WWC Single Study Review

A review of the design and summary of findings for an individual study

ies INSTITUTE OF EDUCATION SCIENCES

March 2018

WWC Review of the Report "Should Students Assessed as Needing Remedial Mathematics Take College-level Quantitative Courses Instead? A Randomized Controlled Trial"[1]

The findings from this review do not reflect the full body of research evidence on mainstreaming.

What is this study about?

The study examined the impact of mainstreaming college freshmen into college-level quantitative courses. College policies usually require students to take remedial courses (also known as developmental education) if their placement exams indicate they are unprepared for college-level work. Such courses can be barriers to college completion because many students delay taking these courses, and because students typically have low passing rates in remedial courses.[2] Mainstreaming refers to the practice of placing students who are in need of remedial courses into regular degree-track courses. The goal of mainstreaming is to help students avoid delays in completing their degree.

The study was conducted at three City University of New York (CUNY) community colleges. Within each community college, the authors randomly assigned students to one of three groups: 297 students were assigned to a mainstream credit bearing statistics course with weekly workshops (Stat-WS), 313 students were assigned to remedial elementary algebra with weekly workshops (EA-WS), and 297 students were assigned to remedial elementary algebra without workshops (EA). Twelve instructors participated in the study, and each instructor taught one section of each study condition (i.e., Stat-WS, EA-WS, and EA).

Students in the mainstreaming (Stat-WS) condition enrolled in a college-level introductory statistics course in the fall of 2013. This course included topics on probability, binomial probability distributions, normal distributions, confidence intervals, and hypothesis testing. The course duration varied by college and ranged from 3 to 6 hours per week over the course of one semester. Weekly workshops lasted 2 hours each.

Students in the comparison groups (EA-WS and EA) were enrolled in a semester-long, traditional remedial algebra course that covered topics such as linear equations, exponents, polynomials, and quadratic equations. Workshops offered to the EA-WS group followed the same structure as those used in the Stat-WS group.

The study assessed the effectiveness of mainstreaming on academic achievement (as measured by percent of students who passed their assigned class in the fall of 2013), and credit accumulation and persistence (as measured by fall-to-fall persistence and total number of college credits earned over three semesters).[3] Since the focus of the report was on mainstreaming, this review will focus on the most direct test of mainstreaming's effects: the comparison between the Stat-WS group and the EA-WS group.

WWC Rating

The research described in this report meets the WWC group design standards without reservations

The study is a randomized controlled trial with low attrition.

Features of the Mainstreaming Intervention

The mainstream college-level statistics course (Stat-WS) was delivered in the fall of 2013 and lasted between 3 and 6 hours per week, depending on the institution. Students were taught probability, binomial probability distributions, normal distributions, confidence intervals, and hypothesis testing.

The course required students to attend a 2-hour workshop every week for supplemental instruction on statistical concepts and problems. The workshops, which were taught by advanced undergraduates or recent graduates of CUNY, had three components: (1) 10–15 minutes of reflection on concepts learned so far and what was difficult; (2) about 100 minutes of individual and group work on difficult topics and problems, and (3) a final 5 minutes of reflection on whether the difficult issues were addressed or not.

Students who completed the college-level statistics class fulfilled their institution's quantitative graduation requirement.

What did the study find?

The study found, and the WWC confirmed, statistically significant and positive effects of mainstreaming on course passing rates (effect size = 0.30). Students in the Stat-WS group had a 12% higher passing rate than students in the EA-WS group. The study also found that students in the Stat-WS group had earned 5.4 additional credits on average than the EA-WS group by the end of the fall 2014 semester (i.e., 1 year after the end of the intervention; effect size = 0.37). This finding was statistically significant. The Stat-WS group also had higher enrollment rates in the fall 2014 semester (60% for Stat-WS vs. 51% for EA-WS), and this finding was statistically significant as well (effect size = 0.24).

More detail about these findings is presented in Appendix C. Supplemental findings not reported here are shown in Appendix D.

