Article

© Author(s)

Volume 20 (2026), pp. 1-28
American Journal of STEM Education: Issues and Perspectives
eISSN 3065-1190 | Print ISSN: 3069-0072
Star Scholars Press
https://doi.org/10.32674/xcpag319

Improving Undergraduate Statistics Education: Educational Lessons from Pedagogical Experiments

Anushka Karkelanova
Dr. Bing Zhang Department of Statistics, University of Kentucky, USA

Xin Ma
Department of Educational, School, and Counseling Psychology, University of Kentucky, USA

William Rayens
Dr. Bing Zhang Department of Statistics, University of Kentucky, USA

ABSTRACT

We explored instructional ways of engaging students in introductory statistics courses, with two pedagogical experiments. The first (N = 96) examined the effects of traditional versus online manipulatives. After control for gender, age, and high school ACT (American College Testing) mathematics scores, there were no differences in course average between students using traditional and online manipulatives (immediate pedagogical effects), and there were no differences in GPA (grade point average) one year later (prolonged pedagogical effects). The second (N = 270) investigated effects of traditional versus inverted classrooms. After controlling for students individual backgrounds, high school, and university program backgrounds, students in the conventional classroom did better than students in the inverted classroom in midterm grades, while the two groups did equally well in other outcomes.

Keywords: college students, inverted classroom, statistics achievement,

virtual manipulatives,

INTRODUCTION

On March 29, 2018, the prestigious journal, Science, published the largest observational study concerning undergraduate STEM (science, technology, engineering, and mathematics) education. This study monitored about 550 faculty members teaching more than 700 courses at 25 higher education institutions across North America (Stains et al., 2018). Among more than 2,000 college classes, lectures are prominent even though practices vary. Although many faculty members were observed to adopt different teaching styles throughout a semester in an effort to promote learning among students, the mainstream understanding from the North American higher education community is summarized effectively in the title of a newspaper article appeared on Nebraska Today (May 7, 2018): *Massive Study Finds Lectures Still Dominate STEM Education* (by Scott Schrage, a science writer and editor with the University of Nebraska at Lincoln). The surprisingly strong reliance on conventional lecturing represents an urgent global call for educational reforms seeking innovative ways to promote active student involvement in undergraduate teaching and learning process. Stains et al. (2018) and Martin-Alguacil et al. (2024) argued that faculty members in general lack pedagogical knowledge to work with educational reform concepts to turn their classrooms into a student-centered learning environment. This argument, given today's artificial intelligence (AI) enhanced instructional approaches often offering personalized instructional feedback and immersive learning experiences, becomes more valid to challenge the idea that there is a one-size-fits-all approach to teaching and learning.

As a STEM field, the ultimate educational goal of statistics is to create a statistically literate society in which people can appropriately use statistical thinking (Gao et al., 2025; Immekus, 2019; Souza et al., 2020). There has been a growing movement to teach statistics in a more innovative way with a strong emphasis on improving students' ability to reason statistically (Gal & Ograjenšek, 2017; National Governors Association, 2010). Institutions at all educational levels strive to make statistical reasoning a critical part of statistics education around the world (Burrill et al., 2023). Many universities in the United States (US) become involved in reforming the teaching of undergraduate statistics (e.g., Zieffler et al., 2018). To inform statistics educators of potentially innovative instructional practices, we pursued two independent educational experiments to explore alternative ways to engage undergraduate students in introductory statistics courses, the implementation of online virtual (versus traditional concrete) manipulatives in the first experiment and the introduction of inverted (versus traditional) classroom in the second experiment (inverted classroom refers to the instructional practice where activities or events that take place traditionally inside of the classroom now take place outside of the classroom and vice versa). For undergraduate students in introductory statistics courses, we asked whether using online virtual manipulatives, a precursor to AI driven adaptive learning, could improve the level of statistical understanding over using traditional concrete manipulatives (the first study), and whether inverted classroom setting, a very similar concept to small class setting and open classroom layout as advocated in Stains et al. (2018), could improve learning participations and outcomes over traditional classroom setting (the second study). These research questions were examined using experimental design because the (US) National Research Council has confirmed multiple times (e.g., 2010) that randomized clinical trials are the primary tool for evaluating the effectiveness of interventions (innovations).

THE CHALLENGE

In today's data-rich society, statistical reasoning is one of the "must-have" competencies essential to "thrive in the modern world" (Franklin et al., 2005, p. 4) (also Dole & Geiger, 2020; Gao et al., 2025). Literature is full of students' inability to understand statistical concepts and procedures, a strong indication for reform in statistics education. Garfield and Ben-Zvi (2008) explained why statistics is a challenging discipline to teach and to learn. Many statistical ideas and rules are complex, abstract, and even counterintuitive. Many students have difficulties with the underlying mathematics needed to learn statistics (e.g., algebraic manipulation). The context in many statistical issues misleads students to rely on experiences and even faulty intuitions to produce a solution. Students equate statistics with mathematics and expect the focus to be on numbers, computations, and formulas with one and only one correct answer. Finally, students' inadequate experiences make it difficult to prepare them for the massiveness of the real-world data, to pursue different possible interpretations based on different statistical assumptions, and to appreciate the extensive reliance on communication skills to convey statistical information.

Specifically, the challenges in working with statistics are three-fold (i.e., cognitive, affective, and pedagogical). Pedagogically, statistics is a unique discipline and requires a unique way of teaching and learning (see Geller, 2011, presidential address to the American Statistics Association). The famous statistician John Tukey believed that statistics is more of a branch of science than mathematics in that it is sufficient for a mathematical theorem to be elegant (true and beautiful) but statistics is held to an additional standard imposed by science. A statistical model, no matter how elegantly derived, must be discarded if it does not fit the data (Sawka, 2020), often referred in science as "the great tragedy of science" (Huxley, 1870). Mathematics stems logically from its axiomatic structure dictating the order in which mathematics is taught and making any mathematics course self-contained; this course design principle does not work in statistics, and statistics educators who do not understand the unique nature of the discipline often disregard the implications of that uniqueness for teaching and learning, resulting in a statistics course fairly confusing to students (e.g., Rossman & De Veaux, 2016).

Affectively, students' negative attitudes toward statistics (common among students) contribute to their difficulties in statistics courses (e.g., Emmioglu Sarikaya et al., 2022; Chiesi & Primi, 2015; Quane, 2025). According to these authors, students believe that statistics is merely a part of mathematics so that negative attitudes toward mathematics get transferred. So, students, particularly those poor at mathematics, lack self-confidence and have high anxiety in statistics courses. Because students question the usefulness of statistics in their life, they are not motivated to engage in statistics.

