

WWC Review of the Report “Are Tenure Track Professors Better Teachers?”¹

The findings from this review do not reflect the full body of research evidence on having tenure track versus non-tenure track faculty for first-term freshman-level courses.

What is this study about?

The study examined whether taking a course with a tenure track professor versus a non-tenure track professor² for first-term freshman-level courses (e.g., introductory economics) had an impact on students’ future enrollment and performance in classes in the same subject.

The authors used data from 15,662 students who entered Northwestern University, IL, as freshmen between fall 2001 and fall 2008. The intervention group was comprised of students who took a first-term freshman-level class with a tenure track professor. The comparison group was comprised of students who took a first-term freshman-level class with a non-tenure track professor. The study examined the impact of having a tenure track professor versus a non-tenure track professor on students’ future course enrollment and performance in subsequent classes in the same subject.

The authors used transcript data to examine the primary outcomes of enrollment in a subsequent class in the same subject area and grade point average in the next class taken in the same subject area (see Appendix B for more information on the outcome measures). The study did not measure any immediate academic achievement outcomes during the first-term freshman semester.

What did the study find?

The authors investigated the impact of having a tenure track professor (vs. a non-tenure track professor) in a first-term freshman-level class on two outcomes. The authors reported, and the WWC confirmed, that students who took a freshman-level class with a tenure

Features of Having Tenure Track Faculty Teach First-Term Freshman-Level Courses

The study used a quasi-experimental research design to compare two groups that were formed by self-selection: students who enrolled in a first-term freshman-level class taught by a tenure track faculty member versus those enrolled in similar courses taught by a non-tenure track faculty member.

WWC Rating

The research described in this report meets WWC evidence standards with reservations

This study used a quasi-experimental design and established baseline equivalence between groups on student SAT scores.

Although the study established baseline equivalence between groups on student SAT scores, students were not randomly assigned to the intervention and comparison conditions. Therefore, evidence in this study meets WWC standards with reservations.

track professor were statistically significantly less likely to take another class in the same subject (approximately 72% for students taking the introductory course with a tenure track professor, compared to 79% for students taking the course with a non-tenure track professor). In addition, the authors reported, and the WWC confirmed, that among students who did take another class in the same subject, those whose introductory course was taken with a tenure track professor earned slightly (but statistically significantly) lower grades (about one-tenth of a grade point, e.g., 3.1 to 3.0 on a 0–4.0 scale).

Appendix A: Study details

Figlio, D. N., Schapiro, M. O., & Soter, K. B. (2013). *Are tenure track professors better teachers?* (NBER Working Paper 19406). Cambridge, MA: National Bureau of Economic Research.

Setting	The study took place at Northwestern University, IL, a mid-sized private research university that consistently ranks among the most selective undergraduate institutions in the United States.
Study sample	The sample included 15,662 first-term freshmen who entered Northwestern University between fall 2001 and fall 2008. The study used a quasi-experimental research design to compare groups of students that were formed by self-selection: students who enrolled in at least one first-term freshman-level class taught by a tenure track professor ($n = 12,518$) versus those enrolled in similar courses taught by a non-tenure track professor ($n = 3,144$). The demographic composition of the analytic sample is unknown. However, in general, students in this sample were high achievers, with an average SAT score of 1392 (this score indicates that the typical student in this study had a higher SAT score than about 95% of all students taking the SAT).
Intervention group	The intervention group was comprised of students who took a first-term freshman-level class with a tenure track professor. Tenure track professors were those who were identified by individual academic departments, or the department of human resources at Northwestern University, as being tenure track or tenured.
Comparison group	The comparison group was comprised of students who took a first-term freshman-level class with a non-tenure track professor. Although non-tenure track professors could include temporary lecturers and adjuncts, the authors note that almost all classes taught by non-tenure track professors were taught by those with longer-term relationships with the university. The comparison group did not include graduate students or visiting professors who held faculty appointments at other institutions.
Outcomes and measurement	The authors reported findings for two eligible outcomes: whether students enrolled in the next class in a subject, and their grade in the next class taken in the subject. The authors also reported sensitivity analyses for two similar outcomes: whether students enrolled in the next class in a subject only for courses outside of students' intended majors, and their grade in the next class taken in the subject only for courses outside of students' intended majors. All outcome data came from student transcripts from the registrar's office at Northwestern University. For a more detailed description of these outcome measures, see Appendix B.
Support for implementation	Intervention implementation information was not applicable because the intervention was whether students took courses with tenure track versus non-tenure track professors.
Reason for review	This study was identified for review by the WWC because it received significant media attention.