Appendix A: Study details

Logue, A. W., Watanabe-Rose, M., & Douglas, D. (2016). Should students assessed as needing remedial mathematics take college-level quantitative courses instead? A randomized controlled trial. *Educational Evaluation and Policy Analysis, 38,* 578–598. http://dx.doi.org/10.312/0162373716649056

Additional source:

Logue, A. W., Watanabe-Rose, M., & Douglas, D. (2015, April). *Elementary algebra or statistics: A randomized controlled trial with students assessed as needing remedial mathematics.* Paper presented at the meeting of the American Educational Research Association, Chicago, IL.

Setting

The study was conducted at three City University of New York (CUNY) community colleges, one in each of three New York City boroughs: the Bronx, Queens and Manhattan.

Study sample

The study sample included 907 first-time college freshmen assessed as needing remedial elementary algebra and who were majoring in disciplines that did not require algebra. Within each community college, students who consented to be in the study were randomly assigned to one of three interventions: (1) a mainstream, credit-bearing college-level statistics course with weekly workshops (Stat-WS; n = 297 students), (2) a non-credit bearing remedial elementary algebra course with weekly workshops (EA-WS; n = 313 students), and (3) a non-credit bearing remedial elementary algebra course without workshops (EA; n = 297 students). Twelve instructors participated in the study, and each instructor taught one section of each study condition (i.e., Stat-WS, EA-WS, and EA).

Across all three groups, 55% of students were female (55% in Stat-WS, 58% in EA-WS, and 51% in EA), and 86% of students were underrepresented minorities (84% in Stat-WS, 88% in EA-WS, and 87% in EA). In each of the three groups, the majority of students (56%) reported that their first language was English. The average age of study participants was 21 years old.

Intervention group

The intervention (Stat-WS) was a mainstream, credit-bearing, college-level introductory statistics course, delivered in the fall of 2013. Course topics included probability, binomial probability distributions, normal distributions, confidence intervals, and hypothesis testing. The course was held over one semester and lasted between 3 and 6 hours per week, depending on the college. The course required students to attend a 2-hour workshop every week for supplemental instruction on statistical concepts and problems. The workshops had three components: (1) 10–15 minutes of reflection on concepts learned so far and what was difficult; (2) about 100 minutes of individual and group work on difficult topics and problems, and (3) a final 5 minutes of reflection on whether the difficult issues were addressed or not. The 24 class sections that included workshops were taught by 21 workshop leaders, who were either advanced undergraduates or recent graduates of CUNY.

Comparison group

The primary comparison group was a traditional non-credit-bearing remedial algebra course that included supplemental weekly workshops (EA-WS). The course covered topics such as linear equations, exponents, polynomials, and quadratic equations. Students in both comparison groups took the mandatory CUNY-wide elementary final and received their grade based on a CUNY-wide elementary algebra-grading rubric. The weekly workshops delivered to the EA-WS group followed the same three-component structure implemented for the workshops in the intervention group.

Outcomes and measurement

Since the focus of the article is on the effect of mainstreaming, the most direct test of this intervention is the comparison between the Stat-WS group and the EA-WS group. Findings from the other contrasts (i.e., Stat-WS vs. EA and EA-WS vs. EA) are presented as supplementary outcomes in Appendix D.

The study reported findings on three primary outcomes, which are presented in Appendix C. For the academic achievement domain, the primary outcome was the passing rate of students assigned to a given course in the fall 2013 semester. For the credit accumulation and persistence domain, the primary outcome measures were the total number of college credits earned by the end of the fall 2014 semester (i.e., three semesters after implementation of the intervention) and fall-to-fall enrollment persistence (i.e., from fall 2013 to fall 2014). For a more detailed description of these outcome measures, see Appendix B.

The study also reported findings on three supplemental outcomes that are presented in Appendix D: (1) total number of college credits earned excluding statistics during the intervention's semester plus the calendar year following the end of the intervention (credit accumulation and persistence domain); (2) mean number of courses passed out of two other STEM courses needed to meet CUNY's graduation requirement by the end of the fall 2014 semester, one calendar year after implementation of the intervention (academic achievement domain); and (3) mean number of courses passed out of six non-STEM categories needed to meet CUNY's graduation requirement by the end of the fall 2014 semester, one calendar year after implementation of the intervention (academic achievement domain). The supplemental findings do not factor into the study's rating of effectiveness.