Cognitively, students get used to the scientific reasoning of physically cognate disciplines (e.g., mathematics and physics) which are inherently causal, but statistics is essentially acausal. This shift from familiar disciplines with causal interpretations to one largely acausal represents a fundamental paradigm shift, a major cognitive challenge to students (Kurniawan & Wahyuningsih, 2018). Misconceptions on correlation and causality (Janzing, 2019), conditional probability (Fabby, 2021), independence of observations (Nabbout-Cheiban, 2017), randomness and generalizability (Batanero, 2020), and weighted average (Huck, 2015), just to name a few, all have their roots in this fundamental paradigm shift. Inappropriate reasoning about statistical ideas is widespread and persistent across all educational levels (Biehler et al., 2018).

THE SOLUTION

To combat the challenge, statistics educators have called for the development of statistical literacy and interpretive skills as the universal goal of statistics education, advocating content and pedagogy that reflect a change in educational focus from computations and procedures to reasoning and literacy (e.g., Head et al., 2020; Schoen et al., 2025). A balance of content and pedagogy with the assistance of technology moves statistics education from focusing on the “what” to the “how” and “why” of statistics (see Huck, 2015; Tintle et al., 2015). In such a reform, improving instructional methods has been offered to reshape statistics education (Kovacs et al., 2021; Stowell & Addison, 2017; Weiland, 2017). Our research came along with this line of effort, looking into the potential merit of virtual manipulatives and inverted classroom.

Use of Manipulatives

Swan and Marshall (2010) defined manipulatives as “an object that can be handled by an individual in a sensory manner during which conscious and unconscious mathematical thinking will be fostered” (p. 14). Physical (traditional) manipulatives are a set of concrete materials that can be manipulated physically by hand. A physical probability spinner allows students to make predictions, spin for outcomes by hand, and then explore probabilities. On the other hand, virtual (online) manipulatives are a set of imagines that can be manipulated electronically on a computer screen. An online probability spinner allows students to make predictions, click a mouse pointer to spin on a computer screen, and then explore probabilities. Both types of manipulatives achieve the same educational goal of learning the concept of probability, one physically and one virtually.

Confucius (551 – 479 BC) said, “I hear and I forget. I see and I remember. I do and I understand.” Use of manipulatives is considered one of the most effective strategies in statistics education, deepening statistical understanding of students and enhancing their abilities to reason and communicate statistically (Barbosa & Vale, 2025; Cockett & Kilgour, 2015; Larbi & Mavis, 2016). Manipulatives are moveable, create tactile experiences that add a new dimension of learning, give students more control, are traceable in terms of thinking process, relate to real-world applications, promote creativities among students, and allow information to be kept visually and kinesthetically (Carbonneau & Marley, 2015; Delport, 2021).

Hartley (2021) compared physical and virtual manipulatives for mathematics education. Physical manipulatives are easier to move around and work in groups, but they are costly and provide no immediate feedback. Virtual manipulatives do not need time to dispense materials and allow students to save their results for recall and comparison, but they are neither personal nor engaging. Some researchers believe that virtual manipulatives have advantages over physical manipulatives in that they are more interactive (Moyer-Packenham & Bolyard, 2016) and can do things that are impossible with physical manipulatives (e.g., multiple representations of a single concept at the same time) (Sarama & Clements, 2016). Although inconclusive, some differences in learning outcomes do exist between using physical and virtual manipulatives (e.g., Hartley, 2021; Satsangi et al., 2016; Schmit, 2022). While some researchers report that students using physical manipulatives outperform students using virtual manipulatives in mathematics (e.g., Vessonen et al., 2021), others claim the opposite (e.g., Palazuelos, 2017). There are also findings that physical and virtual manipulatives are equally effective in promoting a deeper conceptual understanding (e.g., Justo et al., 2022; Pires et al., 2019).

Inverted Classroom

Only recently have statistics educators begun to investigate whether critical statistical concepts can be developed through a carefully designed sequence of learning activities and how this strategy can be implemented effectively in a classroom setting (e.g., Ben-Zvi, 2018). There are good reasons to value the idea of an inverted classroom that replaces the traditional model of lecturing in class and assigning practice problems after class with a model of assigning learning activities (as homework) before class and discussing practice problems in class (Olakanmi, 2017). Students in an inverted condition are required to watch video lectures as their homework before coming to class (this traditionally happens in class). When students come to class, they are required to complete activities designed to engage them in discovery learning of what has already been experienced by watching the video lectures (this traditionally happens after class). Obviously, the inverted classroom emphasizes student-centered, activity-based learning with increased responsibility of students to become active participants in their own learning (e.g., Amalia & Irwanto, 2025; Mazlan et al., 2025; Yildirim & Kiray, 2016). The inverted classroom involves an important transformation of the teacher's role from being a sage on the stage to a guide on the side (Langdon et al., 2018; McLean & Attardi, 2023). To make an inverted classroom successful, educators need to create an opportunity for students to gain first exposure prior to class, an incentive for students to prepare for class, a mechanism to assess student understanding, and in-class tasks that focus on higher level cognitive activities (see Brame, 2019; Mazlan et al., 2025).

There are three motivations for using an inverted classroom (Mason et al., 2013). It frees class time for interactive activities and for reinforcing course materials without sacrificing content, allows an educator to present course materials in different formats to engage students of different learning styles, and encourages students to become self-regulated learners.

Fulton (2012) listed advantages of the inverted classroom as that students move at their own pace; doing homework in class gives teachers better insights into students' learning difficulties and learning styles; teachers can more easily customize and update the curriculum; classroom time can be used more effectively and creatively; teachers can expect increased levels of student achievement, interest, and engagement; learning theories support this method; there is more time with students on authentic research; students who miss a class can watch the video lectures anywhere; students think both inside and outside of the classroom; and students are more actively involved in the learning process (see also Maciejewski, 2016).

Research has emerged in good quantity to compare the inverted classroom with the traditional classroom in educational outcomes. In the latest research synthesis with a focus on college education, Al-Samarraie et al. (2020) summarized "85 solid articles" (judged as quality assessments, p. 1020) completed between 2015 and 2019. Most articles revealed that the inverted classroom is effective across disciplines (arts, education, social sciences and humanities, natural sciences, medical and health sciences, mathematics, and engineering and technology) to promote achievement, attitude, engagement, metacognition, and understanding. The challenges of adopting the inverted classroom common across disciplines are the length of the video lectures and the time required for instructors to prepare them and for students to master them. In a separate literature review of the same period (2015 to 2019) comparing the inverted classroom with the traditional classroom, Rahman et al. (2020) concluded based on a synthesis of 19 articles (focused on undergraduate education) that the inverted classroom is an effective pedagogical approach regardless of the field of study.

