4-2021

To Prove or Improve, That is the Question: The Resurgence of Comparative, Confounded Research Between 2010 and 2019

Peter C. Honebein
Boise State University

Charles M. Reigeluth
Indiana University

This is a post-peer-review, pre-copyedit version of an article published in *Educational Technology Research & Development*. The final authenticated version is available online at https://doi.org/10.1007/s11423-021-09988-1

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *Educational Technology Research & Development*, published by Springer. Copyright restrictions may apply. https://doi.org/ 10.1007/s11423-021-09988-1.

To Prove or Improve, That is the Question: The Resurgence of Comparative, Confounded Research Between 2010 and 2019

Peter C. Honebein*
Customer Performance Group

and

Indiana University

and

Boise State University
peter@honebein.com

Charles M. Reigeluth
Indiana University

Abstract

Between 2010 and 2019, ETR&D experienced increased publication of a specific type of research that does not provide useful knowledge to the instructional design field. This type of research is *research to prove*, which entails pitting an incumbent, "traditional" learning experience against a new, innovative learning experience that lacks maturity. Additionally, under closer inspection, these new, innovative learning experiences show significant gaps of good design judgment, in terms of their alignment with the instructional theory framework. This type of research robs the instructional design field of important and useful data associated with effectiveness, efficiency, and appeal outcomes. To provide evidence for our claims, we reviewed 39 ETR&D articles between 1980 and 2019 and 41 articles in non-ETR&D journals between 2009 and 2018 that represented traditional instruction comparisons. Our conclusion is that a change in ETR&D editorial policies around 2010, such as reviewers having more power than editors in determining which papers get published, led to the unintended consequences this paper reports. We provide recommendations for addressing this situation.

Keywords: instructional theory framework, Culture Four, Culture Five, traditional instruction, traditional methods, design research, design judgment

In our roles of instructional designers, peer-reviewers, and researchers, we are concerned about an accelerated acceptance of comparative research studies published between 2010 and 2019 in *Educational Technology Research and Development (ETR&D)* and other journals. This concern grows when we see a comparative research paper where one treatment is "traditional" instruction (or traditional method, or traditional approach). The term "traditional instruction" is typically seen as instruction that is didactic, face-to-face, teacher-centered, and reliant on textbooks and lectures (Raja & Najmonnisa, 2018; Stacker & Horn, 2012). Yet researchers have stretched this definition to describe incumbent solutions as well. Cunningham (1986) used the terminology "bad guy" to describe traditional, incumbent instructional treatments. A research study then introduces, with a fanfare of theoretical justification, the "good guy," or "novel" instruction, which usually involves some mixture of instructional methods (for example, problem-based learning) and media methods (for example, virtual reality) that the researcher expects will outperform traditional instruction. Such research is the epitome of what Reigeluth and An (2009) call *research to prove.*

Comparative research pitting the traditional bad guy against the novel good guy has been an intense subject of debate and a well-known pariah in our field for four decades, yet researchers and reviewers seem to pay little attention to reducing it, and no one seems to do anything about it except complain every so often. Robert Ebel, a past president of the American Education Research Association (AERA), suggested basic research in education is limited, as "the process of education is not a natural phenomenon" (Farley, 1982, p. 18). Thus, research to prove is a pariah because it is typically an easy, one-off, useless piece of research that pads one's curriculum vitae (Driscoll and Dick, 1990;

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *Educational Technology Research & Development*, published by Springer. Copyright restrictions may apply. https://doi.org/ 10.1007/s11423-021-09988-1.

Tanner, 1998). It is research that emphasizes a traditional view of scientific rigor over relevance (Phillips et al., 2012; Stokes, 1997), rather than the much harder task of finding a more complementary balance between the two, which Salomon (1991) calls the analytic and the systemic. Or, as Ebel suggests, education "is in need of creative invention to make it work better" (Farley, 1982, p. 18).

Comparative research to prove has several disadvantages. First, methods are situational and can have many variations (Reigeluth & Carr-Chellman, 2009), which means instructional methods work well in some situations but not others. Comparative studies rarely fully describe situational variables and variations, nor do they replicate in different situations or different variations.

Second, every method undergoes a pattern of development called the S-curve (Branson, 1987). Figure 1 shows that research to improve tends to be more useful during the first (slow improvement) and second (rapid improvement) phases of development. Research to prove tends to be more useful when a method moves into the third phase (maturity), which reflects a diminishing rate of improvement (Phillips et al., 2012). In comparative research, the "traditional" method has typically reached maturity, but the "novel" method is typically relatively new (in phase 1 or 2), which is rarely acknowledged. Comparing the two methods at such disparate points in their development can lead to prematurely rejecting a promising new method.

Insert Figure 1 About Here

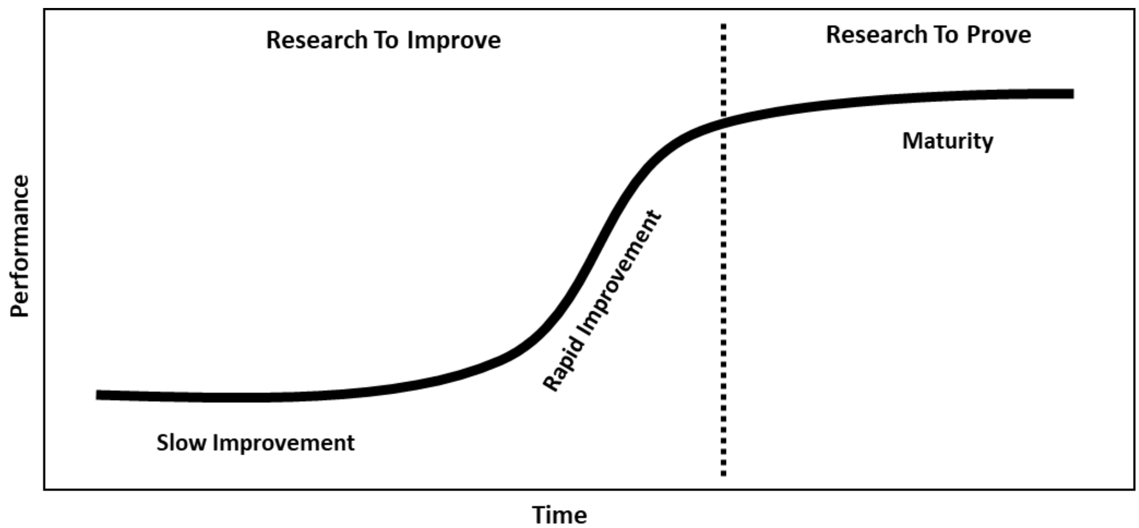

Figure 1: The S-curve of development for a method or technology

----

Third, and perhaps most important, is that comparative research to prove does not provide data about the ways that relatively new methods can move up the S-curve – ways to improve the method. Instructional designers and teachers would find this knowledge very useful, which means it is an important focus area for researchers.

The last active debate about comparative studies in ETR&D started in the mid-1980's and continued, sporadically, until 2000 (Cobb, 1997; Clark, 1983, 1985, 1986, 1989, 1994; Cunningham, 1986; Driscoll and Dick, 1999; Jonassen et al., 1994; Kozma, 1994, 2000; Lockee, Burton, & Cross, 1999; Morrison, 1994; Parkhurst, 1992; Reigeluth, 1989; Reiser, 1994; Richey, 1998; Ross, 1994; Shrock, 1994; Tennyson, 1994; Ullmer, 1994). If there was a last word, it was from Kozma (2000), who said, "traditional experiments often are not able to accommodate the complexity of these real-world situations" and "this confounding makes it difficult, if not impossible, to disentangle one component of a design from another because the various components are designed to work together (p. 10). These situational factors have ultimately evolved into significant interest in design-based research methods at the 2020 AECT conference, as well as those promoting "improvement science" (see e.g., Lewis, 2015; LeMahieu, Edwards, & Gomez, 2015).

Then, around 2010, something happened that accelerated a resurgence of comparative studies in ETR&D, as the data we report below will show. As ETR&D peer reviewers, we began to notice this in articles we were reviewing in 2017. What had changed was ETR&D's editorial policies (Spector, 2017). Power for deciding what gets published had shifted from editors to reviewers, and editors encouraged more contributions from non-North American researchers.

Let us be clear that we strongly support this policy innovation and we commend Dr. Spector, the ETR&D advisory board, other editors, and reviewers for facilitating the adoption of this innovation. However, as with all innovations, there is a risk of unintended consequences, and we feel this is one of those situations, specifically in terms of the resurgence of comparative research. As the reader will see in our data, most of the comparative study resurgence in ETR&D comes from non-North American sources. As North American designers, researchers, and reviewers, our values about research and design were perhaps shaped by the literature presented above; on Salomon's (1991) scale, we North Americans are likely weighted more towards the "systemic" side due to direct experiences with the comparative studies debate. But then again, perhaps non-North American designers', researchers', and reviewers' values were shaped by other forces weighted towards Salomon's "analytic side," due to lack of experience with the comparative studies debate and/or other cultural norms and influences.

The purpose of this paper is to recommend ways we, as a field, can resolve the unintended consequence we described above. The primary contributions that this paper makes are 1) linking the instructional theory framework to the comparative study debate as a way to identify gaps in comparative research design fundamentals and outcome measures, 2) providing quantitative data regarding the resurgence of comparative, traditional-versus-novel, research designs, and 3) recommending to editors, reviewers, and authors actionable, relevant, training and non-training interventions that could lead to shifting comparative studies from research to prove to research to improve. We expect, and welcome, subsequent debate as this topic evolves.

Our research questions are:

1. What are the trends of comparative, empirical research studies in ETR&D that include one or more of the phrases "traditional method," "traditional instruction," and "traditional approach" between 1980 and 2019?
2. What are the trends of comparative research studies in non-ETR&D journals publishing instructional design research that include one or more of the phrases "traditional method," "traditional instruction," and "traditional approach" between 2009 and 2018?
3. How are elements of the instructional theory framework and Culture Four – specifically conditions, values, and methods – represented in comparative, empirical, peer-reviewed studies that assess the relative efficacy of the learning experiences therein?

We organized the remainder of this paper in the following structure, starting with a conceptual framework, then presenting the data, and then providing synthesis and recommendations.

1. Conceptual Framework
2. Study 1: ETR&D, 1980 to 2019
3. Study 2: Flipped Classrooms, 2009 to 2018
4. Synthesis
5. Recommendations
6. Limitations

Conceptual Framework

<u>Research Classification</u>

The purpose of this section of the paper is to describe the research classifications that we used as dependent variables in our two studies. As cited above, our field has substantial literature about what makes good research, as well as frameworks for the various research categories and methods relevant to instructional technology research (Reeves &

Oh, 2017). Reeves and Oh specify six research categories. One of those categories, *design and development*, is relevant to our present study since it involves research goals associated with the creation and improvement of learning experiences.

For greater precision, we expand the Reeves and Oh structure to accommodate Reigeluth and An's (2009) conceptions of research to prove and research to improve. We suggest that there are subcategories within Reeves and Oh's design and development category that are useful for classifying different types of design and development research. These subcategories are research to prove, research to improve, research to describe, and research to test feasibility.

Research to Prove

Research to prove is confirmatory research (see, e.g., Burbach, Matkin, & Fritz, 2004). It is quantitative, and, within the scope of the instructional theory framework (Honebein & Reigeluth, 2020, 2021), it has two theoretical branches, descriptive and design (Reigeluth, 1999):

For descriptive theory, this kind of research to prove advances descriptive theories, which aim to *describe* cause-and-effect relationships or natural processes, such as what happens when a person processes information (information enters short-term memory, then enters long-term memory). Researchers can use research to prove for descriptive theory to test predictions (what will be the effects?) or explanations (what were the causes?), but it is typically insufficient to inform design decisions (Reigeluth, 1999; Reigeluth & Schwartz, 1989), due primarily to the highly situational and interactional nature of instructional method variables. Examples of this kind of research include studies of short-term memory, such as studies of Miller's "magic number" of seven items plus or minus two (see, e.g., Schweickert & Boruff, 1986).