The study includes a measure that indicates whether or not students passed a college-level quantitative course one calendar year after implementation of the intervention. This outcome is endogenous to the intervention since students who were in the Stat-WS intervention had a much greater opportunity to satisfy the course requirement than did students in the comparison group (i.e., they could satisfy this outcome by passing the present course, while students in the EA and EA-WS groups could only enroll in college-level quantitative courses after they passed their remedial course). Therefore, the measure is ineligible for inclusion in the review.

Additionally, several measures of students' attitudes about mathematics were assessed, including students' reports of their mathematical ability, interest, growth, and utility. These outcomes are not eligible for review under any protocol-specified outcome domain.

Support for implementation

Instructors attended a 6-hour orientation workshop, met monthly with researchers, and met weekly with the workshop leaders assigned to their two sections (Stat-WS and EA-WS). The study's 21 workshop leaders completed 10 hours of training focused on the details of the study as well as methods for conducting the workshops. Workshop leaders also met monthly with the researchers to discuss concerns and other issues as needed.

Reason for review

This study was identified for review due to significant media attention.

Appendix B: Outcome measures for each domain

Academic achievement	
Passed course	Academic achievement was measured by whether students passed their assigned course during the fall 2013 semester. Passing rates for math classes were reported for students whether or not they were enrolled in their assigned course at the end of the semester. Students who did not start any math course in the fall of 2013 were counted as not passing.
Number of STEM courses passed	To satisfy CUNY's general education graduation requirement, students had to pass two other college-level courses in science, technology, or engineering, excluding mathematics. Examples of courses that qualified for this requirement include courses in the natural and physical sciences or computer science. This measure is reported as the mean number of STEM courses passed to satisfy this requirement by the end of the fall 2014 semester, one calendar year after the intervention was implemented. This outcome is supplemental and is not included in calculating the effectiveness of the intervention.
Number of non-STEM courses passed	CUNY also has a graduation requirement for students to pass six college-level courses in non-STEM subjects. Courses in the social sciences, arts, and humanities could satisfy this requirement. This measure is the mean number of non-STEM courses passed by the end of the fall 2014 semester; one calendar year after the intervention was implemented. This outcome is supplemental and is not included in calculating the effectiveness of the intervention.
Credit accumulation and persistence	
Total number of college credits earned	Credit accumulation was measured by adding the credits earned during the experiment's semester and the calendar year following the end of the experiment (i.e., from fall 2013 through fall 2014).
Total number of college credits earned, excluding statistics	This measure is a sum of the total number of college credits earned, excluding statistics, during the experiment's semester and the calendar year following the end of the experiment (i.e., from fall 2013 through fall 2014). This outcome is supplemental and is not included in calculating the effectiveness of the intervention.
Fall-to-fall enrollment persistence	This binary outcome measure indicates whether a student was enrolled in fall 2014 semester.

Table Notes: The study includes a measure that indicates whether students passed a college-level quantitative course by the end of the fall 2014 semester, one calendar year after implementation of the intervention. This outcome is endogenous to the intervention since Stat-WS students had the opportunity to pass the course and satisfy this course requirement during the fall 2013 semester while students in the comparison conditions were in non-credit bearing courses and could only fulfill this requirement after the end of the fall 2013 semester. Therefore, the measure is ineligible for inclusion in the review. In addition, several attitudinal outcome measures were assessed in the study, including students' report of their mathematical ability, interest, growth, and utility. These outcomes are not eligible under any WWC protocol-specified outcome domain.

Appendix C: Study findings for each domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Academic achievement								
Passed course (%)	Stat-WS vs. EA-WS	610 students	48 (na)	36 (na)	12	0.30	+12	< .01
Domain average for academic achievement						**0.30**	**+12**	**Statistically significant**
Credit accumulation and persistence								
Fall-to-fall enrollment persistence (%)	Stat-WS vs. EA-WS	610 students	60.30 (na)	50.50 (na)	9.80	0.24	+10	.02
Total number of college credits earned	Stat-WS vs. EA-WS	610 students	20.04 (15.33)	14.66 (13.90)	5.38	0.37	+14	< .001
Domain average for credit accumulation and persistence						**0.30**	**+12**	**Statistically significant**

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on individual outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of the study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding. na = not applicable.