Weaknesses in Literature

We highlight two weaknesses of empirical studies comparing physical and virtual manipulatives as well as traditional and inverted classrooms. First, there is a serious lack of attention to statistics education. Most empirical studies concern about mathematics education. Even the few empirical studies focusing on statistics education get "buried" in mathematics education. In Al-Samarraie et al. (2020), not as a separate category, statistics is classified into mathematics. Rahman et al. (2020) did the same. This makes no sense given the recognition that concrete ways to explore statistics create a deeper understanding of statistical concepts and the argument that inverted classroom is an effective instructional format in statistics education (Burgoyne & Eaton, 2018; Goracke, 2009; Zulkipli et al., 2025). Second, empirical studies are rare at the college level, with most of the prior research focusing on kindergarten to eighth-grade classrooms.

Reimer and Moyer (2005) complained that "the amount of research on high-quality virtual manipulatives is so limited that a judgment about their potential uses in mathematics instruction is entirely speculative" (p. 8). Garfield and Ben-Zvi (2009) acknowledged the growing interest in improving statistics education through inverted classroom but complained that few direct (evidence based) connections have been established between research and practice. Such statements accurately reflect the status of empirical research concerning both issues in undergraduate statistics education. Because of these weaknesses, we were unable to provide sufficient insights on how these instructional methods support the core goals of statistics education, which became the motivation for our research. Finally, the "thin" literature concerning both issues in statistics education adds strongly to the significance of our research.

STUDY 1: EFFECTS OF VIRTUAL MANIPULATIVES

The Experiment

The (US) University of Kentucky began to reform its general education program in 2009. As a part of this effort, Dr. William Rayens carried out a cluster-randomized pedagogical experiment in the fall of 2009. STA 200 is an introductory statistics course (with four sections) required of all undergraduate students who did not take calculus. For the experiment, two sections (with 48 students) were randomly selected as the experimental group (EXP) and the other two sections (with 48 students) were treated as the control group or business as usual group (BAU), for a total sample of 96. STA 200 is set up as two large lectures plus a recitation, thus meeting three times a week. Students in all four sections attended the same lectures. This meant that a single calendar of learning events was created for all four sections, and all students had the same lectures. The recitations followed the same calendar but differed in the type of manipulatives. This was the only difference between the two groups. Overall, the purpose of the design was to give all students (in all four sections) the same lectures by the same instructor at the same time and location and then to have them practice using two different types of manipulatives in recitation to recall and understand concepts and procedures taught in the lectures.

Specifically for recitation, students in the two EXP sections used online virtual manipulatives, and students in the two BAU sections used traditional concrete manipulatives. For example, when students studied patterned repeated sampling, students in BAU spun hand spinners and stacked pegs to create histograms, while students in EXP spun virtual spinners on computer screens and stacked virtual pegs to create histograms. Essentially, the manipulatives were used as helping-to-learn tools for key statistical concepts such as central limit theorem (through spinning bells activity), experimental design (through whacking moles activity), probability (through corn hole likelihood activity), and confidence intervals (through confidence in repetition activity). Dr. Rayens designed all activities used in the course (including workbooks). Detailed lesson plans between EXP and BAU are available from Dr. Rayens (rayens@uky.edu). These lesson plans help clarify that the learning activities were identical between EXP and BAU and highlight the difference in the learning activities between the two groups (i.e., the type of manipulatives only).

Dr. Rayens also made every effort to ensure that the learning conditions were the same for both EXP and BAU throughout the course. For example, in the case of spinning bells activities, Dr. Rayens customized the spinner online so that it had the same number of landing colors of the same colors as the physical spinners. Efforts like this further ensured that the only difference was the type of manipulatives between the two groups, with the learning process made entirely identical. Finally, the experiment went on for the entire course (i.e., one semester).

The Data

As one of the two outcome variables, statistics achievement was measured as the course average combining with equal weights the midterm and final tests as well as several two-minute quizzes focusing on the understanding of major statistical concepts and procedures (statistical vocabulary, confidence intervals, hypothesis testing, experimental design, sampling distributions, generic normal calculations, and correlation). Years before the experiment, the final and midterm tests and the two-minute quizzes were developed, piloted, and revised for use in STA 200 (i.e., they were used for many years). In the paper-and-pencil format with both multiple-choice and short-answer items, the tests and quizzes were given to students in both conditions following the same procedure at the same time and location. The second outcome variable was cumulative GPA (grade point average) for the second year of higher education at the University of Kentucky. This experiment sought to examine both short-term effects (i.e., immediate pedagogical effects) and long-term effects (i.e., prolonged pedagogical effects) of using virtual manipulatives (compared with concrete manipulatives) on academic outcomes among students. The course average was used to measure short-term effects (during completion of STA 200), and the cumulative GPA was used to measure long-term effects (one year after completion of STA 200).

The dichotomous variable of virtual (online) vs concrete (traditional) manipulatives was the key independent variable. Because randomization was done at the course section level (no randomization at the individual student level), there was a need to control for selection bias among students. Data on student background were collected, including continuous variables of age and high school ACT (American College Testing) mathematics scores as well as the dichotomous variable of gender. Race-ethnicity was also collected but not used because there were very few minority students. These variables functioned mainly as control variables in data analysis.

The Analysis

Statistical analysis was aimed at assessing the short-term and long-term effects of the type of manipulatives. To examine the short-term effects (i.e., at the end of the semester when the experiment was implemented), a multiple regression approach to ANOVA (analysis of variance) was adopted. Statistics achievement (i.e., the course average), as described above, was the dependent variable. Test of homogeneity of variances was not statistically significant between EXP and BAU (F = 1.67, p = .20). Student characteristics included gender, age, and high school ACT mathematics scores. The interaction effects between the type of manipulatives and student characteristics were also considered for more control of student effects. To examine the long-term effects (i.e., one year after completing the course), a multiple regression approach to ANOVA was adopted. Students' GPA one year later was the dependent variable. Test of homogeneity of variances was statistically significant between EXP and BAU (F = 6.14, p = .02). The type of manipulatives (virtual vs concrete) and the course average (for the experimental semester) were the independent variables, with student characteristics of gender, age, and high school ACT mathematics scores as control variables.

The Results

Table 1 presents, for both EXP and BAU, descriptive statistics of the two outcome variables, course average and second-year GPA, together with student characteristics of gender, age, and high school ACT mathematics scores. Students in EXP and BAU were quite similar in terms of gender, age, and high school ACT mathematics scores. In the regression model, neither student characteristics nor their interactions with the type of manipulatives were statistically significant, and so they were removed to achieve a parsimonious model concerning the effects of virtual vs concrete manipulatives on course average. Table 2 presents the parsimonious results (the short-term effects). In such a model with the dummy variable (virtual vs concrete manipulatives) as the only independent variable, the course average of students in EXP (virtual manipulatives) was not statistically significantly different from the course average of students in BAU (concrete manipulatives) (effect = -1.95, SE = 2.03). Because of this statistically nonsignificant presence, the regression model accounted for just 1% of the variance in course average.

