What Works Clearinghouse™

ies INSTITUTE OF EDUCATION SCIENCES

May 2013

# WWC Review of the Report "Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment"[1]

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

## What is this study about?

The study examined whether four different types of teacher incentive strategies affected students' math achievement. Authors experimented with the idea of "loss aversion," or people's preference to avoid loss rather than acquire gains, and individual and team responsibility.

Researchers analyzed student and teacher data from nine K–8 public schools in Chicago Heights, Illinois in 2010–11.

Within each of the nine participating schools, 147 teachers were randomly assigned either to one of four intervention groups with different incentive strategies or to a comparison group that received no incentives. See the blue box for more details on the four different incentive strategies used in the intervention groups.

Researchers examined the effect of each type of incentive by comparing the test scores of students whose teachers were in one of the intervention groups to those of students whose teachers were in the comparison group.

### Features of the Teacher Incentives Examined in this Study

Four different incentive models were used in this study. In all four, teachers earned between $0 and $8,000 based on student achievement.

- Teachers in the "individual loss aversion group" received $4,000 at the beginning of the school year. If their own students' end-of-year performance was above average, the teachers were given an additional payment of up to $4,000. If their students' performance was below average, they were required to return the difference between $4,000 and their earned bonus.
- Teachers in the "team loss aversion group" received their bonuses the same way as the individual loss aversion group, but their earned bonus amount was based on the performance of both their own students and those of a similar teacher in the same school, grade, and subject.
- Teachers in the "individual gain group" received their total bonuses all at once at the end of the school year. The bonus amount was determined by their own students' achievement.
- Finally, teachers in the "team gain group" also received their total bonuses at the end of the school year, but the bonus amount was determined by the performance of both their own students and those of a similar teacher in the same school, grade, and subject.

The net payments and the criteria for award levels were the same across the four types of intervention groups.

## What did the study find?

The study authors found that offering teacher incentives using loss aversion had statistically significant positive effects on student math achievement; this was true for both the individual and team loss aversion groups (effect sizes ranging from 0.15 to 0.27).

The study authors did not find an effect of either of the two gain groups on student math achievement.

WWC Rating

***The research described in this report meets WWC evidence standards without reservations***

***Strengths:*** This study is a well-implemented randomized controlled trial.

## Appendix A: Study details

**Fryer, R. G., Levitt, S. D., List, J., & Sadoff, S. (2012). *Enhancing the efficacy of teacher incentives through loss aversion: A field experiment (Working Paper 18237)*. Cambridge, MA: National Bureau of Economic Research.**

### Setting

The study was conducted in the Chicago Heights school district located thirty miles south of Chicago.

### Study sample

Nine high-poverty K–8 schools were included in the random assignment process. Approximately 160 teachers were eligible to participate in the experiment, and 150 of them volunteered to participate. Teachers' homerooms were randomly assigned either to a comparison group (37 homerooms) or to one of four intervention groups: individual loss aversion (34), team loss aversion (28), individual gain (24), or team gain (24). In most cases, a teacher has one homeroom, but at the middle school level, teachers could have multiple homerooms; the students of teachers with multiple homerooms were grouped by class and subject and then randomized into one of the five groups. All the teachers and their homerooms were included in the analysis as part of their original randomly assigned condition, regardless of whether they ultimately participated in the incentive program.

### Intervention group

Teachers in each of the four intervention groups were eligible for incentive payments tied to student performance on end-of-year tests. For each student in the study, the authors identified the nine students who had the most similar pre-intervention test scores; these nine students could be in any of the other eight schools participating in the study, but not the student's own school. These ten students composed a "bin." The authors then computed the difference between each student's end-of-year test score and pre-intervention test score and ranked each student in the bin according to the amount of improvement each demonstrated; the authors refer to this as a "percentile rank." Finally, all of a teacher's students' percentile ranks were averaged to determine an "overall percentile" for each teacher; each percentile was worth $80, for a maximum possible reward of $8,000.

The way in which the incentives were distributed varied across the groups. In the two loss aversion intervention groups, teachers received $4,000 at the beginning of the school year. If a teacher's reward at the end of the school year exceeded $4,000, he or she received another payment for the difference. If a teacher's reward at the end of the school year was less than $4,000, he or she was required to return the difference between $4,000 and his or her final reward amount. Teachers in the two gain intervention groups received their bonus based on the same criteria as the loss aversion intervention groups, but the incentive payments were given all at once at the end of the school year. The net payments and the criteria for award levels were the same across the four types of intervention groups.

