Randomized Controlled Trials in Online Higher Education Research

Kathryn Becker-Blease
Ameer Almuaybid
Oregon State University

Research Fellows | January 2018

Oregon State University

“Not all education research is equal.” – *What Works Clearinghouse*

“In education, there are no federal or state laws protecting consumers from bad educational practices.” – Daniel Willingham, *When Can You Trust the Experts*?

Educators need fact-based, objective, scientific evidence on what works and what does not work. Faculty, students, and increasingly society at large, agree: no one can afford to waste time or money on ineffective educational schemes or fads. Yet, no laws protect students and faculty from bad educational practices and products (Willingham, 2012). When patients visit the doctor, they assume the advice given is based on evidence about what is safe and effective. Students cannot be as sure of that in school.

Comparatively, the U.S. Department of Education runs the *What Works Clearinghouse* solely to evaluate the quality of educational research, and educate the public about the strengths and weaknesses of research methodologies. They provide a user-friendly flow chart for evaluating educational practices and products that begins with the question “Are groups randomly assigned?” (See Appendix A). Appropriately, it is impossible to earn the highest rating of quality if the answer is no.

In online higher education, Bowen, Chingos, Lack, and Nygren (2012, p. 8) summarized research on randomized controlled trials (RCTs) on interactive online learning at public universities this way:

> Very few of the studies use randomized assignment techniques to create ‘treatment’ and ‘control’ groups that can be used to reduce otherwise ubiquitous selection effects that make it hard to interpret findings.

A year later, Lack similarly concluded, “It is unfortunately the case that there have been few rigorous efforts to produce compelling evidence of the learning outcomes associated with online courses at the post-secondary level,” (Lack, 2013, p. 1).

Online Education Allows for Uniquely Powerful RCTs

Online education allows scientist instructors to use RCTs to pinpoint particularly important factors to better understand the answer to a crucial question: Which factors, exactly, are responsible for improving learning, which are detrimental, and which make no difference? Specifically, scientist instructors can use the online platforms that power online classes to:

1. set up natural experiments
2. randomize efficiently
3. set up blind and double-blind studies

Natural experiments

A natural experiment is a type of quasi-experiment that occurs when conditions in the real world serendipitously allow for a comparison between two groups, often a treatment-as-usual group and a group that receives some new treatment. For example, when there are more people who qualify for a program, such as health insurance (Baicker et al., 2013) or charter school admission (Angrist, Bettinger, Bloom, King, & Kremer, 2002), researchers can compare those who were admitted by chance through a lottery to those who were not admitted. In some circumstances, it may be possible to compare students who are admitted to those on a waitlist, but typically the number of waitlisted students is much lower than the number admitted, and students are typically allowed to register in a particular order (e.g., by seniority) which makes students on the waitlist differ in meaningful ways from those who are admitted.

In face-to-face classes, it may be possible to yoke two classes taught by the same instructor, and/or at the same time, and compare one teaching method to another (e.g., Becker-Blease, 2013; Becker-Blease, Bostwick, Almuaybid, & Soicher, 2017; SRI Education, 2016). The problem is that it is not possible to simultaneously control for even two of the most relevant variables (e.g., instructor and time, or instructor and term, or room and time) when classes are taught face-to-face in a physical location simultaneously.

In asynchronous online classes, when an instructor is scheduled to teach two sections of the same course, it is possible to control for more factors, controlling for the instructor and digital learning space (e.g., online homework platforms) and time (i.e., students are all working asynchronously at times of their choosing). However, even in online classes, students are typically not randomly assigned to sections. Earlier enrollees tending to enroll in the first listed section and there could be meaningful differences between groups at baseline that will affect performance in class. In these cases, pre-test scores and demographic variables can be used to assess the similarity of the participants in each class. In a recent study, we were pleasantly surprised to find that students in sections listed first in the schedule of classes (i.e., sections 001 and sections 002) had identical knowledge pre-test scores at baseline, even though the first section on the list (i.e., section 001) filled before the one listed below it (i.e., section 002, Becker-Blease, Almuaybid, & Soicher, 2017). Thus, although we did not have truly random assignment into the experimental and control sections, we were able to rule out differences in baseline content knowledge and test-taking ability.

Randomize efficiently

In face-to-face classes, it is quite rare to be able to randomly assign students to sections, instructors, or classes. There are some cases where this might occur. Some examples we have heard of include the military academies in which students' entire daily schedules are controlled by the school; limited cases where the Registrar might help facilitate a study; cases where high enrollment, eligibility for a high enrollment class is determined by last name; or cases where a school uses quasi-random assignment to put first year students into learning communities in which they take a bloc of classes together. Randomization is always preferable when the goal is to ascertain whether a particular activity, approach or intervention causes more of an improvement than class-as-usual, but often is not possible.