For design theory, this kind of research to prove advances design theories, which aim to *prescribe* what method(s) are preferable to achieve a goal. Thus, this type of research guides choices about whether one method is better than another, or which approach is the "best available route." Research to prove for design theory is prescriptive, but in a probabilistic rather than deterministic way. The conclusions section reports win, lose, or draw results. Comparative research, such as Kuo and Hooper (2004), is a typical example of this kind of research.

Research to Improve

Research to improve is exploratory research (Reigeluth & Frick, 1999; Stebbins, 2001). It is largely qualitative, and, within the scope of the instructional theory framework, it uses formative evaluation, often with mixed methods, to improve an instance (case) of instruction as an example of an instructional method, model, or theory. Typical research methods include action research (Efron & Ravid, 2020; Stringer, 2008; Stringer & Aragon, 2021), design experiments (Cobb et al. 2003), design-based research (Barab & Squire, 2004; Collins et al., 2004; Design-Based Research Collective, 2003; Wang & Hannafin, 2005), evaluation research (Phillips et al., 2012), and formative research (Reigeluth & An, 2009; Reigeluth & Frick, 1999). It often involves two or more iterations through which researchers can observe a method, make changes based on formative data, observe the effects of those changes, and propose additional possible improvements based on stated criteria for effectiveness, efficiency, and appeal. The conclusions section in a research paper proposes possible improvements based on the data from research participants, such as learners, teachers, and other stakeholders, for a particular set of situation variables. English and Reigeluth (1996); Honebein, 1994; Kakos-Kraft, Honebein, Prince, and Marrero, 1997; Lee and Reigeluth (2003); and Lee, Jahnke, and Austin (2021) are examples of research to improve.

Research to Describe

Research to describe does not attempt to prove or improve. This approach typically uses naturalistic inquiry to describe and characterize a learning experience in the way that things happened. It is typically ethnographic, with a narrative-oriented product. The conclusions section does not provide any recommendations. Lubin and Ge (2012) is an example of research to describe.

Research to Test Feasibility

Research to test feasibility also does not attempt to prove or improve. This approach uses a variety of mixed methods to answer the question, "is a case of an instructional method workable?" Thus, the focus is on viability and feasibility of a method within a specific situation. The conclusions section provides a conclusion about whether a method is feasible or not. Lee and Thomas (2011) is an example of research to test feasibility.

Thus, in our two studies, we classify design and development research using five subcategories:

1. Research to prove for descriptive theory
2. Research to prove for design theory
3. Research to improve
4. Research to describe
5. Research to test feasibility

Instructional Theory Framework

The purpose of this section of the paper is to describe the elements of the instructional theory framework (Honebein & Reigeluth, 2020, 2021; Reigeluth, 1983; Reigeluth 1999; Reigeluth & Carr-Chellman, 2009) that we used as dependent variables in our two studies. The purpose of the instructional theory framework (Figure 2) is to guide instructional theorists and researchers in creating and improving instructional theories or understanding instructional theories. The framework represents three key sequential parts. First, there is the situation, which represents the conditions and values for which the instructional theory is intended. *Conditions* represent matters of fact. *Values* represent matters of opinion. Thus, together, they represent objective and subjective ways of knowing.

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *Educational Technology Research & Development*, published by Springer. Copyright restrictions may apply. https://doi.org/ 10.1007/s11423-021-09988-1.

--

Insert Figure 2 About Here

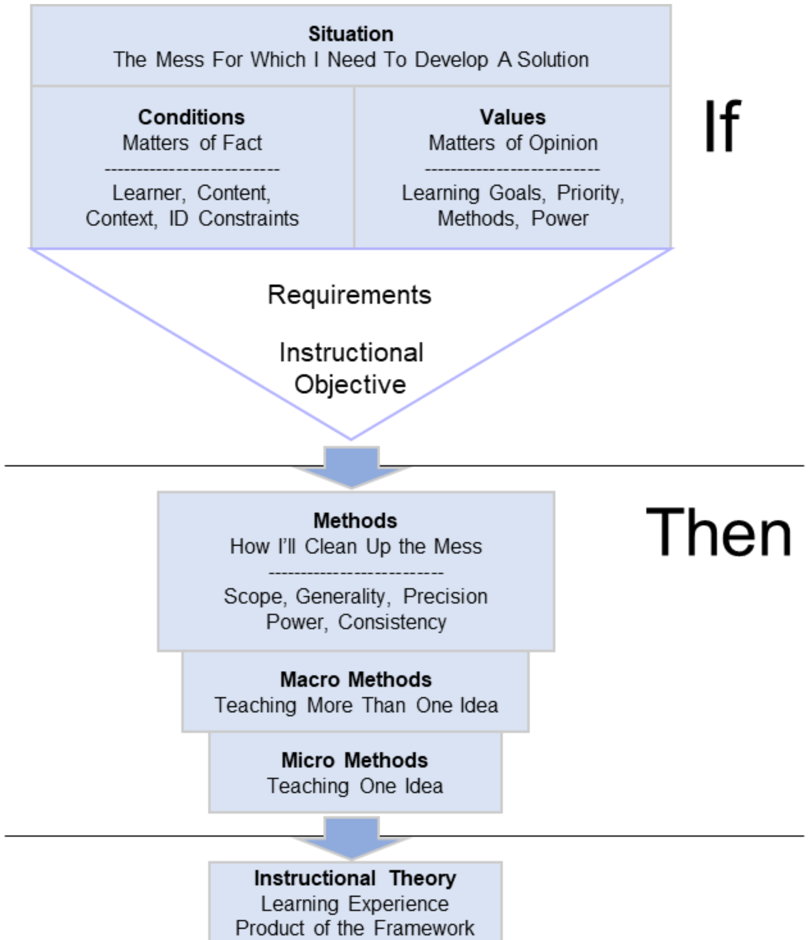

Figure 2: The instructional theory framework

--

In instructional situations, conditions and values represent the type of data that one would collect through a front-end analysis. For conditions, this means assembling information about the learner, content, context, and instructional design constraints. For values, this means assembling information about how stakeholders feel about learning goals, the priority of outcomes (effectiveness, efficiency, and appeal), possible instructional methods, and who will have power to influence or control the nature of the instruction (Reigeluth & Carr-Chellman, 2009).

One can describe the synthesis of the situation in many ways, such as front-end analysis reports and design documents that specify instructional objectives, as well as organizational, functional, and non-functional requirements (Honebein, 2018). These requirements typically specify the need for systematic instructional design and formative evaluation.

But ultimately, if you imagine a funnel, a designer refines the synthesis of the situation into an instructional objective in the form of a *designer objective*, which is typically more detailed and elaborated than the instructional objective(s) a designer presents to learners as part of the learning experience. There are many forms of designer objectives (Czeropski & Pembroke, 2017; Merrill, 1983; Reigeluth & An, 2021), but for the purposes of this paper, we use Mager's (1984) formula of *condition*, *behavior*, and *criterion* to represent the common information that should be present in a designer objective.

Second, with the organizational requirements and well-formed design objectives elicited from the situation, a designer continues work by determining which methods best fit the situation. Methods in this context typically represent instructional methods and media methods, but can also include methods associated with the "layers" specified by Gibbons and Rogers (2009): *content, strategy, message, control, representation, media-logic, and data management*. Kozma (2000) used the phrase "learning environment," and we use the phrase "learning experience" to represent a unique instance that incorporates one or more of these seven layers for a given situation. Method characteristics include scope (macro or micro), generality, precision, power, and consistency (Reigeluth & Carr-Chellman, 2009).

Third, what the instructional theory framework ultimately delivers is an *instructional theory* (also known as a learning experience), which is *prescriptively* a collection of methods (instructional, media, or otherwise) that best fit a certain set of situations. An instructional theory is different from a *learning theory*, such as behaviorism, cognitivism, or constructivism, since a learning theory 1) descriptively explains the learning process, typically what might be going on in one's head, and 2) does not include any methods. An instructional theory is also different from an *instructional model*, since an instructional model, while it does include methods, lacks any situational connection. Anyone can create, modify, improve, or "mash up" instructional theories. The key principle to remember is that instructional theories are situational, thus, any creation or change must fit the situation(s).

The intersection of the instructional theory framework and the design/development research category is best illustrated by Leslie Briggs' Culture Four aspects (Briggs, 1984) (**Error! Reference source not found.**). These aspects, which Briggs used to prescribe key design/develop activities, align nicely with elements of the instructional theory framework.

Insert Table 1 About Here

Table 1: Utilization of the Instructional Theory Framework factors in each of the Culture Four aspects.

Culture Four Aspects	Instructional Theory Framework Factors
1. The researchers do accurately classify the type of learning outcome being studied, and they supply objectives and test items so we can check the classification.	• **Conditions**: Learner, Content, Context, ID Constraints • **Values**: Learning Goals, Priority, Power
2. The passages of materials are similar to textbook chapters in length, and they are real curriculum materials.	• **Methods**: Scope, Generality, Precision, Power, Consistency • **Instructional Theory**: Product of the Framework
3. The materials were systematically designed and formatively evaluated.	• **Conditions**: Learner, Content, Context, ID Constraints • **Values**: Learning Goals, Priority, Methods, Power • **Methods**: Scope, Generality, Precision, Power, Consistency
4. The tests really measure the learner's ability to perform the behaviors to the standard specified in the objective.	• **Values**: Learning Goals, Priority • **Methods**: Scope, Generality, Precision, Power, Consistency

----

Thus, in our two studies, we observed learning experiences through the following instructional theory framework and Culture Four concepts/prescriptions:

1. Presence of instructional objectives (conditions, behavior, standard of performance)
2. Presence of instructional outcomes (effectiveness, efficiency, and appeal)
3. Presence of an improvement focus involving formative evaluation (instructional methods, media methods, and so on)

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *Educational Technology Research & Development*, published by Springer. Copyright restrictions may apply. https://doi.org/ 10.1007/s11423-021-09988-1.

Study 1: ETR&D 1980 to 2019

Method

Our literature review method generally followed that of Driscoll and Dick (1999), but in a less arbitrary manner. Like Driscoll and Dick, we focused solely on ETR&D articles. Unlike Driscoll and Dick, we selected a much longer timeframe spanning 1980 to 2019, which 1) included the foundational literature discussed in this paper's introduction, and 2) enabled our analysis and results to be represented in terms of four equal time spans (decades).

Our procedure to find relevant articles to answer our second research question was as follows:

1. Select one of the phrases: "Traditional Approach," "Traditional Method," "Traditional Instruction"
2. Go to the ETR&D website provided by our institution.
3. Enter the phrase into the search field, bounded by quotation marks: For example, "traditional approach"
4. Select "Newest First"
5. Enter 1980 as the Start Year, and 2019 as the End Year
6. Click the search icon (Figure 3).
7. Click the Download Search Results (CSV) button to generate a spreadsheet of all the results.

Insert Figure 3 About Here

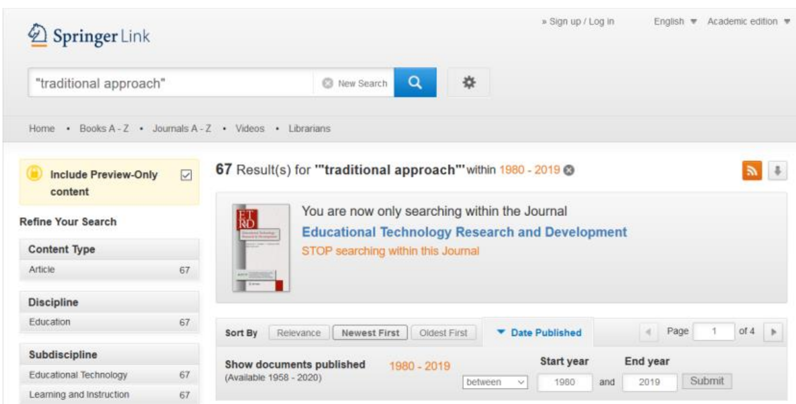

Figure 3: The literature search process in ETR&D

Our procedure to determine if an article met our inclusion criteria was to:

1. Open each article's PDF file of the manuscript.
2. Search the article for the word "traditional."
3. Review the article's abstract and method section for information about whether the article involved the comparison of two or more treatments.
4. If both conditions were true, indicate "Yes" in the Comparison column on the Search Results spreadsheet; otherwise, indicate No.
5. Enter any relevant notes about both Yes and No articles on the spreadsheet.