Table 1: Descriptive Statistics of Outcome Variables and Student Characteristics

	BAU (n = 48)		EXP (n = 48)	
Variable	Mean	SD	Mean	SD
Course average (continuous)	84.44	1.08	82.28	1.74
Second-year grade point average (GPA) (continuous)	3.17	.07	3.17	.09
Male (= 1 versus female = 0)	.21	.06	.33	.07
Age (continuous)	24.15	.23	24.04	.26
High school ACT mathematics scores (continuous)	22.24	.67	23.41	.71

Note. Students in BAU use traditional concrete manipulatives. Students in EXP use online virtual manipulatives.

Table 2: Short-Term Effects of the Type of Manipulatives (on Course Average)

	Effects	SE
Online virtual (vs traditional concrete) manipulatives	-1.95	2.03
Proportion of variance explained	.01	

* $p < .05$.

Table 3 presents the long-term effects of the type of manipulatives and the course average on cumulative GPA one year later. GPA of students who used virtual manipulatives was not statistically significantly different from GPA of students who used concrete manipulatives after controlling for the course average, gender, age, and high school ACT mathematics scores (effect = -.08, SE = .07). However, the course average made a statistically significant difference on GPA one year later (effect = .03, SE = .01). Equivalently, a ten-point increase in the course average was associated with an increase of .30 in GPA one year later (holding other independent variables constant). The regression model accounted for 69% of the variance in GPA one year later. Caution was necessary for the results on the long-term effects because of the statistically significant test on the homogeneity of variances between EXP and BAU. Nonetheless, a direct comparison of GPA (without control variables) was not statistically significant between the two groups after statistical adjustment for their heterogeneity of variances.

Table 3: Long-Term Effects of the Type of Manipulatives (on Second-Year GPA)

	Effects	SE
Online virtual (vs traditional concrete) manipulatives	-0.08	0.07
Course average	0.03*	0.01
Gender	-0.27*	0.09
Age	0.01	0.03
High school ACT mathematics	0.03*	0.01
Proportion of variance explained	0.69	

* $p < .05$.

STUDY 2: EFFECTS OF INVERTED CLASSROOM

The Experiment

Created in 2010 at the University of Kentucky, STA 210 (Introduction to Statistical Reasoning) is an algebra-based conceptual statistics course that emphasizes statistical concepts rather than mathematical manipulations. Euphemistically and perhaps somewhat incorrectly, STA 210 belongs to the genre of courses often labeled as "statistics for poets" across the US, requiring much more writing, reading, and pursuing conceptual ideas than what a traditional statistics course expects. This course attempts to avoid passive learning among students by reducing course content but requiring students to shoulder substantially more responsibility for learning the content. Such a less-is-more idea focuses on three modules of human inference, confidence intervals, and hypothesis testing, using the workbook entitled *Beyond the numbers: Student-centered activities for learning statistical reasoning* (Rayens, 2014). Course content is carefully recorded and publicly placed on YouTube (a total of 18 videos).

Especially, STA 210 addresses the concern that conceptual statistics is too difficult for first-year teaching assistants to handle in their recitations by moving discussions and discoveries (i.e., assignments) into classrooms. This move makes STA 210 inverted because lectures are removed from the classroom to make room for discussions and discoveries (i.e., lectures outside classrooms and assignments inside classrooms). Specifically, students need to participate at some level before they come to class (i.e., they need to study video lectures). Once in class, they apply their knowledge and skills learned from studying video lectures to work with the workbook (that consists of a series of prompts, applications, and hands-on activities) in plenty of interactions with the instructor and teaching assistants who provide assistance to the learning experience of students. Approximately 4,000 undergraduates are taught statistical reasoning in this inverted classroom setting each calendar year, a coordinated effort across more than 70 sections of the course with normally eight faculty instructors and 20 teaching assistants.

An educational experiment was designed for the fall semester of 2014 with students who enrolled in STA 210 to compare the effects of two different instructional methods for presenting statistical content (inverted vs traditional classrooms) on a range of educational outcomes. The inverted instruction was the experimental condition (EXP) and the traditional instruction was the control condition (business as usual or BAU). A total of 270 students participated in the experiment under one instructor with ten first-year teaching assistants (135 students in EXP and 135 students in BAU). Participation of students was not exactly random but blind in that when students signed up for the course, they did not have any knowledge about the experiment. A power analysis indicated that this sample size would allow the detection of a medium effect size (.50) with a statistical power of .80 (alpha set as .05) (see Cohen, 1992 where .50 is considered medium effect size).

To reduce confronting effects, students in both EXP and BAU were taught the same content by the same instructor who carefully trained teaching assistants to perform their roles in either EXP or BAU. Students were taught in the same physical classroom environments (i.e., a typical lecture hall) at the same daytime hours. All students were given the same assignments throughout the semester. Finally, an extensive day-by-day comparison between EXP and BAU was made for the entire semester to ensure that the same topics were covered to the same depth with the same expectation for students (i.e., students were expected to achieve the same level of mastery). Detailed day-by-day lesson plans between EXP and

BAU are available from Dr. Anushka Karkelanova (anushka.karkelanova@uky.edu).

Structurally, for EXP, each week, students met two times as a full class with the instructor and a third time as a small group (recitation) with a teaching assistant (five teaching assistants, one for each group). These meetings were for discussions and discoveries with the instructor or the teaching assistant (i.e., assignments moved into class), after they studied video lectures (i.e., lectures moved out of class). For BAU, each week, students met two times as a full class with the instructor and a third time as a small group (recitation) with a teaching assistant (five teaching assistants, one for each group). Not instructed to study video lectures prior to class, these meetings were life lectures by the instructor or the teaching assistant (i.e., lectures in class). After class, they used the workbook to work on assignments (i.e., assignments after class).

The Data

Comprehensive information was collected to measure students' academic behaviors in the course. The dependent variables included students' academic performance in statistics based on tests and exams, completion of assignments, and major projects. Tests and exams contained both multiple-choice and open-ended questions, and students were given study guides with answers ahead of time. Tests, exams, and projects were developed with periodical revision under the direction and guidance of a faculty advisory committee in the (statistics) department. Other dependent variables captured factual behaviors of students in the course. In total, there were seven dependent measures including projects average, tests average, classwork, midterm attendance average, class final attendance average, midterm grade, and class final grade.

Independent variables portrayed various background characteristics associated with students. Student information was obtained from the university to adjust for various background characteristics of students between EXP and BAU. Given that randomization was only partially achieved in this experiment, these variables were important to control individual and practical differences between EXP and BAU when examining the treatment effects. These variables came from three categories. Student individual background included gender, age, and race-ethnicity. High school background included high school ACT mathematics scores and high school GPA. University program background included university cumulative GPA and study major.

The Analysis

Like Study 1, a multiple regression approach to ANOVA was used to compare students' behaviors in the course between the EXP condition (inverted classroom) and the BAU condition (traditional classroom). While the test of homogeneity of variances between EXP and BAU indicated statistical significance in projects average ($F = 4.60, p = .03$), the other six outcome measures met the assumption ($F = .14, p = .71$ for tests average; $F = 1.32, p = .25$ for classwork; $F = 2.05, p = .15$ for midterm attendance average; $F = 1.17, p = .28$ for class final attendance average; $F = 1.65, p = .20$ for midterm grade; and $F = .19, p = .66$ for class final grade).