In face-to-face classes, it is possible to randomly assign students to conditions within the same class. For example, students can number off and then move to rooms or areas of the room based on number. The instructor or TA can then walk around and distribute different versions of an assignment that is being studied. Or, instructors or TAs can create different versions of tests or homework, collate them into piles (Version A, Version B, Version C, etc.) and pass out the stacks (Bostwick & Becker-Blease, under review). This does become challenging in small classes, where students may prefer to work with certain people or the differences between groups may be obvious. It can also be challenging in large classes. Even passing out papers in rooms that seat hundreds of students can take a lot of time (especially if instructors are not assigned TAs), and there is often neither enough smaller break out rooms or room in the class for students to move into small groups. In both large and small classes, an additional step is needed for students who are not present in class. In contrast, most online learning management systems (LMS) have built-in systems for randomly assigning students to groups, assignments, and/or and documents in a few seconds. These methods approach true randomization, and also avoid the practical problems with papers and people moving around classrooms. Thus, the instructor could use randomization to study the efficacy of certain modules, for example.

Blind and double-blind studies

One of the biggest challenges to knowing if educational products and practices are beneficial, harmful, or have no effect relates to a set of related powerful confounds: instructor, experimenter, and participant expectancy. In educational research, quite often the instructor in one way or another helped to choose or create the practice or product to be tested. In fact, we have heard researchers who included their own students in their studies who say they directly tell students not only that the intervention will work, but how. For example, an instructor might say,

> I'm trying out a new type of homework this term, and I want to make sure it works, so I'm comparing it to regular homework in my other section. This new homework is based on the concept of distributed practice, which has been shown to work in classes like this

because of the way it helps students build memories over time.

This kind of statement is normal in classrooms, but highly problematic for efficacy research because it introduces the instructors' expectation of a certain outcome and induces participants to share that expectation. Expectancy and placebo effects are often not addressed, even in highly publicized educational studies (e.g., Freeman et al., 2014). To get around this, in face-to-face classes, a teaching or research assistant may be able to introduce the study or particular assignment, but this requires another person to be available, who preferably does not know the hypotheses or which condition is which.

In online classes, instructors can set up conditions through the learning management system that introduce the conditions to students. Instructors can control the way the conditions are presented to make sure they are presented without bias, and can even pre-test (manipulation check) the instructions to demonstrate they are neutral. Students log in and access particular activities or documents like usual, helping them to remain blind to conditions and hypotheses. In fact, in at least one study, the names of instructors were manipulated to show that students give lower teaching evaluation ratings when they believe their instructor is a woman compared to a man (MacNell, Driscoll, & Hunt, 2015), a technique close to impossible to pull off in a face-to-face classroom. In addition, researchers and program coordinators can set up courses in the LMS in a way that keeps instructors blind to conditions and hypotheses as well. In these double-blind studies, researchers can rule out instructor and student expectancy effects. Often researchers have no interaction with students in these studies, essentially allowing for triple-blind studies that control for instructor, student, and researcher expectancy effects.

Practical Considerations for Conducting RCTs in Online Education

What follows are some specific techniques and practical suggestions for setting up RCTs in online classes.

Randomization through Learning Management Systems (LMS)

There are two main ways of randomizing through a LMS. First, some systems allow instructors to create groups randomly, and then assign students to groups without allowing individual students to see each other's work or affecting their overall experience in any way. The fact that they were part of a group that received a particular manipulation is invisible to them.

Other systems allow students to be randomly assigned to groups, but assigning those groups an assignment automatically makes those groups visible to the members. Whether or not this is acceptable depends on the study design. Because it is not necessarily clear at the outset which way a particular LMS system will work, it is important to test out any randomization feature ahead of time to see how it works. Unfortunately, this can be difficult to adequately test with the default "student view mode" or with a single "test student" that some LMSs offer by default. It is better to get a sandbox course setup and enroll several fake students, and log in as each fake student to fully understand how the system will operate. It is also important for instructors to check the gradebook to understand how these assignments will appear. (We once had students asking us what the "Group A" assignment was in the gradebook and why they didn't have the "Group B" assignment like their friends. We forgot to check a box.) See this video, RCT in Canvas that presents a walkthrough of how to set up randomized assignments in the Canvas LMS.