We repeated this procedure for each of the three phrases. Of the 217 articles we reviewed, 39 articles met our criteria. We then conducted an in-depth analysis of each of the 39 articles. To do this, we created a spreadsheet containing columns representing our coding categories shown in Appendix A. Then, we reviewed each article, entering the appropriate coding into each column. We reviewed each article multiple times, with latter reviews often utilizing relevant search terms to determine if an article included a specific concept (for example, some articles referred to "formative evaluation" as "pilot testing"). Additionally, to facilitate discussion between the authors and to expedite illustrative examples for publication, we created a Comment for some spreadsheet cells, in which we then pasted text copied directly from the article that further elaborated the coding. For example, Van Eck and Dempsey (2002) was one of the few studies that reported conducting a relatively robust pilot test/formative evaluation of their treatment, so we copied and pasted the following text into the comment associated with the Formative Evaluation cell: "Schools A and B were used for pilot testing and field trials (respectively) of the game, and School D was ultimately unable to participate" (p. 26).

The researchers discussed and agreed upon the critical characteristics of the five research classifications: 1) research to prove for descriptive theory, 2) research to prove for design theory, 3) research to improve, 4) research to describe, and 5) research to test feasibility. Each researcher independently classified the 39 studies using the data in the spreadsheet, as well as consulting individual articles when necessary. Inter-rater reliability was 82%, representing "almost perfect" on the Landis and Koch (1977) scale. The researchers then discussed the seven studies in which their ratings differed. The researchers resolved the differences, leading to 95% of the studies matching the original rating of one or both of the researchers.

The researchers then used pivot tables to analyze the data, focusing on historical trends and specific instructional design practices consistent with Culture Four.

Results

Table 2 presents a summary of primary statistics elicited from the research data. Of the 39 "traditional instruction" comparison articles, more than two thirds, 69%, were published between 2010 and 2019 (Figure 4). The most substantial increase of published articles between 2010 and 2019 (N=27) came from non-North American regions (Figure 5). The ETR&D section to which the editors originally assigned the studies (Research or Development) appeared relatively balanced. Note that older issues of ETR&D and its predecessor ECTJ did not designate specific sections for articles.

Insert Table 2 About Here

Table 2: ETR&D Summary Statistics

Number of Traditional, Comparative Studies By Decade	1980-1989	4	10%
	1990-1999	3	8%
	2000-2009	5	13%
	2010-2019	27	69%
Number of Studies by Regional Source Between 2010-2019	Asia	11	41%
	Europe	7	26%
	North America	7	26%
	Middle East	1	4%
	South America	1	4%
Number of Studies by Assigned Sections in ETR&D	Development	15	38%
	Research	12	31%
	Cultural Perspectives	1	3%
	No Designation	11	28%
Number of Studies by Research Classification	To Prove For Descriptive Theory	0	0%
	To Prove For Design Theory	36	92%
	To Improve	0	0%
	To Describe	2	5%
	To Test Feasibility	1	3%
Number of Studies Including Instructional Objective Elements	Condition	6	15%
	Behavior	21	54%
	Criterion	3	8%
Number of Studies Including Instructional Outcome Measures	Effectiveness	38	97%
	Efficiency	8	21%
	Appeal	28	72%

Insert Figure 4 About Here

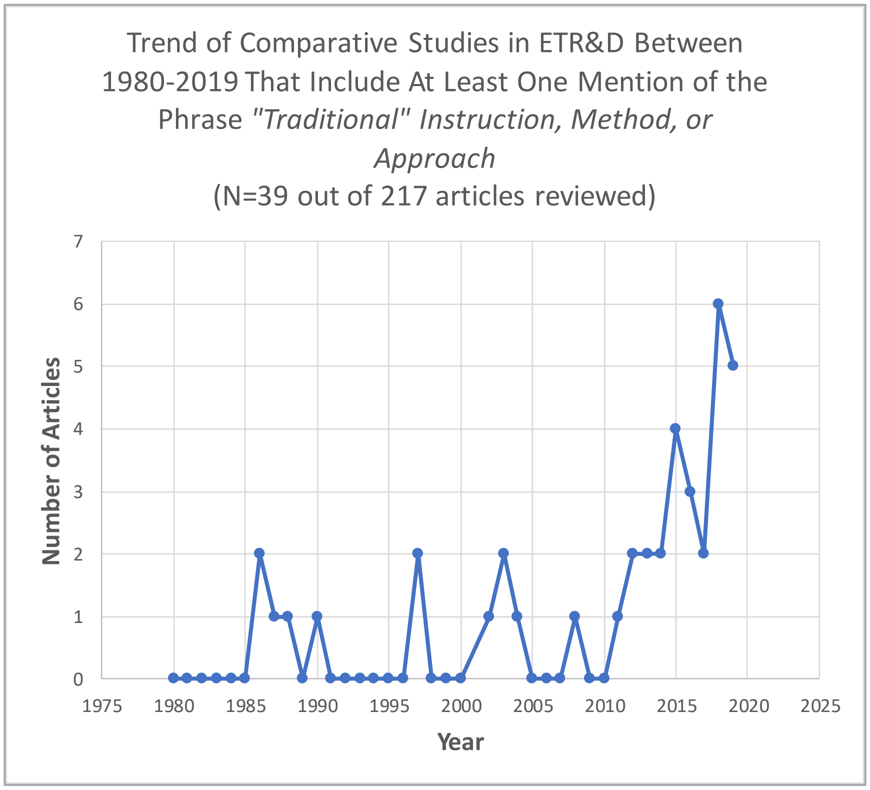

Figure 4: Comparative research study trends in ETR&D between 1980-2019. The number of comparative studies increased substantially between 2010 and 2019.

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *Educational Technology Research & Development*, published by Springer. Copyright restrictions may apply. https://doi.org/ 10.1007/s11423-021-09988-1.

Insert Figure 5 About Here

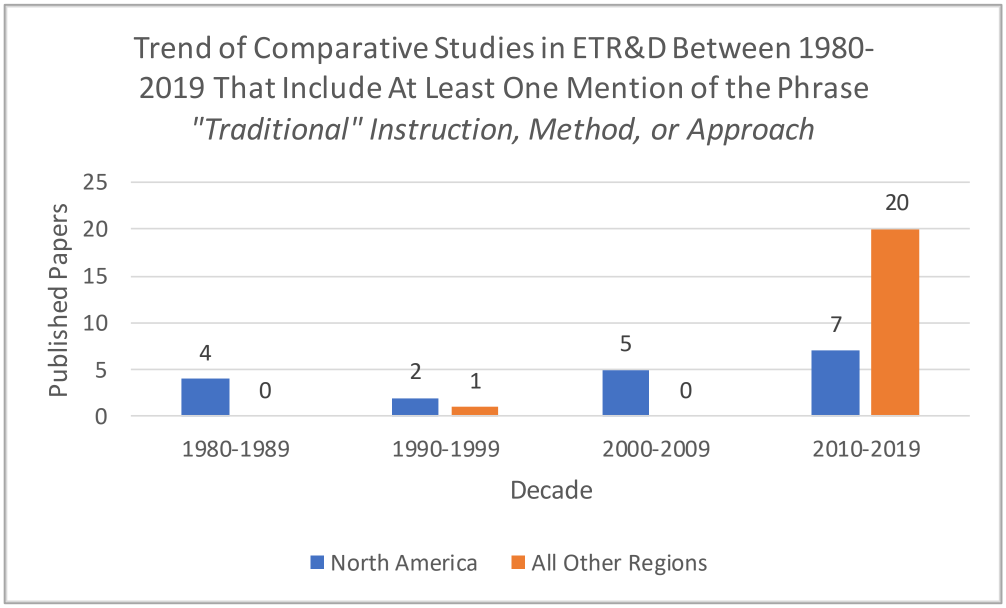

Figure 5: Distribution of published articles comparing North America with All Other Regions. The number of articles published from non-North American regions increased substantially between 2010 and 2019.

None of the articles described instructional objectives in a complete form that reflects Mager's (1984) classic elements (condition, behavior, criterion) or Czeropski and Pembroke's (2017) alternative agile, story-based elements. Of the classic elements, behavior was most frequently specified (54%), distantly followed by conditions (15%) and criterion (8%).

In terms of formative evaluation (or pilot testing), only 13% of the articles described the researcher performing formative evaluation or pilot testing of their novel treatment. None of the articles described performing formative evaluation for any "traditional" treatment.

As expected, nearly all articles (97%) included an effectiveness outcome measure. The efficiency outcome measure was relatively rare (21%), and when it was included, it focused on the differences in instructional time between treatments. The appeal outcome measure was fairly widespread (72%), but in a few articles, researchers collected this measure for only the novel treatment, ignoring the appeal of the "traditional" treatment.

The researchers classified the majority of the articles as *research to prove for design theory* (92%), due to some extent, to our criterion of "traditional" instruction, method, or approach. *Research to describe* accounted for 5%, and *research to test feasibility* accounted for 3%. No articles met the criteria for *research to prove for descriptive theory* or *research to improve*.

Discussion

Our data shows that ETR&D has become much more geographically inclusive in the most recent decade for empirical scholarship that investigates instructional theories and models. This was Michael Spector's (ETR&D's previous editor-in-chief) goal, "to encourage more international contributions from outside North America" (Spector, 2017, p. 1416). Compared to the three previous decades of North American publication dominance, the 2010's showed that non-North American regions have been more productive than North American regions in producing research that compares traditional methods with novel methods.

With that productivity comes a concern. Studies that include media methods as an independent variable were twice as frequent as studies where instructional methods were the only independent variable (26 vs 13 studies). Between 2010 and 2019, 19 articles were media-oriented studies. The 2010's saw "traditional" classroom learning experiences being compared to computer-based instruction, video, robots, mobile devices, computer games, tablet PCs, and mBots. Does this signal that potentially confounded research is now acceptable?

Kozma (2000) acknowledged the messiness of complex, often real-world, situations in which researchers find themselves working. He makes the point that "Traditional experimental designs are not able to accommodate the complexity of these real-world situations" (p. 10). In other words, no matter how hard a researcher tries, conducting an experimental comparison of two or more learning experiences that use different methods, media, and perhaps unique blends of the other five layers of Gibbons and Rogers (2009) model, in a real-world context, is not useful. The research will be confounded. In such complexity, this type of *research to prove* does not provide a win, lose, or draw result that helps a client make a wise adoption-of-innovations decision. We agree with Kozma (2000) when he writes, "It is the interplay of [method and media] within the learning context that should be the focus of our research and theory" (p. 19).

Study 2: Flipped Classrooms 2009 to 2018

Method

The second study used Al-Samarraie, Shamsuddin, and Alzahrani's (2019) literature review of research on the flipped classroom instructional model. We first observed the Al-Samarraie et al. article in Study 1, since it met the criteria of the ETR&D literature search we performed. However, as it was itself a literature review of mostly non-ETR&D papers, it was incompatible with Study 1's ETR&D focus. Furthermore, because of the reviewer criticism Driscoll and Dick (1999) experienced for relying only on ETR&D articles to understand design and research trends, it seemed to us to be serendipitous that Al-Samarraie et al's. review of 85 studies between 2009 and 2018 included numerous studies that compared traditional and flipped instruction. Thus, we decided to use it as a data set for a second study of non-ETR&D articles (yet articles that were implicitly deemed acceptable by ETR&D reviewers) to further understand the issues associated with the comparative research problem.

Thus, the Al-Samarraie et al. (2019) research enabled us to expand our evidence in a way that blends the narrow band of a single instructional model (flipped instruction) with the wide band of journals investigating instruction in seven disciplines: "Engineering and technology (16.2%); Mathematics (9.4%); Medical and health sciences (23.5%); Natural sciences (20%); Social sciences and humanities (20%); Education (8.2%); and Arts (3.5%)" (p. 4).

