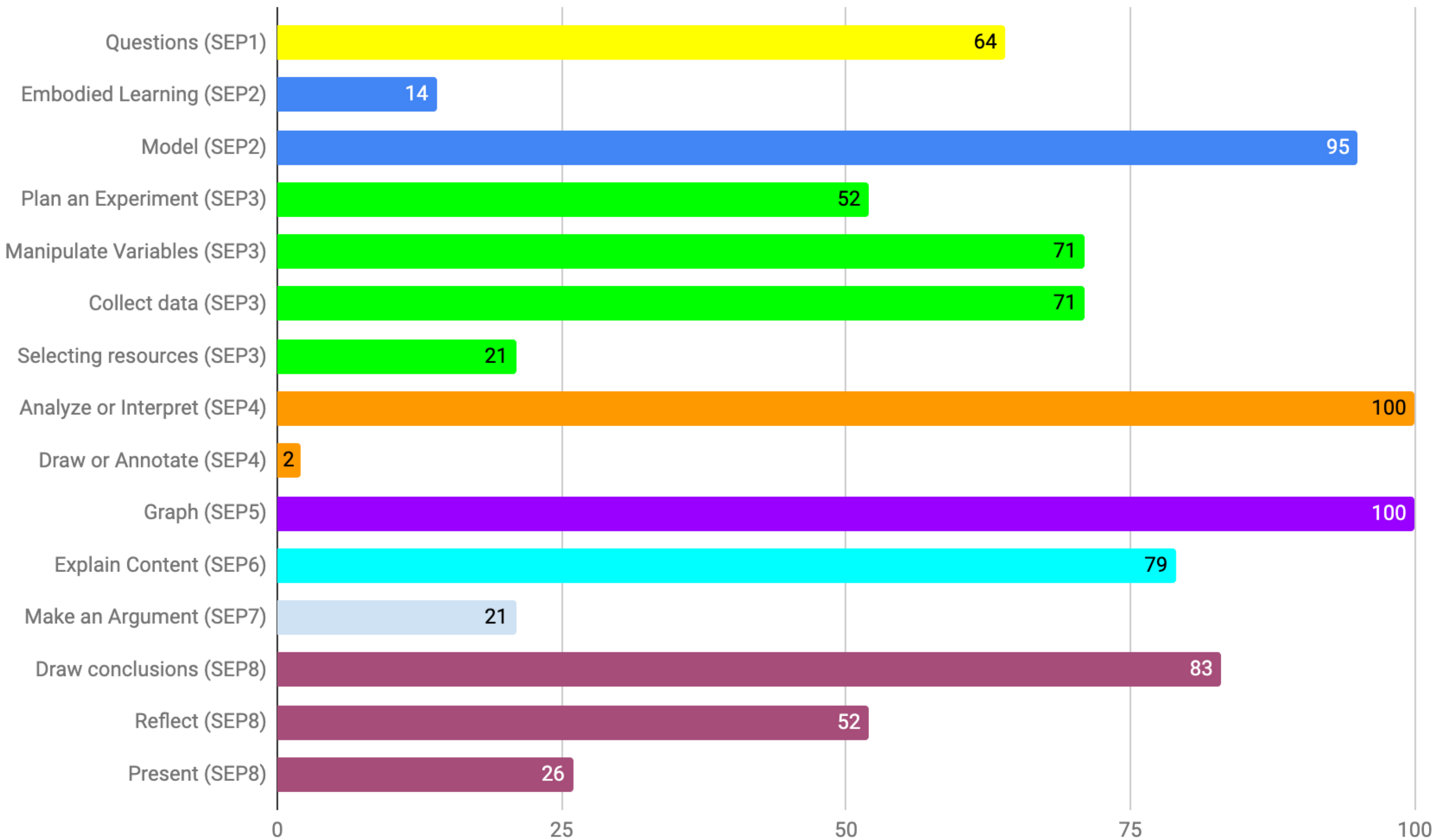
- Searching specific journals is more productive and time-efficient than a database search gi. en the broad use of the word 'graph'.
- Peer-reviewed journals will result in better quality and more consistent studies than research reports, conference papers, etc.

External Factors

Potential factors influencing the overall outcomes include (a) the discipline: science or mathematics, (b) the grade level: elementary or secondary, (c) the quality of the study: randomly-controlled trial or quasi-experimental, (d) the sample size: small (<250) or large (250+), (e) assessment design: researcher-generated or standardized assessments, and (f) assessment type: open-response explanations or non-open-response explanations.

Figure(s)

Figure 2 – Investigation Features and NGSS SEPs



Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1418423. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This material is also based on work supported by the College of Science and Mathematics at California State University Fresno. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of California State University Fresno.

*Highlights (for review)

Research Highlights

- Design studies illustrate that graph technologies improve student learning in general
- Comparison studies show student outcomes are better using graph technologies compared to conventional approaches
- Many studies lack features like planning experiments, arguing from evidence, and evaluating and communicating information