Study Notes: A correction for multiple comparisons was needed for the two outcomes in the credit accumulation and persistence domain, but did not affect whether any of the contrasts were found to be statistically significant. The WWC did not need to make corrections for clustering or to adjust for baseline differences. The mean difference is the covariate adjusted impact, as reported by the author. The comparison group mean is the covariate adjusted mean, as reported by the author. The WWC computed the intervention group mean as the sum of the mean difference and comparison group mean. The *p*-values presented here were reported in the original study, and the *p*-value for fall-to-fall enrollment persistence was obtained through an author query. Means for fall-to-fall enrollment persistence and standard deviations for the total number of college credits earned were also obtained via an author query. Study authors provided the WWC with updated impact estimates for course passing rates after discovering that two students from the Stat-WS group had their grades changed from "D" to "Not Completed" (a failing grade) after the end of the semester. This study is characterized as having a statistically significant positive effect in the academic achievement domain because the estimated effect is positive and statistically significant. The study is characterized as having a statistically significant positive effect in the credit accumulation and persistence domain because the mean effect reported is positive and statistically significant. For more information, please refer to the WWC Standards and Procedures Handbook (version 3.0), p. 26.

Appendix D: Supplemental findings by domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Academic achievement								
Number of STEM courses passed	Stat-WS vs. EA-WS	610 students	0.40 (0.64)	0.33 (0.57)	0.07	0.12	+5	> .05
Number of non-STEM courses passed	Stat-WS vs. EA-WS	610 students	2.19 (2.01)	1.67 (1.96)	0.52	0.26	+10	> .05
Passed course (%)	Stat-WS vs. EA	594 students	48 (na)	31 (na)	17	0.44	+17	< .01
Number of STEM courses passed	Stat-WS vs. EA	594 students	0.40 (0.64)	0.44 (0.68)	–0.04	–0.06	-2	> .05
Number of non-STEM courses passed	Stat-WS vs. EA	594 students	2.19 (2.01)	2.00 (1.93)	0.19	0.10	+4	> .05
Passed course (%)	EA-WS vs. EA	610 students	36 (na)	31 (na)	5	0.14	+5	> .05
Number of STEM courses passed	EA-WS vs. EA	610 students	0.33 (0.57)	0.44 (0.68)	–0.11	–0.18	–7	> .05
Number of non-STEM courses passed	EA-WS vs. EA	610 students	1.67 (1.96)	2.00 (1.93)	–0.33	–0.17	–7	> .05
Credit accumulation and persistence								
Total number of college credits earned, excluding statistics	Stat-WS vs. EA-WS	610 students	18.54 (14.59)	14.66 (13.90)	3.88	0.27	+11	< .001
Fall-to-fall enrollment persistence (%)	Stat-WS vs. EA	594 students	60.30 (na)	54.90 (na)	5.40	0.13	–5	.18
Total number of college credits earned	Stat-WS vs. EA	594 students	19.93 (15.33)	15.53 (13.79)	4.40	0.30	+12	< .01
Total number of college credits earned, excluding statistics	Stat-WS vs. EA	594 students	18.40 (14.59)	15.53 (13.79)	2.87	0.20	+8	< .01
Fall-to-fall enrollment persistence (%)	EA-WS vs. EA	610 students	50.50 (na)	54.90 (na)	–4.40	–0.11	–4	.28
Total number of college credits earned	EA-WS vs. EA	610 students	14.66 (13.90)	15.53 (13.79)	–0.87	–0.06	–3	> .05
Total number of college credits earned, excluding statistics	EA-WS vs. EA	610 students	14.66 (13.90)	15.53 (13.79)	–0.87	–0.06	–3	> .05

Table Notes: The supplemental findings presented in this table are additional findings that meet WWC design standards without reservations, but do not factor into the determination of the study rating. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on individual outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding. na = not applicable.