Articles for inclusion needed to meet the same criteria as for Study 1 in terms of 1) source being a peer-reviewed journal, 2) a mention of "traditional" as it relates to a certain type of instruction, and 3) that the article reports a comparison of some type. The procedure for reviewing and documenting articles was the same as specified for Study 1.

Of the 85 articles we reviewed, 45 met our criteria, but only 41 were usable. One article was duplicated in Al-Samarraie et al's. (2019) list (Simpson & Richards, 2015), one article was a duplicate from Study 1 (Davies et al., 2013), and we could not access two articles (Fraga & Harmon, 2014; Chien & Hsieh, 2018).

Each researcher independently classified the 41 studies based upon research type. Inter-rater reliability was 78% ("substantial" per Landis and Koch (1977)). The researchers then discussed the 9 studies in which their ratings differed. The researchers resolved the differences, leading to 100% of the studies matching one or both of the researcher's original ratings.

The researchers then used pivot tables to analyze the data, focusing on historical trends and specific instructional design practices consistent with Culture Four.

Results

Ninety-five percent of the flipped instruction articles we reviewed included media methods as an independent variable, with 90% including both media methods and instructional methods (Figure 6). Additionally, the data shows the number of flipped instruction articles peaked in 2015, representing 41% of the total articles in our sample.

Insert Figure 6 About Here

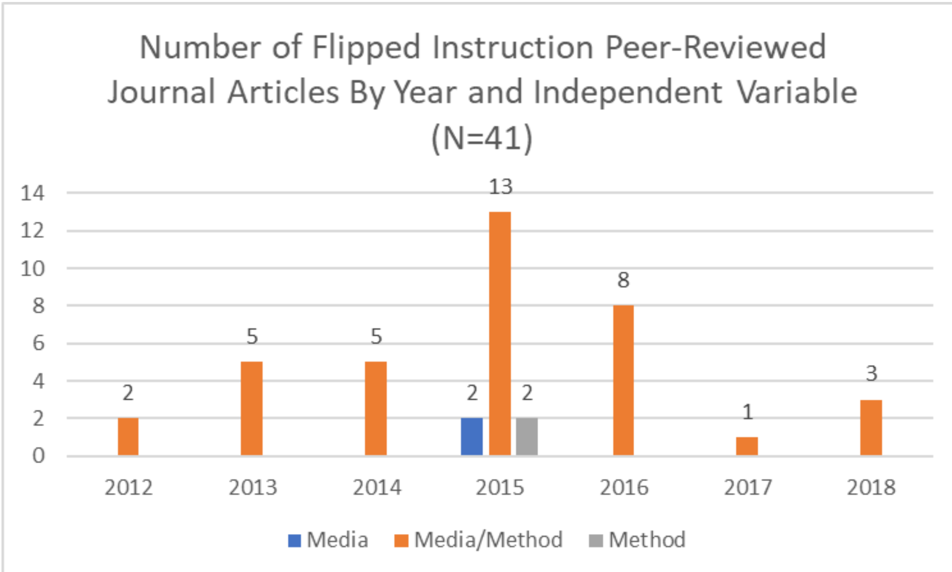

Figure 6: The number of peer-reviewed journal articles reporting data about flipped instruction peaked in 2015. The majority of the articles blended media and method as independent variables, suggesting significant confounding.

Table 3 presents a summary of primary statistics elicited from the research data. Sixty-three percent of the articles were from North America. Similar to the results in Study 1, none of the articles represented instructional objectives in a complete form that reflects Mager's (1984) classic elements (condition, behavior, criterion) or Czeropski and Pembroke's (2017) alternative agile, story-based elements. However, Hung (2015) did provide enough content to enable one to assemble it into a well-formed instructional objective if they so desired. Of the classic elements, behavior was most frequently specified (41%), distantly followed by conditions (10%) and criterion (5%).

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *Educational Technology Research & Development*, published by Springer. Copyright restrictions may apply. https://doi.org/ 10.1007/s11423-021-09988-1.

Insert Table 3 About Here

Table 3: Flipped Summary Statistics

Number of Studies by Regional Source Between 2009-2018	North America	26	63%
	Asia	5	12%
	Europe	4	10%
	Middle East	2	5%
	Oceania	2	5%
	Africa	1	2%
	South America	1	2%
Number of Studies by Discipline 2009-2018	Arts	1	2%
	Education	2	5%
	Engineering	4	10%
	Health	14	34%
	Mathematics	6	15%
	Natural Sciences	5	12%
	Social Science	9	22%
Number of Studies by Research Classification	To Prove For Descriptive Theory	0	0%
	To Prove For Design Theory	34	83%
	To Improve	3	7%
	To Describe	4	10%
	To Test Feasibility	0	0%
Number of Studies Including Instructional Objective Elements	Condition	4	10%
	Behavior	17	41%
	Criterion	2	5%
Number of Studies Including Instructional Outcome Measures	Effectiveness	35	85%
	Efficiency	14	34%
	Appeal	36	88%

--

Only 5% of articles described the researcher performing formative evaluation or pilot testing of their novel treatment. None of the articles described performing formative evaluation for any "traditional" treatment.

The inclusion of effectiveness (85%) and appeal (88%) measures was very high, yet there were several articles that only reported appeal measures for the flipped instruction treatment. Efficiency measures were much less frequent (34%) and focused on instructional and/or instructor time. A total of 12 articles (29%) reported data for all three measures.

Similar to Study 1, the researchers classified the majority of articles as *research to prove for design theory* (83%). *Research to describe* accounted for 10%, and *research to improve* accounted for 7%. No articles met the criteria for *research to prove for descriptive theory* or *research to test feasibility*.

Synthesis

In our review of the 80 papers that comprised Studies 1 and 2, we were very impressed by the innovation that is happening in the instructional design field. Researchers are working to figure out the right recipes that blend various instructional theories, models, and methods with computer-based instruction, videos, robots, mobile devices, computer games, tablet PCs, and mBots in complex contexts to create learning experiences that have the potential to be effective, efficient, and appealing. We want to see these innovations succeed, so they can enable the adoption characteristics of relative advantage, compatibility, simplicity, observability, and trialability that result in stakeholder adoption of these innovations (Rogers, 2003).

It is ironic that the barrier to generating useful knowledge about the innovations we reviewed in Studies 1 and 2 was the reliance on *research to prove* that was often confounded and lacked alignment with the instructional theory framework. Additionally, to answer our research questions, the elements of the instructional theory framework and Culture Four are generally absent from research that compares traditional instruction and novel instruction. The trend in ETR&D and non-ETR&D journals for research-to-prove-type articles appears to be increasing.

A total of 70 articles (86%) reflected *research to prove*. Only three articles (4%) reflected *research to improve*, and even those articles were difficult for us to classify as such. This represents a poor choice of research method. A better option is for a designer to consider the *maturity* of their learning experience (Figure 1) when choosing a research method (Branson, 1987; Phillips et al., 2012).

One hundred percent of the articles we reviewed lacked a well-formed designer objective that identifies all three parts – condition, behavior, and standard of performance. Without this information, a reviewer or reader cannot know whether 1) the learning experience aligns with the conditions, 2) the instructional methods align with the desired behavior or performance, and 3) learner performance as determined by an assessment meets a mastery standard. Research to improve requires well-formed instructional objectives and aligned assessments.

Seventy-seven-point-five percent (77.5%) of the articles in our sample did not collect data for all three outcomes: effectiveness, efficiency, and appeal. Honebein and Honebein (2015) consolidated the three instructional outcomes of effectiveness, efficiency, and appeal into a structure they call the *iron triangle of instructional design* (Figure 7). What the iron triangle represents are the ideas of sacrifice and trade-offs in design. For example, a designer may desire to have all three outcomes equally balanced and at high levels, as depicted in the equilateral triangle (left-hand diagram). Or an instructional designer may want or need to favor two of the outcomes over a third, resulting in the choice of methods that sacrifice one (or even two) of the outcomes. For example, the experiential approach (Lindsey & Berger, 2009) values effectiveness and appeal at the expense of efficiency (right-hand diagram). Why? Because experiential learning experiences are typically more costly to design and develop, and require more time and skill to deliver. If research is to accommodate "complex situations that [are] naturally and intentionally confounded" (Kozma, 2000, p. 10), then revealing the effects on all three outcomes is paramount for research to be called "good."

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *Educational Technology Research & Development*, published by Springer. Copyright restrictions may apply. https://doi.org/ 10.1007/s11423-021-09988-1.

Insert Figure 7 About Here

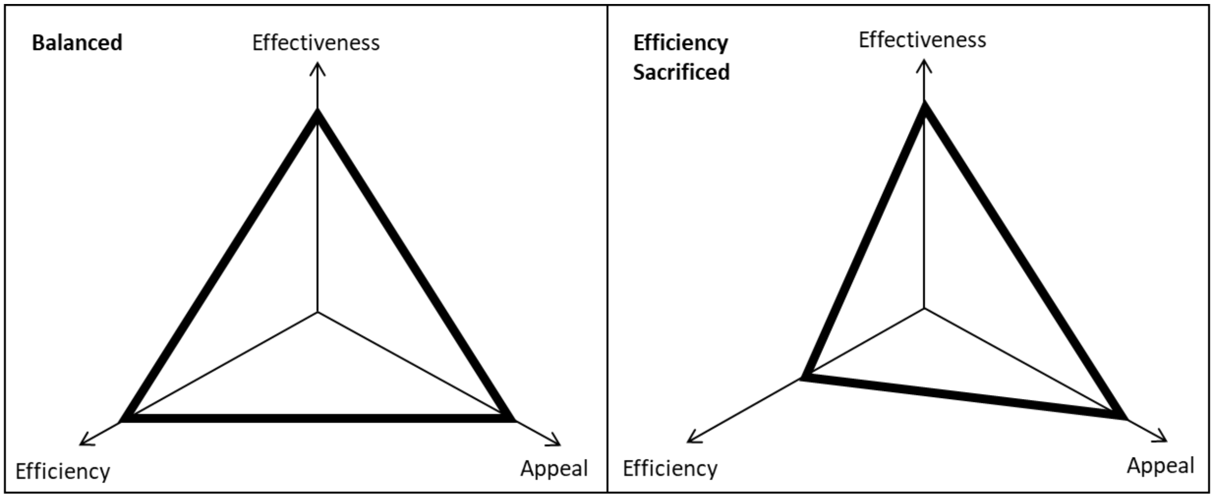

Figure 7: The Instructional Design Iron Triangle depicts the three outcomes (or constraints) associated with instructional methods: effectiveness, efficiency, and appeal. An instructional theory, model, or method typically involves the sacrifice of one or more of the outcomes.

Without data for effectiveness, efficiency, and appeal, it is difficult to know when an instructional medium or method is preferable compared to another, given that different priorities are valued by different stakeholders in different situations. This is a huge gap in our field's research practice, which was noted by one of our anonymous reviewers:

> Imagine if the 40 years of ETR&D papers reviewed by the authors had instead of trying to prove what works in IDT, had focused on solving important educational problems such as finding effective balances among effectiveness, efficiency, and appeal as the authors advise. We would be so much further along in providing effective, efficient, and appealing online learning opportunities to learners forced out of traditional instructional modalities by the current pandemic.

The majority of the research-to-prove articles confounded instructional methods and instructional media, making conclusions, prescriptions, and recommendations less useful in terms of their power to support adoption of the novel learning experience. The principles of good instructional research and good instructional design described above represent good design judgments (Boling et al., 2017; Demiral-Uzan, 2015; Nelson & Stolterman, 2012; Smith & Boling, 2009). Much of the research we reviewed lacks key pieces of information signaling that the researchers are using good design judgments that guide the advancement of knowledge about instruction. Thus, to better guide designers in the future, we suggest refinements to Briggs' (1984) Culture Four aspects. Kozma (2000) seeded some initial ideas for creating a *Culture Five* position, which we now significantly elaborate:

1. The researchers do accurately specify the desired learning outcomes based upon the conditions and values of the situation elicited from stakeholders, and they supply requirements and instructional objectives that include conditions, behaviors, and criteria, along with assessments that align with the situation.
2. The researchers describe their real learning experiences in detail, including improvements suggested by their data, made over time.
3. The researchers describe how they systematically designed learning experiences and formatively evaluated those learning experiences, using good design judgment, prior to conducting the research.
4. The researchers' tests and data really measure effectiveness, efficiency, and appeal.