Randomization in Qualtrics

Some instructors find Qualtrics, a powerful survey tool, to also be a useful teaching tool. Qualtrics makes it easy to randomize students to receive one bloc of questions, pictures, videos, or links versus another, and to ask students questions before and after randomization. For example, we once piloted an intervention in a computer science class in which the instructor sent out a regular weekly email with a link to a Qualtrics "homework pre-course survey assignment" at the end. The survey randomized students at the outset to receive one of two motivational messages before asking students questions about their interests and study strategies for the course. We then tracked students' grades across the term for the two motivational message groups. We will share two important tips. First, make sure the instructor who sets this up for you has a way to track who completes each survey. Second, make sure the instructor conveys clearly that the link to the Qualtrics survey is a graded assignment. Students may pay less attention to it because their other assignments are typically not on Qualtrics. See this video, [RCT in Qualtrics]() that presents a walkthrough of how to set up randomized content in Qualtrics.

Randomization through Publisher's Homework System
Some educational publisher's homework systems function as LMSs that have a built-in way to randomly assign students to groups, assignments, or documents. In addition, some publishers are able to program their systems to create two conditions for research or product testing. This kind of digital product testing is common across the digital industry (e.g., www.optimizely.com).

One particularly elegant design answered a call from both researchers and instructors for more compelling efficacy data for a particular product relatively new to the marketplace. In this study, the regular homework platform was redesigned so that for the first assigned chapter, students were randomly assigned to take a regular, static quiz with all students receiving the same questions or an adaptive quiz in which the question difficulty was automatically adjusted after each question based on student performance. For the second assigned chapter, students completed the static and adaptive quiz in the opposite order. This allowed the publisher to compare performance when students took the adaptive quiz first to when they took it second. Links to the quizzes (modified with random assignment) were provided to the instructor to give to students, so the instructor did not need to set up randomization any other way. A nice report describing the design and results is now prominently displayed on the product's website (http://books.wwnorton.com/books/inquizitive/overview/), one of the few cases where the evidence actually supports the marketing claim for a digital product:

> InQuizitive improved student quiz scores by nearly a letter grade in a recent efficacy study. Read about the experimental design here.

Pre- and Post-Tests
In some recent work, we compared pre- and post-tests in three different settings: 1) fully online classes with online pre- and post-tests, 2) face-to-face class with online pre-test and in-class post-test, and 3) face-to-face classes with in-class pre- and post-tests (Becker-Blease, Almuaybid, et al., 2017; Becker-Blease, Bostwick, et al., 2017). Initially, our colleagues at other institutions and we were thrilled that pre-test scores were statistically the same among all sections. However, once we calculated post-test minus pre-test scores for all students, we uncovered a potential problem.

We discovered that students who took the pre-test online had no time pressure, and students who took the pre- and post-test in the classroom had minimal time pressure, but students taking the post-test online had extreme time pressure. Only the online classes had exams with tight time constraints designed to minimize students' cheating on the unproctored exams. This issue was complicated by the fact that our pre- and post-tests intentionally included some very difficult problems that asked students to read a paragraph summarizing a scientific study and a fairly complicated graph, and then answering questions that were not simple knowledge questions but tapped skills in applying what students had learned about statistics and research design. Not only are these questions challenging, but the cognitive demands are such that the time it takes students to read and think are much more variable than for typical multiple-choice exams, yet they were worth the same number of points as the rest of the questions (which was made visible to students automatically through the learning management system). We discovered that some students skipped these difficult questions under time pressure, and we suspect others chose to guess on those rather than take the time to answer them and risk running out of time for the shorter questions. We had discovered three important and unanticipated confounding variables (time pressure, perceived question difficulty, and student ability), that may have acted independently or together to influence the results.

Further analyses are underway to understand the extent to which this issue differentially affects students with less knowledge, skill, motivation, or meta-cognitive ability (which could limit a student's ability to wisely allocate time), as well as students who differ in motivation on the pre- versus post-test, or when the points possible is made more or less salient, and between students taking exams online versus face-to-face or under time pressure versus no time pressure. All of these factors deserve further research attention. The question of how different students approach tasks under varying real-world conditions is especially urgent because it goes to the validity not only of the research measures, but actual grades. As a preliminary step, we recommend using proctored exams with no time pressure, or separating out easy, medium, and hard questions, allowing more time for the most challenging questions. We believe this could be done with the questions we used because the questions required demonstrating skills rather than knowledge; they cannot look up the answers to these questions. To prevent students from working together or sharing questions, it may be necessary to deploy alternative versions of those questions in particular.