Study Notes: A correction for multiple comparisons was needed for all outcomes listed above, but did not affect whether any of the contrasts were found to be statistically significant. The WWC did not need to make corrections for clustering or to adjust for baseline differences. The mean difference is the covariate adjusted impact, as reported by the author. The comparison group mean is the covariate adjusted mean, as reported by the author. The WWC computed the intervention group mean as the sum of the mean difference and comparison group mean. The *p*-values presented here were reported in the original study, and exact *p*-values for fall-to-fall enrollment persistence were obtained through an author query. Unadjusted means for fall-to-fall enrollment persistence, and standard deviations for number of non-STEM courses passed, total credits earned, total credits earned excluding statistics were also obtained via an author query. Study authors provided the WWC updated impact estimates for course passing rates after discovering that two students from the Stat-WS group had their grades changed from "D" to "Not Completed" (a failing grade) after the end of the semester.

Endnotes

[1] Single study reviews examine evidence published in a study (supplemented, if necessary, by information obtained directly from the authors) to assess whether the study design meets WWC group design standards. The review reports the WWC's assessment of whether the study meets WWC group design standards and summarizes the study findings following WWC conventions for reporting evidence on effectiveness. This study was reviewed using the Review Protocol for Studies of Interventions for Developmental Students in Postsecondary Education (version 3.1). The WWC rating applies only to the study outcomes that were eligible for review under this topic area. The reported analyses in this Single Study Review are only for those eligible outcomes that either met WWC group design standards without reservations or met WWC group design standards with reservations, and do not necessarily apply to all results presented in the study.

[2] Bailey, Jeong, & Cho (2010), as cited in Logue, Watanabe-Rose, & Douglas (2016).

[3] There were four outcomes included in the study that are not described in this WWC report. See the table notes in Appendix B for more information.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2018, March). *WWC review of the report: Should students assessed as needing remedial mathematics take college-level quantitative courses instead? A randomized controlled trial.* Retrieved from https://whatworks.ed.gov

Glossary of Terms

Attrition Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.

Clustering adjustment If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.

Confounding factor A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.

Design The design of a study is the method by which intervention and comparison groups were assigned.

Domain A domain is a group of closely related outcomes.

Effect size The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.

Eligibility A study is eligible for review if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.

Equivalence A demonstration that the analytic sample groups are similar on observed characteristics defined in the review area protocol.

Improvement index Along a percentile distribution of individuals, the improvement index represents the gain or loss of the average individual due to the intervention. As the average individual starts at the 50th percentile, the measure ranges from –50 to +50.

Multiple comparison adjustment When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.

Quasi-experimental design (QED) A quasi-experimental design (QED) is a research design in which study participants are assigned to intervention and comparison groups through a process that is not random.

Randomized controlled trial (RCT) A randomized controlled trial (RCT) is an experiment in which eligible study participants are randomly assigned to intervention and comparison groups.

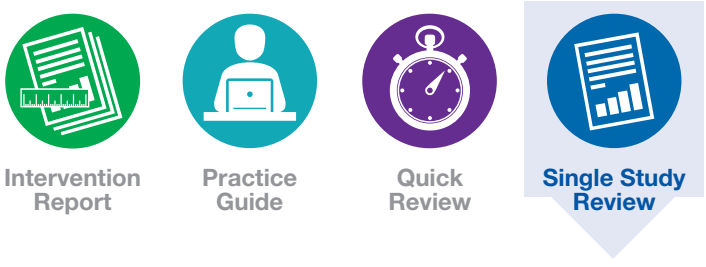
Single-case design (SCD) A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.

Standard deviation The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample are spread out over a large range of values.

Statistical significance Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < .05$).

Substantively important A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the WWC Procedures and Standards Handbook (version 3.0) for additional details.



A **single study review** of an individual study includes the WWC's assessment of the quality of the research design and technical details about the study's design and findings.

This single study review was prepared for the WWC by Abt Associates under contract ED-IES-16-C-0024.