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *Educational Technology Research & Development*, published by Springer. Copyright restrictions may apply. https://doi.org/ 10.1007/s11423-021-09988-1.

Perhaps an appropriate way to wrap-up this synthesis section is for us to put on our peer-review-hat. Peer reviewers serve authors in a variety of helpful ways. They serve as coaches and mentors rather than simply as judge and jury. Thus, in this spirit as we reflect upon the 80 articles we reviewed, we submit the following collective review comments for the majority of the articles we reviewed, which, we hope, will provide inspiration and direction to instructional designers, researchers, and reviewers in the future:

> We feel this collection of research, which compares traditional learning experiences to novel learning experiences, has five key issues that need to be resolved.
>
> 1. This research reflects a heavy dose of research to prove when research to improve is more appropriate. Correcting this involves eliminating the traditional learning experience from the study, and then collecting and analyzing improvement data associated with the novel learning experience. This likely will require revising research questions, research methods, and conclusions to better reflect improvements. One can then focus on advancing knowledge about improvements for particular conditions (the situation) and priorities about outcomes (effectiveness, efficiency, and appeal).
> 2. The novel learning experiences lack maturity because they are novel; hence they are too low on the S-curve to benefit from proof-oriented research. In fact, proof-oriented research can lead to the premature abandonment of a promising new learning experience that needs further development to move it up its S-curve. Therefore, based upon Kozma's (2000) view of modern instructional design research, the current state of knowledge embedded in your novel learning experience may be better served by situational *research to improve* (design-based research, formative research, etc.), which is more useful for advancing instructional design theory and methods. Please see Kakos-Kraft et al. (1997), English and Reigeluth (2006), Lee and Reigeluth (2003), and Lee, Jahnke, and Austin (2021) for examples of *research to improve*, where researchers encourage participants in the research, specifically learners, to suggest learning experience improvements.
> 3. It seems that most of this collection of research focuses primarily on the *effectiveness* of the new treatment, in terms of the pre- and/or post-test scores. The instructional theory framework (Honebein & Reigeluth, 2020, 2021; Reigeluth & Carr-Chellman, 2009) specifies three outcomes, effectiveness, efficiency, and appeal. These three outcomes provide a more robust view of the learning experience's preferability (Driscoll & Dick, 1999) and the nature of what the researchers had to sacrifice in their learning experience (Honebein & Honebein, 2015). If your research has not collected data for all three outcomes, then be sure to collect such data in the future. Alternatively, address this issue in the limitations section of your paper, explaining your reasons for omitting certain outcomes.
> 4. It was very difficult, if not impossible, to discern the primary instructional objective for the learning experiences reported in this collection of research. Briggs (1984) suggests that well-formed instructional objectives are a foundation for good instructional design and good research design. Time, however, seems to have faded this advice. In this age when our field is more aware of design judgment and core judgment (Boling et al., 2017; Stolterman & Nelson, 2000), we ask that the researchers synthesize their instructional situation and elicit from it well-formed instructional *designer* objectives (Reigeluth & An, 2021) that richly describe the condition, behavior, and criteria (Mager, 1984) of the overall learning experience as well as the priorities for instructional outcomes (effectiveness, efficiency, and appeal) for the situation. Note that any objectives you provide to your learners will be much simpler than the designer objectives in your research report. For *research to improve*, the mastery criteria you specify are very important, as they set the effectiveness standard for your learning experience, which you can compare with your dependent measures for effectiveness. Your instructional objectives can assume a variety of forms, from the Mager-style structure suggested above, to the format offered by Reigeluth and An (2021), to agile stories described by Czeropski and Pembroke (2017). You may find

additional inspiration by further exploring the instructional theory framework and the Culture Five prescriptions.

5. The studies introduce significant confounding variables involving the mixture of instructional methods and media methods. The instructional design field has vigorously debated these issues (see Clark, 1983, 1994; Kozma, 1994, 2000; Tennyson, 1994). By eliminating the comparison group (traditional learning experience) and focusing on research to improve, a researcher eliminates the problem of confounding variables. However, the researcher should comment on the extent to which the media methods used may or may not have influenced the outcomes (effectiveness, efficiency, and appeal of the instruction) – in other words, whether alternative media might as well be used.

Recommendations

Spector (2017) set a precedent for ETR&D, where “it should be the reviewers and not the editor who decides which manuscripts are published” and that ETR&D should “encourage more international contributions from outside North America” (p. 1416). As ETR&D peer reviewers, the authors can affirm that this principle appears to be the norm and, as we stated earlier, strongly support it, with one significant modification: that the editors assume more of a coaching role to improve the work of the reviewers. In addition to all the really good research that ETR&D and other journals publish, the two studies reported in this paper have identified that educational research is negatively impacted by

1. heavy use of research to prove when research to improve is more appropriate,
2. failure to conduct or report formative evaluation,
3. failure to address all three outcomes of effectiveness, efficiency, and appeal,
4. failure to include well-formed objectives, and
5. when research to prove is appropriate, inclusion of confounding variables.

This is not just a non-North American issue as the data in Study 1 suggests. Based upon the Study 2 data, it is also a North American issue. Editors, reviewers, and authors affiliated not only with ETR&D, but also with other instructional design and technology journals, should consider this situation, and join us to correct it in the future. New, innovative learning experiences benefit more from *research to improve*, paving the way toward more rapid improvement and adoption.

To develop and present our recommendations, we borrow a useful tool from our sister-science human performance technology, the behavioral engineering model (Chevalier, 2003; Gilbert, 1996; The Performance Thinking Network, 2012).

1. Expectations and Feedback

For Editors: In ETR&D’s *Instructions to Authors*, add a section that

- briefly describes the comparative research concerns (to prove) in our field,
- cautions authors about those comparative research concerns, and
- reinforces the inclusion of well-formed instructional objectives (conditions, behavior, standard of performance), evidence of formative evaluation, and information about all three outcomes (effectiveness, efficiency, appeal).

This section should also provide links to relevant ETR&D articles discussing the comparative study issue. And editors should coach and mentor their reviewers more.

For Authors: Read the revised *Instructions to Authors*.

For Reviewers: Be more aware of these concerns when asked to review a comparative research article. Advise or remind authors of the information in the *Instructions to Authors*. Suggest alternative research goals and methods that repurpose the study to make more valuable contributions to the field.

2. Tools and Resources

For Editors: Continue the policies described in Spector (2017) while monitoring/reporting the submission and publication frequencies of comparative research studies. Additionally, consider classifying comparative, research-to-prove studies that reflect S-curve maturity in ETR&D's research section, while classifying other comparative studies that lack S-curve maturity in ETR&D's development section.

3. Consequences and Incentives

For Authors: ETR&D editors and reviewers communicate to authors the increased risk of rejection for comparative studies.

For Reviewers: ETR&D editors continue to provide Outstanding Research Reviewer and Reviewer Excellence awards, which should consider the issues described in this article.

4. Motives and Preferences (Attitudes)

For Authors: If the local values associated with your country, your funding sources, and/or your department favor research to prove versus research to improve, you have two courses of action. First, add some research to improve into your research designs. Second, use the resources discussed in this article and others to influence your colleagues' thinking.

For Editors and Reviewers: Champion the publication of the best examples of research to improve.

5. Selection and Assignment (Capacity)

For Editors: When assigning comparative research submissions to reviewers, editors should either 1) include at least one reviewer who is very familiar with the disadvantages and issues of comparative research, or 2) provide information (the *Information for Authors* suggested above) and coaching to reviewers who are less familiar with the disadvantages and issues of comparative research.

6. Skills and Knowledge

For Authors, Reviewers, and Professors: Provide open educational resources (OERs) that teach the fundamentals of the instructional theory framework (for example, see https://edtechbooks.org/id/making_good_design (Honebein & Reigeluth, 2021)) and design/development research goals (research to prove descriptive theory, research to prove design theory, research to improve, research to describe, and research to test feasibility). Additionally, ETR&D could sponsor an educational session at the annual AECT conference to discuss these issues.

Limitations

This research only included peer-reviewed articles that mentioned "traditional" method, instruction, or approach. There are likely other peer-reviewed articles that compare learning experiences without referring to one of the treatments as "traditional." We expect those articles may have the same issues as the articles discussed in this research, depending upon their S-curve maturity. Additionally, there are likely many other articles that represent the four research methods other than research to prove, as our inclusion criteria favored research-to-prove articles. Thus, the percentages we report for type of research should not be attributed to the overall state of instructional technology research.

Given the large number of articles examined in this research, it is possible that the researchers missed some details useful for classifying each of the articles. For example, while it is relatively easy to identify if an article contains a well-formed Mager-style instructional objective containing all three elements of condition, behavior, and criterion, it is much harder to identify individual components that may be scattered throughout an article. Furthermore, a common cause of the researchers not initially agreeing on an article's research type classification was mismatches and inconsistencies between the article's research questions, research methods, and conclusions (which authors of future papers should avoid). Furthermore, while some missed details may be due to researcher fatigue, other missed details were due to non-standard or inconsistent article structures that hid details in places the researchers did not expect or used non-standard terminology to describe details.

The opinion we offer regarding the resurgence of comparative, research-to-prove studies is informed judgment on our part and was specifically requested by one of the reviewers of this paper. While we feel we have a reasonable understanding of the values and motives of colleagues in North America, we lack deeper understanding of the values and motives of colleagues in non-North American locations, which is why, before submitting this article, we requested that if reviewers accept this article, ETR&D editors would consider soliciting commentary and critique from colleagues, much like the Clark/Kozma debates in the 1990's. Our feeling is that factors such as complexity, context, design judgment, and culture may have a role in terms of researchers' choice of research methods. A subsequent study or discussion should investigate these factors.

References

Acker, S.R., Klein, E.L. Visualizing spatial tasks: A comparison of computer graphic and full-band video displays. ECTJ 34, 21–30 (1986). https://doi-org.libproxy.boisestate.edu/10.1007/BF02768359

Al-Samarraie, H., Shamsuddin, A., and Alzahrani, A. I. (2019, October). A flipped classroom model in higher education: a review of the evidence across disciplines. *Educational Technology Research and Development*. https://doi-org.proxyiub.uits.iu.edu/10.1007/s11423-019-09718-8

Al-Zahrani, A. M. (2015). From passive to active: The impact of the flipped classroom through social learning platforms on higher education students' creative thinking. *British Journal of Educational Technology, 46*(6), 1133–1148.

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences*, *13*(1), 1-14.

Belfi, L. M., Bartolotta, R. J., Giambrone, A. E., Davi, C., & Min, R. J. (2015). "Flipping" the introductory clerkship in radiology: Impact on medical student performance and perceptions. *Academic Radiology, 22*(6), 794–801.

Blair, E., Maharaj, C., & Primus, S. (2016). Performance and perception in the flipped classroom. *Education and Information Technologies, 21*(6), 1465–1482.

Boling, E. et al. (2017). Core judgments of instructional designers in practice. *Performance Improvement Quarterly, 30*(3), 119-219. DOI: 10.1002/piq.21250

Bonneau, C., Bourdeau, S. Computer-supported collaboration: simulation-based training using LEGO®. Education Tech Research Dev 67, 1507–1527 (2019). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-019-09689-w

Branson, R. K. (1987). Why the schools can't improve: The upper limit hypothesis. *Journal of Instructional Development*, *10*(4), 15-26.

Briggs, L. J. (1984). Trying to straddle four research cultures. *Educational Technology*, 24(8), 33-34. https://www.jstor.org/stable/44424176

Brooks, A. W. (2014). Information literacy and the flipped classroom: Examining the impact of a one-shot flipped class on student learning and perceptions. *Communications in Information Literacy, 8*(2), 225–235.