Both absolute treatment effects and relative treatment effects were examined. Absolute effects refer to the effects with the absence of control variables, whereas relative effects refer to the effects with the presence of control variables. Specifically in this case, absolute treatment effects compared EXP with BAU on a certain dependent measure without adjustment for variables descriptive of student individual background, high school background, and university program background. Relative treatment effects compared EXP with BAU on the same dependent measure with adjustment for those background variables.

The Results

Table 4 presents descriptive information of dependent and independent variables for students in both EXP and BAU conditions. The independent variables functioned as control variables to adjust for selection bias from different perspectives (student individual background, high school background, and university program background). Students in EXP and BAU were similar across the independent variables. In Table 5, seven dependent variables were examined for absolute treatment effects of inverted classroom against traditional classroom. There were statistically significant (negative) absolute treatment effects on projects average, classwork, midterm attendance average, midterm grade, and class final grade, all in favor of students in BAU (traditional classroom). Caution was necessary for the results on projects average because of the statistically significant test on the homogeneity of variances between EXP and BAU. Nonetheless, the comparison of projects average remained statistically significant between the two groups after statistical adjustment for their heterogeneity of variances.

Table 4: Descriptive Statistics of Dependent and Independent Variables

	BAU (n = 130)		EXP (n =135)	
	Mean	SD	Mean	SD
Dependent variables				
Projects average (continuous)	87.04	10.87	83.43	13.40
Tests average (continuous)	80.11	12.64	78.76	11.57
Classwork (continuous)	87.28	16.34	82.33	15.93
Midterm attendance average (continuous)	89.85	13.61	86.20	15.56
Class final attendance average (continuous)	88.88	14.53	85.41	15.71
Midterm grade (continuous)	86.47	8.36	83.26	8.70
Class final grade (continuous)	85.96	11.98	82.93	11.07
Independent Variables				
Male (= 1, female = 0)	.59	.49	.46	.50
Age (continuous)	20.41	.81	20.72	1.38
Race-ethnicity (white = 1, nonwhite = 0)	.78	.42	.72	.45
High school ACT mathematics scores (continuous)	24.68	3.88	24.39	4.38
High school GPA (continuous)	3.53	.68	3.47	.76
University cumulative GPA (continuous)	3.11	.61	3.02	.59
Engineering (yes = 1, no = 0)	.05	.23	.15	.36
Professional (Education, Nursing) (yes = 1, no = 0)	.24	.43	.16	.36
Economics (yes = 1, no = 0)	.20	.40	.22	.42
Humanities (yes = 1, no = 0)	.18	.39	.21	.41
Sciences (yes = 1, no = 0)	.15	.36	.07	.25
Undecided (yes = 1, no = 0)	.17	.38	.20	.40

Table 5: Absolute Treatment Effects (Inverted Classroom = 1 vs Traditional Classroom = 0)

	Effects	SE
Projects average	-3.61*	1.50
Tests average	-1.36	1.49
Classwork	-4.95*	1.98
Midterm attendance average	-3.65*	1.80
Class final attendance average	-3.47	1.86
Midterm grade	-3.21*	1.05
Class final grade	-3.04*	1.42

* $p < .05$.

Table 6 presents the relative treatment effects of inverted classroom against traditional classroom in terms of projects average, tests average, classwork, midterm attendance average, class final attendance average, midterm grade, and class final grade respectively, all with adjustment for student individual background (gender, age, and race-ethnicity), high school background (high school ACT mathematics scores and high school GPA), and university program background (university cumulative GPA and study major). There were statistically significant (negative) relative treatment effects on midterm grade only (effect = -1.69, SE = .82), in favor of students in BAU (traditional classroom). Finally, the regression model accounted for 38% of the variance in projects average, 60% in tests average, 35% in classwork, 19% in midterm attendance average, 23% in class final attendance average, 54% in midterm grade, and 65% in class final grade.

Absolute treatment effects function to provide background information to help the interpretation of relative treatment effects. A comparison between absolute and relative treatment effects highlights certain "robust" treatment effects. For example, absolute treatment effects were statistically significantly in favor of students in BAU concerning tests average. After controlling student individual background, high school background, and university program background, there were not statistically significant differences in tests average between EXP and BAU (i.e., the absence of relative treatment effects). In contrast, absolute treatment effects were statistically significantly in favor of students in BAU concerning midterm grade. More importantly, even after controlling student individual background, high school background, and university program background, there were still statistically significant differences in midterm grade between EXP and BAU (i.e., the presence of relative treatment effects) in favor of students in BAU. Therefore, the comparison increased our confidence to conclude that treatment effects on tests average were due to selection bias, but treatment effects on midterm grade were robust to selection bias.

Table 6: Relative Treatment Effects (Inverted Classroom = 1 vs Traditional Classroom = 0)

	Effects	SE	R^2
Projects average	-1.81	1.17	.38
Tests average	.75	1.00	.60
Classwork	-3.97	1.65	.35
Midterm attendance average	-2.65	2.10	.19
Class final attendance average	-2.68	1.87	.23
Midterm grade	-1.69*	.82	.54
Class final grade	-1.37	.80	.65

Note. * $p < .05$. Treatment effects are adjusted over individual background (gender, age, race-ethnicity), high school background (high school ACT mathematics scores and high school GPA), and university program background (university cumulative GPA and student major).

DISCUSSION AND CONCLUSIONS

College statistics is a required course for a wide range of study majors, often viewed as a gatekeeper for degree completion by providing critical conceptual understanding as well as foundational knowledge and skills needed for success in subsequent courses (Reyes, 2010). The two pedagogical experiments revealed some interesting findings concerning the treatments (online virtual manipulatives and inverted classroom) in undergraduate statistics education.

Summary of Principal Findings

For the adoption of online virtual manipulatives in comparison with traditional concrete manipulatives, there were not statistically significant differences in statistics achievement as the short-term treatment effects (student background variables of gender, age, and high school ACT mathematics scores as well as their interaction effects with the type of manipulatives were not important to the short-term treatment effects), and there were not statistically significant differences in university GPA one year later as the long-term effects. For the adoption of inverted classroom in undergraduate statistics education in comparison with traditional classroom (lecture), out of seven dependent measures (projects average, tests average, classwork, midterm attendance average, class final attendance average, midterm grade, and class final grade), students in the traditional classroom did better than students in the inverted classroom in midterm grade only, after adjustment for student individual background, high school background, and university program background.

What Can be Said about Virtual Manipulatives?