In conclusion, for educators, instructional designers, higher education administrators, and policymakers who need to know what works in online education, RCTs are essential. Fortunately, instructional designers and educators are well-positioned to do their own scholarship of teaching and learning RCTs (Linder & Dello Stritto, 2017). The process of planning courses ahead, using software to randomly assign students, and deliver separate content simultaneously allow for powerful designs tested in real-world settings.

Additional Resources:

[Martin, Gurung & Wilson (2014). IRBs and Research on Teaching and Learning]

[U.S. Department of Education Educational Technology Rapid Cycle Evaluations Coach]

Acknowledgments

This work was funded by an Oregon State University Ecampus Research Unit Fellowship awarded to the first author. Thanks to OSU Extended Campus; Canvas support team, Lynn Greenough, Katie Linder; and the Oregon State University General Psychology Research team for suggestions and technical assistance. Thanks to the School of Psychological Science Ecampus instructors for helping with implementing the research design. Thanks to all the student participants for consenting and agreeing to participate.

References

Angrist, J., Bettinger, E., Bloom, E., King, E., & Kremer, M. (2002). Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment. *The American Economic Review, 92*(5), 1535-1558.

Baicker, K., Taubman, S. L., Allen, H. L., Bernstein, M., Gruber, J. H., Newhouse, J. P., . . . Finkelstein, A. N. (2013). The Oregon experiment—effects of Medicaid on clinical outcomes. *New England Journal of Medicine, 368*(18), 1713-1722.

Becker-Blease, K. A. (2013). Student Achievement and Satisfaction in Introductory Psychology: No Significant Differences Between Face-to-Face and Hybrid Sections. *L. L. Stewart Faculty Development Award Fund Report*. Retrieved from http://ctl.oregonstate.edu/sites/ctl.oregonstate.edu/files/hybrid-osu-student-achievement-becker-blease.pdf

Becker-Blease, K. A., Almuaybid, A. J., & Soicher, R. N. (2017). *Consent rates but not pre-test scores vary among online, blended, and face-to-face sections of introductory psychology classes: Implications for scholarship on teaching and learning*. Paper presented at the 2nd Biennial International Convention of Psychological Science, Vienna, Austria.

Becker-Blease, K. A., Bostwick, K., Almuaybid, A. J., & Soicher, R. N. (2017). *Three Efficient Ways to Conduct Experimental Research in Large Online, Hybrid, and Face-to-Face Classes.* Paper presented at the 2nd biennial International Convention of Psychological Science., Vienna, Austria.

Bostwick, K., & Becker-Blease, K. A. (under review). Quick, Easy Mindset Intervention Can Boost Academic Achievement in Large Introductory Psychology Classes.

Bowen, W. G., Chingos, M. M., Lack, K. A., & Nygren, T. L. (2012). *Interactive Learning Online at Public Universities: Evidence from Randomized Trials*. Retrieved from http://www.sr.ithaka.org/wp-content/uploads/2015/08/sr-ithaka-interactive-learning-online-at-public-universities.pdf

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410-8415. doi:10.1073/pnas.1319030111

Lack, K. A. (2013, March 21). Current Status of Research on Online Learning in Postsecondary Education. Retrieved from https://doi.org/10.18665/sr.22463

