

## WWC Review of the Report “Daily Online Testing in Large Classes: Boosting College Performance While Reducing Achievement Gaps”<sup>1,2</sup>

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

### What is this study about?

The study examined the impact of frequent online testing and immediate feedback on the academic performance of college students.

Students who registered in two large Introductory Psychology courses at the University of Texas in fall 2011 (n = 982) took short internet-based quizzes at the beginning of every class. After each test, students were provided with immediate feedback on their performance.

Two Introductory Psychology classes at the same university in fall 2008 served as the comparison group (n = 993). Students in the comparison condition were taught in the traditional manner, with lectures and an assigned textbook. They were assessed with four class-long exams over course of the semester.

Study authors examined the program’s impact on student GPA in courses other than Introductory Psychology in the semester during which students were enrolled in Introductory Psychology and overall student GPA in the semester following the Introductory Psychology course. GPA was provided by the University of Texas registrar’s office.<sup>3</sup>

### What did students in the intervention and comparison classes experience?

Intervention classes had the same lecture format as comparison classes. However, students in the intervention group spent the first 10 minutes of each class (in 26 of 28 classes) taking an in-class computerized quiz that covered material from the assigned reading and the previous lecture. After each quiz, students were provided with immediate feedback on their performances.

Students in the intervention group also used readings from online sources instead of an assigned textbook and participated in occasional in-class virtual discussions instead of small group discussions.

### What did the study find?

The study reported that students in the intervention condition earned higher grades in their other courses during both the fall semester and the subsequent spring semester. The authors also reported that the intervention's impact was significantly larger for students whose parents attended "some college or less" than for students whose parents had "some graduate work." The WWC could not confirm the statistical significance of either of these results and does not consider these results to be conclusive because the intervention and comparison groups were not shown to be equivalent at baseline on any of the outcome variables. The reported differences might reflect baseline differences between the two groups, rather than the impact of frequent online testing on student achievement.

### WWC Rating

#### ***The research described in this report does not meet WWC group design standards***

Because students were not randomly assigned to the intervention and comparison groups, the study needed to demonstrate equivalence on baseline measures of the outcomes to meet WWC group design standards with reservations. This information was not available. Therefore, any differences in student outcomes between intervention and comparison groups cannot be attributed solely to the implementation of the intervention.

### Endnotes

<sup>1</sup> Full study citation: Pennebaker, J. W., Gosling, S. D., & Ferrell, J. D. (2013). Daily online testing in large classes: Boosting college performance while reducing achievement gaps. *PLoS ONE*, 8(11), e79774. doi:10.1371/journal.pone.0079774

<sup>2</sup> 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 WWC review protocol for individual studies in the Postsecondary Education topic area, version 2.0. A quick review of this study was released on December 20, 2013, and this report is the follow-up review that replaces that initial assessment.

<sup>3</sup> There were four outcomes included in the study that are not described in this WWC report. Three outcomes—class attendance, instructor course evaluation, and test performance—were not eligible outcomes under the WWC review protocol for individual studies in the Postsecondary Education topic area. The fourth outcome, overall grade in the Introductory Psychology course, did not meet WWC review requirements because it was measured differently across intervention and comparison conditions.

### Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2014, May). *WWC review of the report: Daily online testing in large classes: Boosting college performance while reducing achievement gaps*. Retrieved from <http://whatworks.ed.gov>

### Glossary of Terms

<b>Attrition</b>	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.
<b>Clustering adjustment</b>	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.
<b>Confounding factor</b>	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.
<b>Design</b>	The design of a study is the method by which intervention and comparison groups were assigned.
<b>Domain</b>	A domain is a group of closely related outcomes.
<b>Effect size</b>	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.
<b>Eligibility</b>	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.
<b>Equivalence</b>	A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.
<b>Improvement index</b>	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.
<b>Multiple comparison adjustment</b>	When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.
<b>Quasi-experimental design (QED)</b>	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.
<b>Randomized controlled trial (RCT)</b>	A randomized controlled trial (RCT) is an experiment in which investigators randomly assign eligible participants into intervention and comparison groups.
<b>Single-case design (SCD)</b>	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.
<b>Standard deviation</b>	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.
<b>Statistical significance</b>	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% ( $p < 0.05$ ).
<b>Substantively important</b>	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.