Burbach, M. E., Matkin, G. S., & Fritz, S. M. (2004). Teaching critical thinking in an introductory leadership course utilizing active learning strategies: A confirmatory study. *College Student Journal*, *38*(3), 482–493.

Cabı, E. (2018). The impact of the flipped classroom model on students' academic achievement. *The International Review of Research in Open and Distributed Learning, 19*(3), 202–221.

Chang, CY., Kao, CH., Hwang, GJ. et al. From experiencing to critical thinking: a contextual game-based learning approach to improving nursing students' performance in Electrocardiogram training. Education Tech Research Dev 68, 1225–1245 (2019). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-019-09723-x

Chevalier, R. (2003). Updating the behavior engineering model. *Performance Improvement*, *42*(5), 9-14.

Chien, C.-F., & Hsieh, L.-H. C. (2018). Exploring university students' achievement, motivation, and receptivity of flipped learning in an engineering mathematics course. *International Journal of Online Pedagogy and Course Design (IJOPCD), 8*(4), 22–37.

Choi, J.I.,& Hannafin, M. The effects of instructional context and reasoning complexity on mathematics problem-solving. ETR&D 45, 43–55 (1997). https://doi-org.libproxy.boisestate.edu/10.1007/BF02299728

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445-459.

Clark, R. E. (1985). Evidence for confounding in computer-based instruction studies: Analyzing the meta-analyses. *Educational Communications and Technology Journal, 33*(4), 249-262.

Clark, R. E. (1986). Absolutes and angst in educational technology research: A reply to Don Cunningham. *Educational Communications and Technology Journal, 34*(1), 8-10.

Clark, R. E. (1989). Current progress and future directions for research in instructional technology. *Educational Technology Research and Development, 37*(1), 57-66.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*(2), 21-29.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9-13. https://www.jstor.org/stable/3699928

Cobb, T. (1997). Cognitive efficiency: Toward a revised theory of media. *Educational Technology Research and Development, 45*(4), 21-35.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences*, *13*(1), 15-42. https://doi.org/10.1207/s15327809jls1301_2

Cunningham, D. J. (1986). Good guys and bad guys. *Educational Communications and Technology Journal, 34*(1), 3-7.

Czeropski, S., & Pembroke, C. (2017). E-learning ain't performance: Revising HPT in an era of agile and lean. *Performance Improvement*, 56(8), 37-47. 10.1002/pfi.21728

Danielson, J.A., Bender, H.S., Mills, E.M. et al. A tool for helping veterinary students learn diagnostic problem solving. ETR&D 51, 63–81 (2003). https://doi-org.libproxy.boisestate.edu/10.1007/BF02504553

Danker, B. (2015). Using flipped classroom approach to explore deep learning in large classrooms. *IAFOR Journal of Education*, *3*(1), 171–186.

Davies, R. S., Dean, D. L., & Ball, N. (2013). Flipping the classroom and instructional technology integration in a college-level information systems spreadsheet course. *Educational Technology Research and Development, 61*(4), 563–580.

Demiral-Uzan, M. (2015). Instructional design students' design judgment in action. *Performance Improvement Quarterly*, 28(3), 7-23.

Design-Based Research Collective (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, *32*(1), 5-8.

Driscoll, M. P., & Dick, W. (1999). New research paradigms in instructional technology: An inquiry. *Educational Technology Research and Development, 47*(2), 7-18.

Efron, S. E., & Ravid, R. (2020). *Action research in education: A practical guide* (2nd ed.). New York, NY: The Guilford Press.

Efstathiou, C., Hovardas, T., Xenofontos, N.A. et al. Providing guidance in virtual lab experimentation: the case of an experiment design tool. Education Tech Research Dev 66, 767–791 (2018). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-018-9576-z

Eftekhari, M., Sotoudehnama, E. & Marandi, S.S. Computer-aided argument mapping in an EFL setting: does technology precede traditional paper and pencil approach in developing critical thinking?. Education Tech Research Dev 64, 339–357 (2016). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-016-9431-z

English, R. E., & Reigeluth, C. M. (2003). Formative research on sequencing instruction with the elaboration theory. *Educational Technology Research and Development, 44*, 23–42. https://doi.org/10.1007/BF02300324

Farley, F. H. (1982). The future of educational research. *Educational Researcher, 11*(8), 11-19.

Fabian, K., Topping, K.J. & Barron, I.G. Using mobile technologies for mathematics: effects on student attitudes and achievement. Education Tech Research Dev 66, 1119–1139 (2018). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-018-9580-3

Fiorella, L., Vogel-Walcutt, J.J. & Schatz, S. Applying the modality principle to real-time feedback and the acquisition of higher-order cognitive skills. Education Tech Research Dev 60, 223–238 (2012). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-011-9218-1

Foldnes, N. (2016). The flipped classroom and cooperative learning: Evidence from a randomised experiment. *Active Learning in Higher Education, 17*(1), 39–49.

Fraga, L. M., & Harmon, J. (2014). The flipped classroom model of learning in higher education: An investigation of preservice teachers' perspectives and achievement. *Journal of Digital Learning in Teacher Education, 31*(1), 18–27.

Gibbons, A. S., & Rogers, P. C. (2009). The architecture of instructional theory. In C. M. Reigeluth & A. Carr-Chellman (Eds.), *Instructional-design theories and models: Building a common knowledge base (Vol. III)* (pp. 305-326). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gilbert, T. F. (1996). *Human competence: Engineering worthy performance*. HRD Press.

Gross, D., Pietri, E. S., Anderson, G., Moyano-Camihort, K., & Graham, M. J. (2015). Increased preclass preparation underlies student outcome improvement in the flipped classroom. *CBE-Life Sciences Education*. https ://doi.org/10.1187/cbe.15-02-0040.

Han, J., Jo, M., Hyun, E. *et al.* Examining young children's perception toward augmented reality-infused dramatic play. *Education Tech Research Dev* **63,** 455–474 (2015). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-015-9374-9

Hancock-Niemic, M.A., Lin, L., Atkinson, R.K. et al. Example-based learning: exploring the use of matrices and problem variability. Education Tech Research Dev 64, 115–136 (2016). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-015-9403-8

Harrington, S. A., Bosch, M. V., Schoofs, N., Beel-Bates, C., & Anderson, K. (2015). Quantitative outcomes for nursing students in a flipped classroom. *Nursing Education Perspectives, 36(3),* 179–181.

Honebein, P. C. (1994). The effects of a problem-based learning curriculum for diabetes management and care in a large medical school (Doctoral dissertation). Retrieved from https://search-proquest-com.proxyiub.uits.iu.edu/pqdtglobal/docview/304102555/fulltextPDF/784056FA4C0F49F2PQ/2?accountid=11620

Honebein, P. C. (2018). Specifying human performance solutions through well-formed business requirements. *Performance Improvement, 57*(7), 45-49. DOI: 10.1002/pfi.21788

Honebein, P. C., & Honebein, C. H. (2015). Effectiveness, efficiency, and appeal: pick any two? The influence of learning domains and learning outcomes on designer judgments of useful instructional methods. *Educational Technology Research and Development, 63*(6), 937-955. https://doi.org/10.1007/s11423-015-9396-3

Honebein, P. C., & Reigeluth, C. M. (2020). The instructional theory framework appears lost. Isn't it time we find it again? *Revista de Educación a Distancia, 64*(20). https://revistas.um.es/red/article/view/405871/290451

Honebein, P. & Reigeluth, C. M. (2021). Making good design judgments via the instructional theory framework. In J. K. McDonald & R. E. West (Eds.), *Design for Learning: Principles, Processes, and Praxis*. EdTech Books. https://edtechbooks.org/id/making_good_design

Hotle, S. L., & Garrow, L. A. (2015). Effects of the traditional and flipped classrooms on undergraduate student opinions and success. *Journal of Professional Issues in Engineering Education and Practice, 142*(1), 05015005.

Huang, YM., Huang, YM. A scaffolding strategy to develop handheld sensor-based vocabulary games for improving students' learning motivation and performance. Education Tech Research Dev 63, 691–708 (2015). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-015-9382-9

Huang, YM., Shadiev, R., Sun, A. et al. A study of the cognitive diffusion model: facilitating students' high level cognitive processes with authentic support. Education Tech Research Dev 65, 505–531 (2017). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-016-9475-0

Hung, H.-T. (2015). Flipping the classroom for English language learners to foster active learning. *Computer Assisted Language Learning, 28*(1), 81–96.

Hwang, GJ., Chang, SC., Chen, PY. et al. Effects of integrating an active learning-promoting mechanism into location-based real-world learning environments on students' learning performances and behaviors. Education Tech Research Dev 66, 451–474 (2018). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-017-9567-5

Hwang, GJ., Hung, CM. & Chen, NS. Improving learning achievements, motivations and problem-solving skills through a peer assessment-based game development approach. *Education Tech Research Dev* **62,** 129–145 (2014). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-013-9320-7

Hwang, WY., Zhao, L., Shadiev, R. et al. Exploring the effects of ubiquitous geometry learning in real situations. Education Tech Research Dev 68, 1121–1147 (2019). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-019-09730-y

Jayasinghe, M.G., Morrison, G.R. & Ross, S.M. The effect of distance learning classroom design on student perceptions. ETR&D 45, 5–19 (1997). https://doi-org.libproxy.boisestate.edu/10.1007/BF02299680

Jensen, J. L., Kummer, T. A., & Godoy, P. D. M. (2015). Improvements from a flipped classroom may simply be the fruits of active learning. *CBE-Life Sciences Education*. https ://doi.org/10.1187/cbe.14-08-0129.

Jonassen, D.H., Campbell, J.P. & Davidson, M.E. (1994). Learning with media: Restructuring the debate. *Educational Technology Research and Development, 42*(2), 31–39. https://doi-org.proxyiub.uits.iu.edu/10.1007/BF02299089

Jungić, V., Kaur, H., Mulholland, J., & Xin, C. (2015). On flipping the classroom in large first year calculus courses. *International Journal of Mathematical Education in Science and Technology, 46*(4), 508–520.

Kakos-Kraft, S., Honebein, P. C., Prince, M. J., & Marrero, D. G. (1997). The SOCRATES curriculum: An innovative integration of technology and theory in medical education. *Journal of Audiovisual Media in Medicine, 20*(4), 166-171. https://doi.org/10.3109/17453059709063100

Ke, F. Computer games application within alternative classroom goal structures: cognitive, metacognitive, and affective evaluation. Education Tech Research Dev 56, 539–556 (2008). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-008-9086-5

Koo, C. L., Demps, E. L., Farris, C., Bowman, J. D., Panahi, L., & Boyle, P. (2016). Impact of flipped classroom design on student performance and perceptions in a pharmacotherapy course. *American Journal Of Pharmaceutical Education, 80*(2), 33.

Kozma, R.B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research & Development, 42*, 7–19. https://doi-org.proxyiub.uits.iu.edu/10.1007/BF02299087

Kozma, R. (2000). Reflections on the state of educational technology research and development. *Educational Technology Research and Development, 48*(1), 5-15.

Kuo, M. L. A., & Hooper, S. (2004). The effects of visual and verbal coding mnemonics on learning Chinese characters in computer-based instruction. *Educational Technology Research and Development, 52*(3), 23-38.

Lan, YJ., Fang, SY., Legault, J. et al. Second language acquisition of Mandarin Chinese vocabulary: context of learning effects. Education Tech Research Dev 63, 671–690 (2015). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-015-9380-y

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. http://www.jstor.org/stable/2529310

Lee, G., & Wallace, A. (2018). Flipped learning in the english as a foreign language classroom: Outcomes and perceptions. *TESOL Quarterly, 52*(1), 62–84.

Lee, J., & Reigeluth, C. M. (2006). Formative research on the heuristic task analysis process. *Educational Technology Research and Development, 51*, 5–17. https://doi.org/10.1007/BF02504541

Lee, Y. M., Jahnke, I. & Austin, L. (2021) Mobile microlearning design and effects on learning efficacy and learner experience. *Educational Technology Research and Development*. https://doi-org.libproxy.boisestate.edu/10.1007/s11423-020-09931-w

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

LeMahieu, P. G., Edwards, A. R., & Gomez, L. M. (2015). At the nexus of improvement science and teaching: Introduction to a special section of the Journal of Teacher Education. *Journal of Teacher Education*, *66*(5), 446-449.