It is a common expectation that when students can visualize a statistical concept in action, a deeper level of understanding occurs. With the advancement of technology, online virtual manipulatives are becoming increasingly available in statistics education. This study compared the effectiveness of using online virtual manipulatives against traditional concrete manipulatives in undergraduate statistics education. Virtual manipulatives and concrete manipulatives turned out to be equally effective in terms of short-term effects and long-term effects. This equality is not a negative thing to virtual manipulatives. There are some very attractive advantages concerning the use of virtual manipulatives (see our literature review earlier). They can be highly dynamic, interactive, and flexible; they can provide feedback to students immediately upon rendering their response; and they are also very affordable and quite easy to manage. From all these perspectives, the above equality indicates that virtual manipulatives are a good choice (against often quite expensive concrete manipulatives), especially in a budget-tight environment. Overall, this equality is a piece of welcoming evidence to many educational institutions.

We offer two possible explanations for the lack of superiority of virtual manipulatives over concrete manipulatives: Duration of use and quality of manipulatives. There has been a long-standing recognition that only constant use of manipulatives (as opposed to temporary use) provides meaningful improvement in knowledge and skills (e.g., Sowell, 1989). The duration of only one semester of using online virtual manipulatives might not constitute constant use. A longer duration may separate virtual manipulatives from concrete manipulatives, considering the very attractive advantages associated with virtual manipulatives. What could be equally important is to develop more effective and efficient online virtual manipulatives. The benefits of using virtual manipulatives are not static; instead, they are dynamically changing, with the advent of modern technology such as AI. To some extent, this study has laid the foundation for the worthiness of investment into the development and improvement of modern technology assisted online virtual manipulatives.

What Can be Said about Inverted Classroom?

This study was conducted with the expectation that the inverted instruction would show some advantages over the traditional instruction given so many positive arguments and comments about inverted classroom in the literature (note that empirical evidence is still thin in the literature as pointed out earlier). The finding that midterm grade was in favor of students in the traditional instruction, but class final grade was not different between students in the two conditions may indirectly indicate something positive about inverted classroom. It appears that the inverted instruction did close the gap in statistics achievement during the progression from midterm to the end of the course. Similar to the case of online virtual manipulatives, duration and quality may be conditions to benefit from this instructional method. The duration of one semester might not be sufficient to observe meaningful advantages of inverted classroom. The integration of technology beyond pre-recorded videos (as in the case of this study), especially modern technology (e.g., AI assisted inverted classroom), has great potential to improve the instructional quality of inverted classroom. Overall, the newer (modern) digital approaches for both virtual manipulatives and inverted classroom may fundamentally reshape the learning experiences in undergraduate statistics education, given that platforms (e.g., YouTube, LinkedIn, Khan Academy) have already created a new type of learners for whom the primary sources of information and data are virtual rather than derived from traditional sources.

Limitations and Further Research

The age of our studies is indeed a limitation. They were initially planned as internal evaluations of important reforms for undergraduate statistics education. There was immediate data analysis in both studies, providing feedback on curriculum and instruction in the (statistics) department. Later, educational researchers came to realize the empirical values of these studies and pursued them to disseminate their results. Although we thought about replicating Study 1, current complications of conducting randomized educational experiments have eventually prevented our effort. The University of Kentucky is a regional university, serving mainly students from the state. Although the demographic characteristics of the student population are stable over the years, the environmental impact of advancing technology in society in general and the increasing utilization of modern technology by individuals in specific cannot be overlooked. Today’s students, confronted by technology driven media, may well be better prepared to take advantage of educational reforms such as virtual manipulatives and inverted classroom.

Nonetheless, we justify this limitation by gently suggesting that perhaps some historical insights may also be appropriate in social science research. We now know that the use of virtual manipulatives and concrete manipulatives did not show any difference 15 years ago. If a similar study were conducted today, knowledge on the issue would be empirically enriched from a comparison between current results and those of ours. In fact, Study 1 occurred at a time when Applet became the big news for statistics classrooms, just as AI is becoming the big news for today's statistics classrooms. Applet marked a historical time when some faculty across the US began to devote a lot of time into its application in undergraduate statistics education, just as AI is marking a historical time when some faculty around the world begin to experiment with its great potential for teaching and learning in undergraduate statistics education. Study 1, even though of age, did represent a very important historical record in undergraduate statistics education. We want to acknowledge the historical significance of Study 1, while emphasizing its empirical relevance to the current undergraduate statistics education in which many people are seeking technological innovations.

For further research replicating our studies, we discuss the methodological weaknesses of our studies as a way to guide a better research design. The results of Study 2 may have limited generalizability because of the lack of true randomization, even though randomization was partially achieved. One limitation in both studies is the lack of pretest data. Although we did use high school ACT mathematics scores and high school GPA to function as measures of pretests or at least as measures of prior academic abilities, a pretest and posttest experimental design is highly desirable to investigate the effects of virtual manipulatives and inverted classroom. The use of cumulative GPA in Study 1 to examine the long-term effects of virtual manipulatives on academic outcomes among students can also be discussed. Cumulative GPA was the best available academic measure to us one year after the completion of STA 200, avoiding difficult administrative approval and complex data mining to extract information on the 96 students. Nonetheless, cumulative GPA is influenced by many factors and the conclusion of Study 1 is rather tentative. A more reliable measure on the long-term effects needs to be considered. Further studies may also seek some longer period of using each intervention for treatment effects to emerge. Finally, the use of both high-stake and low-stake assessments (to separate conceptual and procedural learning) may have a critical role to play in further research, given that both interventions aim at conceptual understanding. Overall, future research should focus on replicating our findings in diverse settings, improving research designs, and incorporating modern technology to better align with today's digital learning landscape.

REFERENCES

Al-Samarraie, H., Shamsuddin, A., & Alzahrani, A. I. (2020). A flipped classroom model in higher education: A review of the evidence across disciplines. *Educational Technology Research & Development, 68*(3), 1017-1051. doi:10.1007/s11423-019-09718-8

Batanero, C. (2020). Probability teaching and learning. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 491-496). Springer. doi:10.1007/978-94-007-4978-8_128

Ben-Zvi, D. (2018). Three paradigms to develop students' statistical reasoning. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward: Proceedings of the Tenth International Conference on Teaching Statistics* (pp. 1-6). https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_2E1.pdf?1531364242

Biehler, R., Frischemeier, D., Reading, C., & Shaughnessy, J. M. (2018). Reasoning about data. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 139-192). Springer. doi:10.1007/978-3-319-66195-7

Brame, C. J. (2019). *Science teaching essentials: Short guides to good practice*. Academic Press. doi:10.1016/C2017-0-01010-8

Burgoyne, S., & Eaton, J. (2018). The partially flipped classroom: The effects of flipping a module on "junk science" in a large methods course. *Teaching of Psychology, 45*(2), 154-157. doi:10.1177/0098628318762894

Burrill, G. F., de Oliveria Souza, L., & Reston, E. (Eds.) (2023). *Research on reasoning with data and statistical thinking: International perspectives.* Springer. doi:10.1007/978-3-031-29459-4

Carbonneau, K. J., & Marley, S. C. (2015). Instructional guidance and realism of manipulatives influence preschool children's mathematics learning. *The Journal of Experimental Education, 83*(4), 495-513. doi:10.1080/00220973.2014.989306

Chiesi, F., & Primi, C. (2015). Gender differences in attitudes toward statistics: Is there a case for a confidence gap? *Proceedings of the Nineth Congress of the European Society for Research in Mathematics Education,* 622-628.