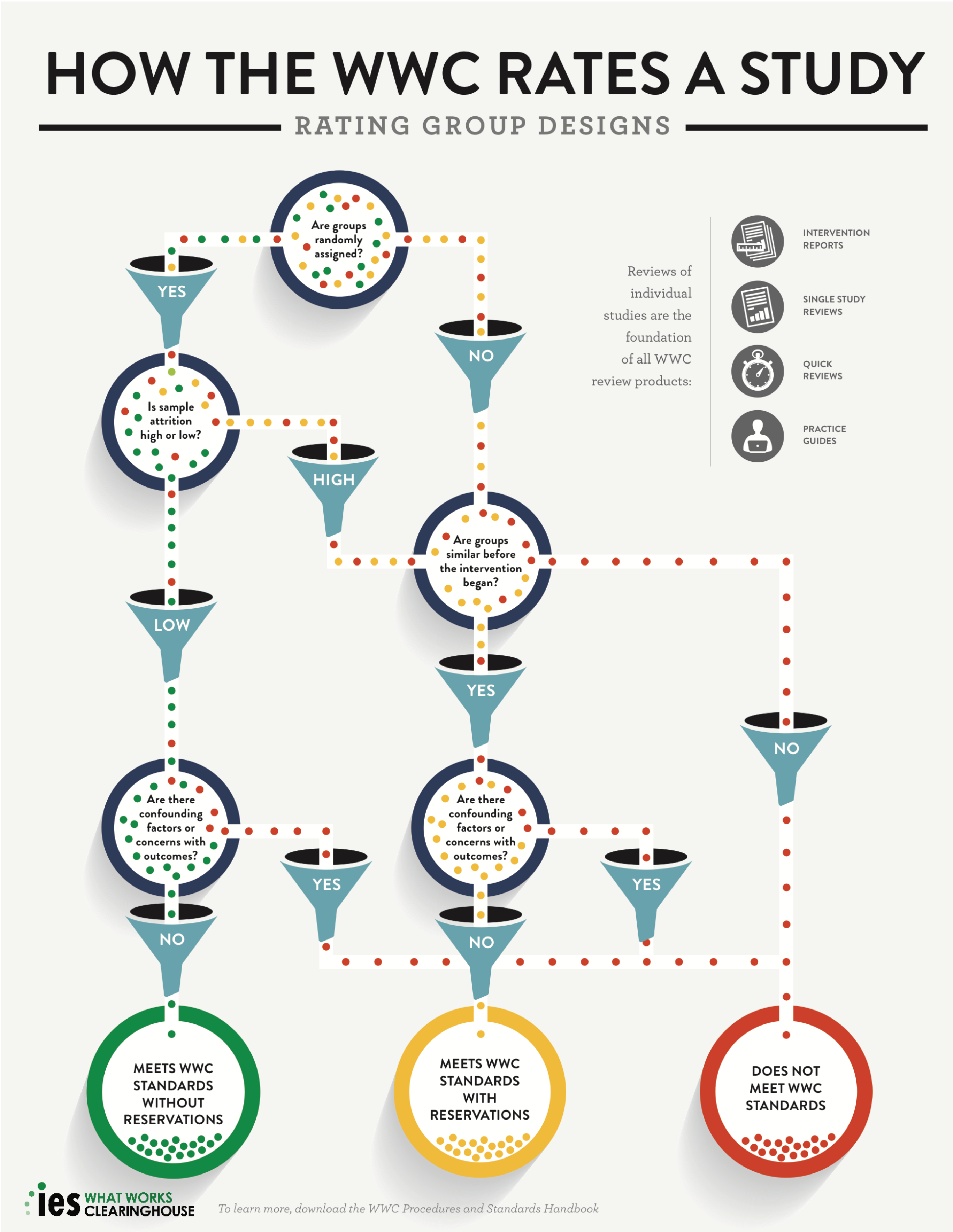
MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a Name: Exposing Gender Bias in Student Ratings of Teaching. *Innovative Higher Education*, *40*(4), 291-303. doi: 10.1007/s10755-014-9313-4

Linder, K. & Dello Stritto, M. E. (2017). *Research Preparation and Engagement of Instructional Designers in Higher Education*. Corvallis, OR: Oregon State University Ecampus.

SRI Education. (2016). *Lessons Learned From Early Implementations of Adaptive Courseware*. Retrieved from https://www.sri.com/sites/default/files/brochures/almap_final_report.pdf

Willingham, D. T. (2012). *When can you trust the experts: How to tell good science from bad in education*: John Wiley & Sons.

Appendix A:



Source: What Works Clearinghouse: https://ies.ed.gov/ncee/wwc/WhatWeDo

About the Research Unit at Oregon State Ecampus

Vision

The Ecampus Research Unit supports Oregon State University's mission and vision by conducting world-class research on online education that develops knowledge, serves our students and contributes to the economic, social, cultural and environmental progress of Oregonians, as well as national and international communities of teachers and learners.

Mission

The Ecampus Research Unit (ECRU) makes research actionable through the creation of evidence-based resources related to effective online teaching, learning and program administration toward the fulfillment of the goals of Oregon State's mission. Specifically, the research unit conducts original research, creates and validates instruments, supports full-cycle assessment loops for internal programs, and provides resources to encourage faculty research and external grant applications related to online teaching and learning.

Contact

Katie Linder, Ph.D.
Director of Research
Oregon State Ecampus
541-737-4629
kathryn.linder@oregonstate.edu

Creative Commons License

Suggested Citation

Becker-Blease. K. & Almuaybid. A. (2018). Randomized controlled trials in online higher education. *White Paper*. Corvallis, OR: Oregon State University Ecampus Research Unit.