Lewis, C. (2015). What is improvement science? Do we need it in education? *Educational Researcher*, *44*(1), 54-61.

Liebert, C. A., Lin, D. T., Mazer, L. M., Bereknyei, S., & Lau, J. N. (2016). Effectiveness of the surgery core clerkship flipped classroom: A prospective cohort trial. *The American Journal of Surgery, 211*(2), 451.e1–457.e1.

Liefeld, J.P., Herrmann, T.F. Learning consequences for university students using computerized mastery testing. ETR&D 38, 19–25 (1990). https://doi-org.libproxy.boisestate.edu/10.1007/BF02298266

Lindsey, L., & Berger, N. (2009). Experiential approach to instructions. In C. M. Reigeluth & A. Carr-Chellman (Eds.), *Instructional-design theories and models: Building a common knowledge base (Vol. III) (pp. 117-142)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lin-Siegler, X., Shaenfield, D. & Elder, A.D. Contrasting case instruction can improve self-assessment of writing. Education Tech Research Dev 63, 517–537 (2015). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-015-9390-9

Liou, WK., Bhagat, K.K. & Chang, CY. The design, implementation, and evaluation of a digital interactive globe system integrated into an Earth Science course. Education Tech Research Dev 66, 545–561 (2018). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-018-9573-2

Lockee, B. B., Burton, J. K., & Cross, L. H. (1999). No comparison: Distance education finds a new use for 'no significant difference'. *Educational Technology Research and Development, 47*(3), 33-42.

Lubin, I. A., & Ge, X. (2012). Investigating the influences of a LEAPS model on preservice teachers' problem solving, metacognition, and motivation in an educational technology course. *Educational Technology Research and Development, 60*, 239-270. https://doi-org.proxyiub.uits.iu.edu/10.1007/s11423-011-9224-3

Mager, R. F. (1984). *Preparing instructional objectives*. Belmont, CA: Lake Publishing.

Malinverni, L., Schaper, MM. & Pares, N. An evaluation-driven design approach to develop learning environments based on full-body interaction. Education Tech Research Dev 64, 1337–1360 (2016). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-016-9468-z

Mason, G. S., Shuman, T. R., & Cook, K. E. (2013). Comparing the effectiveness of an inverted classroom to a traditional classroom in an upper-division engineering course. *IEEE Transactions on Education, 56*(4), 430–435.

Mattis, K. V. (2015). Flipped classroom versus traditional textbook instruction: Assessing accuracy and mental effort at different levels of mathematical complexity. *Technology, Knowledge and Learning, 20*(2), 231–248.

Mavridis, A., Katmada, A. & Tsiatsos, T. Impact of online flexible games on students' attitude towards mathematics. Education Tech Research Dev 65, 1451–1470 (2017). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-017-9522-5

McCallum, S., Schultz, J., Sellke, K., & Spartz, J. (2015). An examination of the flipped classroom approach on college student academic involvement. *International Journal of Teaching and Learning in Higher Education, 27*(1), 42–55.

McLaughlin, J. E., Griffin, L. M., Esserman, D. A., Davidson, C. A., Glatt, D. M., Roth, M. T., & Mumper, R. J. (2013). Pharmacy student engagement, performance, and perception in a flipped satellite classroom. *American Journal of Pharmaceutical Education, 77*(9), 196.

McLaughlin, J. E., Roth, M. T., Glatt, D. M., Gharkholonarehe, N., Davidson, C. A., Griffin, L. M., & Mumper, R. J. (2014). The flipped classroom: A course redesign to foster learning and engagement in a health professions school. *Academic Medicine, 89*(2), 236–243.

Merrill, M. D. (1983). Component display theory. In C.M. Reigeluth (Ed.), *Instructional-design theories and models: An overview of their current status (pp. 279-333)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Missildine, K., Fountain, R., Summers, L., & Gosselin, K. (2013). Flipping the classroom to improve student performance and satisfaction. *Journal of Nursing Education, 52*(10), 597–599.

Moffett, J., & Mill, A. C. (2014). Evaluation of the flipped classroom approach in a veterinary professional skills course. *Advances in Medical Education and Practice, 5*, 415.

Morrison, G.R. (1994). The media effects question: "Unresolvable" or asking the right question. *Educational Technology Research and Development, 42*, 41–44 (1994). https://doi-org.proxyiub.uits.iu.edu/10.1007/BF02299090

Morrison, G.R., Ross, S.M. & O'Dell, J.K. Text density level as a design variable in instructional displays. ECTJ 36, 103–115 (1988). https://doi-org.libproxy.boisestate.edu/10.1007/BF02766618

Murray, L., McCallum, C., & Petrosino, C. (2014). Flipping the classroom experience: A comparison of online learning to traditional lecture. *Journal of Physical Therapy Education, 28*(3), 35–41.

Nelson, H. G., & Stolterman, E. (2012). *The design way: Intentional change in an unpredictable world (2nd ed.)*. Cambridge, MA: MIT Press.

Ojennus, D. D. (2016). Assessment of learning gains in a flipped biochemistry classroom. *Biochemistry and Molecular Biology Education, 44*(1), 20–27.

Parkhurst, P. E. (1992). Criticisms of a report on teaching with interactive videodiscs. *Academic Medicine, 67*(10), 657-658.

Peterson, D. J. (2016). The flipped classroom improves student achievement and course satisfaction in a statistics course: A quasi-experimental study. *Teaching of Psychology, 43*(1), 10–15.

Phillips, R., Kennedy, G., & McNaught, C. (2012). The role of theory in learning technology and evaluation research. *Australasian Journal of Educational Technology, 28*(7), 1103-1118.

Pierce, R., & Fox, J. (2012). Vodcasts and active-learning exercises in a "flipped classroom" model of a renal pharmacotherapy module. *American Journal of Pharmaceutical Education, 76*(10), 196

Ponce, H.R., Mayer, R.E. & Lopez, M.J. A computer-based spatial learning strategy approach that improves reading comprehension and writing. Education Tech Research Dev 61, 819–840 (2013). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-013-9310-9

Porcaro, P. A., Jackson, D. E., McLaughlin, P. M., & O'Malley, C. J. (2016). Curriculum design of a flipped classroom to enhance haematology learning. *Journal of Science Education and Technology, 25*(3), 345–357.

Prashar, A. (2015). Assessing the flipped classroom in operations management: A pilot study. *Journal of Education for Business, 90*(3), 126–138.

Proske, A., Roscoe, R.D. & McNamara, D.S. Game-based practice versus traditional practice in computer-based writing strategy training: effects on motivation and achievement. Education Tech Research Dev 62, 481–505 (2014). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-014-9349-2

Raja, F. U., & Najmonnisa (2018). Comparing traditional teaching method and experiential teaching method using experimental research. *Journal of Education and Educational Development, 5*(2), 276-288.

Reeves, T. C., & Oh, E. (2017). The goals and methods of educational technology research over a quarter century (1989-2014). *Educational Technology Research and Development, 65*(2), 325-339

Reigeluth, C.M. (1983). Instructional Design: What is it and Why is it? In C.M. Reigeluth (Ed.), *Instructional-design theories and models: An overview of their current status (pp. 3-36)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Reigeluth, C. M. (1989). Educational technology at the crossroads: New mindsets and new directions. *Educational Technology Research and Development, 37*(1), 67-80.

Reigeluth, C.M. (1999). What is instructional-design theory and how is it changing? In C.M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory, volume II (pp. 5-29)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Reigeluth, C.M. & An, Y. (2009). Theory building. In C. M. Reigeluth & A. Carr-Chellman (Eds.), *Instructional-design theories and models: Building a common knowledge base (Vol. III) (pp. 365-386)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Reigeluth, C. M., & An, Y. (2021). *Merging the instructional design process with learner-centered theory: The holistic 4D model*. New York: Routledge.

Reigeluth, C.M. & Carr-Chellman, A. (2009). Understanding instructional theory. In C. M. Reigeluth & A. Carr-Chellman (Eds.), *Instructional-design theories and models: Building a common knowledge base (Vol. III) (pp. 3-26)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Reigeluth, C. M., & Frick, T. W. (1999). Formative research: A methodology for creating and improving design theories. In C.M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory, volume II (pp. 633-651)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Reigeluth, C. M., & Schwartz, E. (1989). An instructional theory for the design of computer-based simulations. *Journal of Computer-Based Instruction, 16*(1), 1–10

Reiser, R.A. (1994). Clark's invitation to the dance: An instructional designer's response. *Educational Technology Research and Development, 42*, 45–48. https://doi-org.proxyiub.uits.iu.edu/10.1007/BF02299091

Richey, R. (1998). The pursuit of useable knowledge in instructional technology. *Educational Technology Research and Development, 46*(4), 7-22.

Ritzhaupt, A., & Sommer, M. (2018). Impact of the flipped classroom on learner achievement and satisfaction in an undergraduate technology literacy course. *Journal of Information Technology Education: Research, 17*(1), 159–182.

Rogers, E. M. (2003). *Diffusion of Innovations (5th Ed.)*. New York: Free Press.

Ronimus, M., Eklund, K., Pesu, L. et al. Supporting struggling readers with digital game-based learning. Education Tech Research Dev 67, 639–663 (2019). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-019-09658-3

Ross, S.M. (1994). Delivery trucks or groceries? More food for thought on whether media (will, may, can't) influence learning. *Educational Technology Research and Development, 42*, 5–6. https://doi-org.proxyiub.uits.iu.edu/10.1007/BF02299086

Ross, S.M., Anand, P.G. A computer-based strategy for personalizing verbal problems in teaching mathematics. ECTJ 35, 151–162 (1987). https://doi-org.libproxy.boisestate.edu/10.1007/BF02793843

Ryan, M. D., & Reid, S. A. (2015). Impact of the flipped classroom on student performance and retention: A parallel controlled study in general chemistry. *Journal of Chemical Education, 93*(1), 13–23.

Sáez-López, JM., Sevillano-García, ML. & Vazquez-Cano, E. The effect of programming on primary school students' mathematical and scientific understanding: educational use of mBot. Education Tech Research Dev 67, 1405–1425 (2019). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-019-09648-5

Salomon, G. (1991). Transcending the qualitative-quantitative debate: The analytic and systemic approaches to educational research. *Educational Researcher, 20*(6), 10-18

Schweickert, R., & Boruff, B. (1986). Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(3), 419–425. https://doi.org/10.1037/0278-7393.12.3.419

Shadiev, R., Hwang, WY. & Liu, TY. Investigating the effectiveness of a learning activity supported by a mobile multimedia learning system to enhance autonomous EFL learning in authentic contexts. Education Tech Research Dev 66, 893–912 (2018). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-018-9590-1

Shrock, S.A. (1994). The media influence debate: Read the fine print, but don't lose sight of the big picture. *Educational Technology Research and Development, 42*, 49–53. https://doi-org.proxyiub.uits.iu.edu/10.1007/BF02299092