Cockett, A., Kilgour, P. W., Cockett, A., & Kilgour, P. (2015). Mathematical manipulatives: Creating an environment for understanding, efficiency, engagement, and enjoyment. *Teach Collection of Christian Education, 1*(1), 47-54. https://research.avondale.edu.au/teachcollection/vol1/iss1/5

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1*(3), 98-101. doi:10.1111/1467-8721.ep10768783

Delport, D. (2021). The impact of math manipulatives as a multi-sensory teaching technique in statistics. *Journal of Mathematics Education, Science and Technology, 6*(2), 186-206. doi:10.30651/must.v6i2.10168

Dole, S., & Geiger, V. (2020). *Numeracy across the curriculum: Research-based strategies for enhancing teaching and learning.* Routledge.

Emmioglu Sarikaya, E., Alkan, M. F., Cladera Munar, M., Leavy, A., Maviş Sevim, Ö., Paul, C. I., Primi, C., Schau, C., & Ustun, U. (2022). A systematic review and meta-analysis of the relationships among post-secondary students' attitudes toward statistics and statistics achievement. *Social Science Protocols, 5*(1), 1-9. doi:10.7565/ssp.v5.6949

Fabby. C. (2021). *Identifying student difficulties in conditional probability within statistical reasoning* (Unpublished doctoral dissertation). University of Cincinnati.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2005). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) report.* Alexandria, VA: American Statistical Association. https://www.amstat.org/education/guidelines-for-assessment-and-instruction-in-statistics-education-(gaise)-reports

Fulton, K. P. (2012). 10 reasons to flip. *Phi Delta Kappan, 94*(2), 20-24.

Gal, I., & Ograjenšek, I. (2017). Official statistics and statistics education: bridging the gap. *Journal of Official Statistics, 33*(1), 79-100. doi:10.1515/jos-2017-0005

Garfield, J., & Ben-Zvi, D. (Eds.) (2008). *Developing students' statistical reasoning: Connecting research and teaching practice.* Springer. doi:10.1007/978-1-4020-8383-9

Garfield, J., & Ben-Zvi, D. (2009). Helping students develop statistical reasoning: Implementing a statistical reasoning learning environment. *Teaching Statistics: An International Journal for Teachers, 31*(3), 72-77. doi:10.1111/j.1467-9639.2009.00363.x

Geller, N. L. (2011). Statistics: An all-encompassing discipline. *Journal of the American Statistical Association, 106*(496), 1225-1229. doi:10.1198/jasa.2011.ap11592

Goracke, M. A. (2009). *The role of manipulatives in the eighth-grade mathematics classroom* (Action Research Projects, 71). University of Nebraska. https://digitalcommons.unl.edu/mathmidactionresearch/71/

Hartley, K. N. (2021). *Comparing virtual and concrete manipulatives effect on the conceptual understanding of the FOIL method* [Unpublished manuscript]. Georgia College and State University. https://www.gcsu.edu/sites/default/files/documents/2021-06/hartley.pdf

Head, A. J., Fister, B., & MacMillan, M. (2020). *Information literacy in the age of algorithms: Student experiences with news and information, and the need for change.* Project Information Research Institute. https://www.projectinfolit.org/uploads/2/7/5/4/27541717/algoreport.pdf

Huck, S. (2015). *Statistical misconceptions*. Routledge. doi:10.4324/9781315651941

Huxley, T. H. (1870). Address to the British Association for the Advancement of Science. *American Journal of Science & Arts, 50*(150), 383-402.

Immekus, J. C. (2019). Flipping statistics courses in graduate education: Integration of cognitive psychology and technology. *Journal of Statistics Education, 27*(2), 79-89. doi:10.1080/10691898.2019.1629852

Janzing, D. (2019). The cause-effect problem: Motivation, ideas, and popular misconceptions. In I. Guyon, A. Statnikov, & B. Bakir Batu (Eds.), *Cause effect pairs in machine learning* (pp. 3-26). Springer. doi:10.1007/978-3-030-21810-2

Justo, E., Delgado, A., Llorente-Cejudo, C., Aguilar, R., & Cabero-Almenara, J. (2022). The effectiveness of physical and virtual manipulatives on learning and motivation in structural engineering. *Journal of Engineering Education, 111*(4), 813-851. doi:10.1002/jee.20482

Kovacs, P., Kuruczleki, E., Kazar, K., Liptak, L., & Racz, T. (2021). Modern teaching methods in action in statistical classes. *Statistical Journal of the IAOS, 37*(3), 899-919. doi:10.3233/SJI-210843

Kurniawan, D., & Wahyuningsih, T. (2018). Analysis of student difficulties in statistics courses. *International Journal of Trends in Mathematics Education Research, 1*(2), 53-55. doi:10.33122/ijtmer.v1i2.39

Langdon, J., Sturges, D., & Schlote, R. (2018). Flipping the classroom: Effects on course experience, academic motivation, and performance in an undergraduate exercise science research methods course. *Journal of the Scholarship of Teaching and Learning, 18*(4), 13-27. doi:10.14434/josotl.v18i4.22729

Larbi, E., & Mavis, O. (2016). The use of manipulatives in mathematics education. *Journal of Education and Practice, 7*(36), 53-61.

Maciejewski, W. (2016). Flipping the calculus classroom: An evaluative study. *Teaching Mathematics and Its Applications, 35*(4), 187-201. doi:10.1093/teamat/hrv019

Martin-Alguacil, N., Avedillo, L., Mota-Blanco, R., & Gallego-Agundez, M. (2024). Student-centered learning: Some issues and recommendations for its implementation in a traditional curriculum setting in health sciences. *Education Sciences, 14*(11), Article 1179. doi:10.3390/educsci14111179

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. *Transactions on Education, 56*(4), 430-435. doi:10.1109/TE.2013.2249066

McLean, S., & Attardi, S. M. (2023). Sage or guide? Student perceptions of the role of the instructor in a flipped classroom. *Active learning in higher education*, *24*(1), 49-61. doi:10.1177/1469787418793725

Moyer-Packenham, P. S., & Bolyard, J. J. (2016). Revisiting the definition of a virtual manipulative. In P. S. Moyer-Packenham (Ed.), *International perspectives on teaching and learning mathematics with virtual manipulatives* (pp. 3-23). Springer. doi:10.1007/978-3-319-32718-1_1

Nabbout-Cheiban, M. (2017). Intuitive thinking and misconceptions of independent events: A case study of us and French pre-service teachers. *International Journal of Research in Undergraduate Mathematics Education, 3*(2), 255-282. doi:10.1007/s40753-016-0038-x

National Governors Association. (2010). *Common core state standards.* Washington, DC.