Appendix B: Outcome measures for each domain

Enrollment	
<i>Enrollment in another class in subject (all classes)</i>	This outcome is based on data from student transcripts received from the registrar's office at Northwestern University. This binary outcome measures whether students enrolled in a class in the same subject as their first-term freshman-level introductory course. For this analysis, the study authors included enrollment in all subsequent classes in a subject, regardless of whether the class was within or outside of students' intended majors.
<i>Enrollment in another class in subject (non-major classes)</i>	This outcome was used for sensitivity analyses and is based on data from student transcripts received from the registrar's office at Northwestern University. This binary outcome measures whether students enrolled in a class in the same subject as their first-term freshman-level introductory course. For this analysis, the study authors included enrollment in only those classes in a subject that were outside of students' intended majors.
Academic achievement	
<i>Grades in another class in subject (all classes)</i>	This outcome is based on data from student transcripts received from the registrar's office at Northwestern University. The outcome measures students' grades in the next class they enrolled in that was in the same subject as their first-term freshman-level introductory course, and ranges from 0 (F-) to 4.0 (A+). For this analysis, the study authors included grades in all subsequent classes in a subject, regardless of whether the class was within or outside of students' intended majors.
<i>Grades in another class in subject (non-major classes)</i>	This outcome was used for sensitivity analyses and is based on data from student transcripts received from the registrar's office at Northwestern University. The outcome measures students' grades in the next class they enrolled in that was in the same subject as their first-term freshman-level introductory course, and ranges from 0 (F-) to 4.0 (A+). For this analysis, the study authors included grades in only those classes in a subject that were outside of students' intended majors.

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	
Enrollment								
<i>Enrollment in another class in subject (all classes)</i>	College students	15,662 students	nr	nr	7.3%	0.29	+11	< 0.01
<i>Enrollment in another class in subject (non-major classes)</i>	College students	15,661 students	nr	nr	12.0%	0.31	+12	< 0.01
Domain average for enrollment						0.30	+12	Statistically significant
Academic achievement								
<i>Grade in another class in subject (all classes)</i>	College students	11,579 students	nr	nr	0.06	0.20	+8	< 0.01
<i>Grade in another class in subject (non-major classes)</i>	College students	11,412 students	nr	nr	0.08	0.23	+9	< 0.01
Domain average for academic achievement						0.22	+9	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 student outcomes, representing the average change expected for all students 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 student’s percentile rank that can be expected if the student 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; for example, the study is characterized as having a statistically significant positive effect because univariate statistical tests are reported for each outcome measure, the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant, accounting for multiple comparisons. In both outcome domains, the study authors measured outcomes for all subsequent classes in a subject area (the primary outcome of interest), and for classes outside of students’ majors (for sensitivity analysis). This table includes findings for all outcomes, but the single study review focuses on the overall analyses focusing on all subsequent classes. nr = not reported.

Study Notes: A correction for multiple comparisons was needed and resulted in a WWC-computed p-value of < 0.01 for all four outcomes; therefore, the WWC confirmed that all results were statistically significant. The p-values presented here were reported in the original study.

Endnotes

¹ Single study reviews examine evidence published in a study (supplemented, if necessary, by information obtained directly from the author[s]) to assess whether the study design meets WWC evidence standards. The review reports the WWC's assessment of whether the study meets WWC evidence standards and summarizes the study findings following WWC conventions for reporting evidence on effectiveness. This study was reviewed using the Postsecondary Education topic area review protocol, version 2.0. A quick review of this study was released on December 5, 2013, and this report is the follow-up review that replaces that initial assessment. The WWC rating applies only to the results that were eligible under this topic area and met WWC standards with reservations, and not necessarily to all results presented in the study.

² This single study review uses the term “tenure track” to refer to professors who are either on the tenure track (e.g., are not yet tenured but are eligible to be considered for tenure) or are already tenured. The term “non-tenure track” refers to professors who are not tenured and are not eligible for tenure consideration.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2014, April). *WWC review of the report: Are tenure track professors better teachers?* Retrieved from <http://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 analysis sample groups are similar on observed characteristics defined in the review area protocol.
Improvement index	Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student 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 subjects 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 investigators randomly assign eligible participants into 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% ($p < 0.05$).
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 \(version 2.1\)](#) for additional details.