Simpson, V., & Richards, E. (2015). Flipping the classroom to teach population health: Increasing the relevance. *Nurse Education in Practice, 15*(3), 162–167.
Smith, K. M., & Boling, E. (2009). What do we make of design? Design as a concept in educational technology. *Educational Technology, 49*(4), 3-17.
Sohrabi, B., & Iraj, H. (2016). Implementing flipped classroom using digital media: A comparison of two demographically different groups perceptions. *Computers in Human Behavior, 60*, 514–524.
Spector, J. M. (2017). Reflections on educational technology research and development. *Educational Technology Research and Development*, *66*(4), 893-912.
Staker, H., & Horn, M. B. (2012). Classifying K-12 blended learning. Retrieved from https://www.christenseninstitute.org/wp-content/uploads/2013/04/Classifying-K-12-blended-learning.pdf
Stebbins, R. A. (2001). *Exploratory Research in the Social Sciences*. Sage: Thousand Oaks, CA.
Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Washington, DC: The Brookings Institution.
Stolterman, E., & Nelson, H. (2000). The guarantor of design. Proceedings of IRIS 23. Laboratorium for Interaction Technology, University of Trollhättan Uddevalla, 2000. L. Svensson, U. Snis, C. Sørensen, H. Fägerlind, T. Lindroth, M. Magnusson, C. Östlund (eds.)
Strayer, J. F. (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. *Learning Environments Research, 15*(2), 171–193.
Stringer, E. T. (2008). *Action research in education* (2nd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
Stringer, E. T., & Aragon, A. O. (2021). *Action research* (5th ed.). Thousand Oaks, CA: Sage Publications.
Tanner, D. (1998). The social consequences of bad research. *Phi Delta Kappan, 79*(5), 345-349.
Tanner, M., & Scott, E. (2015). A flipped classroom approach to teaching systems analysis, design and implementation. *Journal of Information Technology Education: Research, 14*, 219–241.
Tennyson, R. D. (1994). The big wrench vs. integrated approaches: The great media debate. *Educational Technology Research and Development*, *42*(3) 15-28.
Tessmer, M., Driscoll, M.P. Effects of a diagrammatic display of coordinate concept definitions on concept classification performance. ECTJ 34, 195–205 (1986). https://doi-org.libproxy.boisestate.edu/10.1007/BF02767401
The Performance Thinking Network (2012). Six boxes model. https://www.sixboxes.com/Six-Boxes-Model.html
Tune, J. D., Sturek, M., & Basile, D. P. (2013). Flipped classroom model improves graduate student performance in cardiovascular, respiratory, and renal physiology. *Advances in Physiology Education, 37*(4), 316–320.
Ullmer, E. J. (1994). Media and learning: Are there two kinds of truth? *Educational Technology Research and Development, 42*(1), 21-32.
Uribe, D., Klein, J.D. & Sullivan, H. The effect of computer-mediated collaborative learning on solving III-defined problems. ETR&D 51, 5–19 (2003). https://doi-org.libproxy.boisestate.edu/10.1007/BF02504514
Van Eck, R., & Dempsey, J. (2002). The effect of competition and contextualized advisement on the transfer of mathematics skills a computer-based instructional simulation game. *Educational Technology Research and Development, 50*, 23–41 (2002). https://doi-org.proxyiub.uits.iu.edu/10.1007/BF02505023
Van Vliet, E., Winnips, J., & Brouwer, N. (2015). Flipped-class pedagogy enhances student metacognition and collaborative-learning strategies in higher education but effect does not persist. *CBE-Life Sciences Education*. https ://doi.org/10.1187/cbe.14-09-0141.
Velegol, S. B., Zappe, S. E., & Mahoney, E. (2015). The evolution of a flipped classroom: Evidence-based recommendations. *Advances in Engineering Education, 4*(3), 1–37.
Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development, 53,* 5–23. https://doi.org/10.1007/BF02504682
Whillier, S., & Lystad, R. P. (2015). No differences in grades or level of satisfaction in a flipped classroom for neuroanatomy. *Journal of Chiropractic Education, 29*(2), 127–133.
Wong, T. H., Ip, E. J., Lopes, I., & Rajagopalan, V. (2014). Pharmacy students' performance and perceptions in a flipped teaching pilot on cardiac arrhythmias. *American Journal of Pharmaceutical Education, 78*(10), 185.
Wilson, S. G. (2013). The flipped class: A method to address the challenges of an undergraduate statistics course. *Teaching of Psychology, 40*(3), 193–199.
Yeh, YL., Lan, YJ. Fostering student autonomy in English learning through creations in a 3D virtual world. Education Tech Research Dev 66, 693–708 (2018). https://doi-org.libproxy.boisestate.edu/10.1007/s11423-017-9566-6

This is an author-produced, peer-reviewed version of this article. The final, definitive version of this document can be found online at *Educational Technology Research & Development*, published by Springer. Copyright restrictions may apply. https://doi.org/ 10.1007/s11423-021-09988-1.

Appendix A

Coding categories for ETR&D articles between 1980 and 2019

Element/Column	Criteria
Order Number	Sequential order in which the article was collected and reviewed
Date	Year of publication
Document	File name of the article's PDF
Region	Region where the study was conducted or where first author is based.
Subjects	Description of the subjects for the study
Research Questions/ Hypotheses	Indicates presence of research Questions, Hypothesis, Objective, Derived, or some mixture of these four concepts. Cell Comment provides data from the paper. "Derived" means that no specific research questions or hypotheses were specified, but could be derived from statements made in the manuscript.
Classification	Classifies studies based upon the available data: 1 = Research to prove – for descriptive theory 2 = Research to prove – for design theory 3 = Research to improve 4 = Research to describe 5 = Research to test feasibility
Instructional Theory/Model	The type instructional theory/model described in the literature section of the paper (macro, meso, micro).
Traditional Method	The primary instructional method employed by the perceived inferior solution
Traditional Media	The primary media method employed by the perceived inferior solution
New Method	The primary instructional method employed by the perceived superior solution
New Media	The primary media method employed by the perceived superior solution
Independent Variable	The primary focus of the independent variable: Method, Media, or Mixed
Other Method	Other instructional methods received by all participants
Treatments	Number of treatments in the study.
Research Method	As specified by author or implied by researcher: experimental, quasi-experimental, case study, etc.
Task	Description of the primary task of the learner.
Conditions	Yes or No depending upon whether the author provided at least one condition, signaled by words such as "given," "using," or similar.
Behavior	Yes or No depending upon if the author specified at least one logical statement that includes an observable action verb followed by a behavioral statement.
Criterion Specification	Yes or No or Partial depending upon a specification of a quantified, target criterion for at least one of effectiveness, efficiency, or appeal.
Formative Evaluation	Yes or No or Partial depending upon whether an instructional design was pilot tested prior to the running of the experiment. Search terms: formative, pilot.
Effectiveness	Yes or No for specific measure present in the Method section for assessing the effectiveness of the instructional designs.
Efficiency	Yes or No for specific measure present in the Method section for assessing the efficiency of the instructional designs.
Appeal	Yes or No for specific measure present in the Method section for assessing appeal (liking, satisfaction, motivation) of the instructional designs.
Other Measures	Other measures beyond effectiveness, efficiency, and appeal

Result	Identification of the treatment that performed “best”
Source	Link to the research paper
Title	Title of the research paper

Appendix B

Study 1 Articles – From ETR&D 1980-2019

Author	Region	Research Type
Tessmer & Driscoll (1986)	North America	Research to Prove for Design
Acker & Klein (1986)	North America	Research to Prove for Design
Ross & Anand (1987)	North America	Research to Prove for Design
Morrison et al. (1988)	North America	Research to Prove for Design
Liefeld & Herrmann (1990)	North America	Research to Prove for Design
Jayasinghe et al. (1997)	North America	Research to Prove for Design
Choi & Hannafin (1997)	Asia	Research to Prove for Design
Van Eck & Dempsey (2002)	North America	Research to Prove for Design
Uribe et al. (2003)	North America	Research to Prove for Design
Danielson et al. (2003)	North America	Research to Prove for Design
Kuo & Hooper (2004)	North America	Research to Prove for Design
Ke (2008)	North America	Research to Prove for Design
Lee & Thomas (2011)	North America	Research to Test Feasibility
Lubin & Ge (2012)	North America	Research to Describe
Fiorella et al. (2012)	North America	Research to Prove for Design
Ponce et al. (2013)	South America	Research to Prove for Design
Davies et al. (2013)	North America	Research to Prove for Design
Hwang et al. (2014)	Asia	Research to Prove for Design
Proske et al. (2014)	Europe	Research to Prove for Design
Lin-Siegler et al. (2015)	North America	Research to Prove for Design
Huang & Huang (2015)	Asia	Research to Prove for Design
Lan et al. (2015)	Asia	Research to Prove for Design
Han et al. (2015)	Asia	Research to Prove for Design
Malinverni et al. (2016)	Europe	Research to Prove for Design
Eftekhari et al. (2016)	Middle East	Research to Prove for Design
Hancock et al. (2016)	North America	Research to Prove for Design
Hwang et al. (2018)	Asia	Research to Prove for Design
Mavridis et al. (2017)	Europe	Research to Prove for Design
Huang et al. (2017)	Asia	Research to Prove for Design
Fabian et al. (2018)	Europe	Research to Prove for Design
Shadiev et al. (2018)	Asia	Research to Prove for Design
Efstathiou et al. (2018)	Europe	Research to Prove for Design
Liou et al. (2018)	Asia	Research to Prove for Design
Yeh & Lan (2018)	Asia	Research to Describe
Chang et al. (2019)	Asia	Research to Prove for Design
Ronimus et al. (2019)	Europe	Research to Prove for Design
Hwang et al. (2019)	Asia	Research to Prove for Design
Sáez-López et al. (2019)	Europe	Research to Prove for Design
Bonneau & Bourdeau (2019)	North America	Research to Test Feasibility

Appendix C

Study 2 Articles – See Al-Samarraie et al. (2019)

Author	Region	Discipline	Research Type
Pierce and Fox (2012)	North America	Health	Research to Prove for Design
Strayer (2012)	North America	Mathematics	Research to Prove for Design
Tune et al. (2013)	North America	Health	Research to Prove for Design
Wilson (2013)	North America	Mathematics	Research to Prove for Design
McLaughlin et al. (2013)	North America	Health	Research to Prove for Design
Missildine et al. (2013)	North America	Health	Research to Prove for Design
Mason et al. (2013)	North America	Engineering	Research to Prove for Design
McLaughlin et al. (2014)	North America	Health	Research to Prove for Design
Wong et al. (2014)	North America	Health	Research to Prove for Design
Brooks (2014)	North America	Social Science	Research to Prove for Design
Murray et al. (2014)	North America	Health	Research to Prove for Design
Moffett and Mill (2014)	Europe	Health	Research to Prove for Design
Whillier and Lystad (2015)	Oceania	Health	Research to Prove for Design
Harrington et al. (2015)	North America	Health	Research to Prove for Design
McCallum et al. (2015)	North America	Mathematics	Research to Describe
Prashar (2015)	Asia	Social Science	Research to Prove for Design
Ryan and Reid (2015)	North America	Natural Sciences	Research to Prove for Design
Velegol et al. (2015)	North America	Engineering	Research to Improve
Gross et al. (2015)	North America	Natural Sciences	Research to Prove for Design
Tanner and Scott (2015)	Africa	Social Science	Research to Improve
Mattis (2015)	North America	Mathematics	Research to Prove for Design
Van Vliet et al. (2015)	Europe	Natural Sciences	Research to Prove for Design
Jensen et al. (2015)	North America	Natural Sciences	Research to Prove for Design
Hotle and Garrow (2015)	North America	Engineering	Research to Prove for Design
Al-Zahrani (2015)	Middle East	Education	Research to Improve
Hung (2015)	Asia	Social Science	Research to Prove for Design
Jungić et al. (2015)	North America	Mathematics	Research to Describe
Danker (2015)	Asia	Arts	Research to Describe
Belfi et al. (2015)	North America	Health	Research to Prove for Design
Porcaro et al. (2016)	Oceania	Health	Research to Prove for Design
Koo et al. (2016)	North America	Health	Research to Prove for Design
Peterson (2016)	North America	Mathematics	Research to Prove for Design
Foldnes (2016)	Europe	Social Science	Research to Prove for Design
Blair et al. (2016)	South America	Engineering	Research to Prove for Design
Liebert et al. (2016)	North America	Health	Research to Prove for Design
Ojennus (2016)	North America	Natural Sciences	Research to Prove for Design
Sohrabi and Iraj (2016)	Middle East	Social Science	Research to Describe
Chen & Hsieh (2018)	Asia	Social Science	Research to Prove for Design
Cabi (2018)	Europe	Social Science	Research to Prove for Design
Lee and Wallace (2018)	Asia	Social Science	Research to Prove for Design
Ritzhaupt and Sommer (2018)	North America	Education	Research to Prove for Design