National Research Council. (2010). *The prevention and treatment of missing data in clinical trials.* National Academies Press. doi:10.17226/12955

Olakanmi, E. E. (2017). The effects of a flipped classroom model of instruction on students' performance and attitudes towards chemistry. *Journal of Science Education and Technology, 26*(1), 127-137. doi:10.1007/s10956-016-9657-x

Palazuelos, K. (2017). *The impact of virtual manipulatives on student achievement in mathematics* [Unpublished doctoral dissertation]. Utah State University. https://digitalcommons.usu.edu/teal_facpub/1430/

Pires, A. C., González Perilli, F., Bakała, E., Fleisher, B., Sansone, G., & Marichal, S. (2019). Building blocks of mathematical learning: Virtual and tangible manipulatives lead to different strategies in number composition. *Frontiers in Education, 4,* Article 81. doi:10.3389/feduc.2019.00081

Rahman, S. F. A., Yunus, M. M., & Hashim, H. (2020). The uniqueness of flipped learning approach. *International Journal of Education and Practice, 8*(3), 394-404. doi:10.18488/journal.61.2020.83.394.404

Rayens, W. (2014). *Beyond the numbers: Student-centered activities for learning statistical reasoning* (4th ed.). Van-Griner Publishing Company.

Reimer, K., & Moyer, P. S. (2005). Third graders learn about fractions using virtual manipulatives: A classroom study. *Journal of Computers in Mathematics and Science Teaching, 24*(1), 5-25.

Reyes, C. (2010). Success in algebra among community college students. *Community College Journal of Research and Practice*, *34*(3), 256-266. doi:10.1080/10668920802505538

Rossman, A., & De Veaux, R. (2016). Interview with Richard De Veaux. *Journal of Statistics Education, 24*(3), 157-168. doi:10.1080/10691898.2016.1263493

Sarama, J., & Clements, D. H. (2016). Physical and virtual manipulatives: What is "concrete"? In P. S. Moyer-Packenham (Ed.), *International perspectives on teaching and learning mathematics with virtual manipulatives* (pp. 71-93). Springer. doi:10.1007/978-3-319-32718-1_4

Satsangi, R., Bouck, E. C., Taber-Doughty, T., Bofferding, L., & Roberts, C. A. (2016). Comparing the effectiveness of virtual and concrete manipulatives to teach algebra to secondary students with learning disabilities. *Learning Disability Quarterly, 39*(4), 240-253. doi:10.1177/0731948716649754

Sawka, K. (2020). The use and misuse of statistics. In C. W. Gruber (Ed.), *The theory of statistics in psychology: Applications, use,*

and misunderstandings (pp. 95-110). Springer. doi:10.1007/978-3-030-48043-1
Schmit, M. (2022, July 13). *Physical vs. virtual manipulatives: Is there a difference?* Brainingcamp. https://www.brainingcamp.com/blog/posts/physical-vs-virtual-manipulatives-is-there-a-difference
Schrage, S. (2018, May 7). Massive study finds lectures still dominate STEM education. *Nebraska Today.* https://www.facultyfocus.com/articles/teaching-and-learning/massive-study-finds-lectures-still-dominate-stem-fields/
Souza, L. de O., Lopes, C. E., & Fitzallen, N. (2020). Creative insubordination in statistics teaching: Possibilities to go beyond statistical literacy. *Statistics Education Research Journal, 19*(1), 73-91. doi:10.52041/serj.v19i1.120
Sowell, E. J. (1989). Effects of manipulative materials in mathematics instruction. *Journal for Research in Mathematics Education, 20*(5), 498-505. doi:10.2307/749423
Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., Eagan, M. K., Jr., Esson, J. M., Knight, J. K., Laski, F. A., Levis-Fitzgerald, M., Lee, C. J., Lo, S. M., McDonnell, L. M., McKay, T. A., Michelotti, N., Musgrove, A., Palmer, M. S., Plank, K. M., & Rodela, T. M. (2018). Anatomy of STEM teaching in North American universities. *Science, 359*(6383), 1468-1470. doi:10.1126/science.aap8892
Stowell, J. R., & Addison, W. E. (Eds.) (2017). *Activities for teaching statistics and research methods: A guide for psychology instructors.* Washington, DC: American Psychological Association. doi:10.1037/0000024-000
Swan, P., & Marshall, L. (2010). Revisiting mathematics manipulative materials. *Australian Primary Mathematics Classroom, 15*(2), 13-19.
Tintle, N., Chance, B., Cobb, G., Roy, S., Swanson, T., & VanderStoep, J. (2015). Combating anti-statistical thinking using simulation-based methods throughout the undergraduate curriculum. *The American Statistician, 69*(4), 362-370. doi:10.48550/arXiv.1508.00543
Vessonen, T., Hakkarainen, A., Väisänen, E., Laine, A., Aunio, P., & Gagnon, J. C. (2021). Differential effects of virtual and concrete manipulatives in a fraction intervention on fourth and fifth grade students' fraction skills. *Investigations in Mathematics Learning, 13*(4), 323-337. doi:10.1080/19477503.2021.1982586

Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. *Educational Studies in Mathematics, 96*(1), 33-47. doi:10.1007/s10649-017-9764-5

Yildirim, F. S., & Kiray, S. A. (2016). Flipped classroom model in education. *Research Highlights in Education and Science, 2*(6), 1-8.

Zieffler, A., Garfield, J., & Fry, E. (2018). What is statistics education? In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 37-70). Springer. doi:10.1007/978-3-319-66195-7_2

Note: This paper draws on information from a dissertation by the first author, supervised by the second and third authors, in the completion of a Doctor of Philosophy (PhD) program at the University of Kentucky.

The authors did not use OpenAI's ChatGPT or any other AI tools in the drafting, editing, or refining of this manuscript. All content was generated, reviewed, and refined solely by the authors.

ANUSHKA KARKELANOVA, PhD, is a Lecturer at the Dr. Bing Zhang Department of Statistics, University of Kentucky. Her major research interest is mathematics (statistics) education.
Email: anushka.karkelanova@uky.edu

XIN MA, PhD, is Full Professor of Quantitative and Psychometric Methods at the Department of Educational, School, and Counseling Psychology, University of Kentucky. His major research interests include advanced statistical methods, large-scale assessment, program evaluation, policy analysis, organization effectiveness and improvement, and psychology of mathematics education.

WILLIAM RAYENS, PhD, is Full Professor of Statistics and Department Chair at the Dr. Bing Zhang Department of Statistics, University of Kentucky. His major research interest is statistics education.