The loss aversion and gain groups were further divided by the students on which the incentive payments were based. In the individual reward group, a teacher's reward was based on the performance of all of his or her own students. In the team reward group, the reward was based on the performance of the teacher's own students and those of a similar teacher in the same school, grade, and subject.

| | |
|---|---|
| **Comparison group** | Comparison group teachers were not offered the opportunity to participate in the incentive program. |
| **Outcomes and measurement** | Study authors examined student math performance on ThinkLink Predictive Assessment for grades K–8 and state standardized tests (the Illinois Standards Achievement Test [ISAT] for grades 3–8, and the Iowa Test of Basic Skills [ITBS] for grades K–2). For a more detailed description of these outcome measures, see Appendix B. |
| **Support for implementation** | Teachers could adapt their regular practices as they thought was appropriate to attain their incentive payments. No direct instructional support was provided as part of the program. |
| **Reason for review** | This study was identified for review by the WWC by receiving significant media attention. |

## Appendix B: Outcome measures for the mathematics achievement domain

| Mathematics achievement | |
|---|---|
| *Illinois Standards Achievement Test (ISAT)* | The ISAT mathematics achievement test was administered to students in grades 3–8 in Illinois public schools. The math test included items on arithmetic, measurement, algebra, geometry, statistics, probability, and other areas. Scores were obtained from student-level administrative records from the school district. They were standardized by grade level and academic year to have a mean of zero and standard deviation of one. |
| *Iowa Test of Basic Skills (ITBS)* | The ITBS was administered to students in grades K–2 in the nine schools participating in the study. The math test included items on numeration, geometry, measurement, and problem solving using addition and subtraction for the younger students. All questions were presented orally; responses were pictures or numerals. Scores were obtained from student-level administrative records from the school district. They were standardized by grade level and academic year to have a mean of zero and standard deviation of one. |
| *ThinkLink Predictive Assessment* | The ThinkLink Predictive Assessment is a low-stakes standardized diagnostic assessment that is typically used to monitor whether students hit benchmarks and predict how well they will do on the standardized state tests. It was administered to students in grades 3–8 in September, November, January, and May of the 2010–11 school year. Scores were obtained from student-level administrative records from the school district. They were standardized by grade level and academic year to have a mean of zero and standard deviation of one. |

**Table Notes:** Some outcomes were included in the study but are not included in this review. These include: teacher survey results (excluded because they examine attitudes about the program and not effectiveness of the program) and reading results on ThinkLink, ISAT, and ITBS tests (excluded because they were presented in an appendix and the authors did not describe them as major outcomes).

## Appendix C: Study findings for the mathematics achievement 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 | |
| **Mathematics achievement, individual loss vs. comparison** | | | | | | | | |
| *Illinois Standards Achievement Test (ISAT)/Iowa Test of Basic Skills (ITBS)* | K–8 schools | 69 homerooms/ 963 students | 0.10 (nr) | –0.07 (nr) | 0.17 | 0.15 | +6 | > 0.05 |
| *ThinkLink Predictive Assessment* | K–8 schools | 70 homerooms/ 1,016 students | 0.19 (nr) | –0.15 (nr) | 0.33 | 0.22 | +9 | < 0.05 |
| **Domain average for mathematics achievement, individual loss vs. comparison** | | | | | | **+0.19** | **+7** | **Statistically significant** |
| **Mathematics achievement, team loss vs. comparison** | | | | | | | | |
| *ISAT/ITBS* | K–8 schools | 63 homerooms/ 900 students | 0.11 (nr) | –0.07 (nr) | 0.18 | 0.27 | +11 | < 0.05 |
| *ThinkLink Predictive Assessment* | K–8 schools | 64 homerooms/ 965 students | 0.07 (nr) | –0.15 (nr) | 0.21 | 0.22 | +9 | < 0.05 |
| **Domain average for mathematics achievement, team loss vs. comparison** | | | | | | **+0.24** | **+10** | **Statistically significant** |
| **Mathematics achievement, individual gain vs. comparison** | | | | | | | | |
| *ISAT/ITBS* | K–8 schools | 59 homerooms/ 917 students | 0.00 (nr) | –0.07 (nr) | 0.07 | –0.04 | –2 | > 0.05 |
| *ThinkLink Predictive Assessment* | K–8 schools | 60 homerooms/ 988 students | 0.02 (nr) | –0.15 (nr) | 0.17 | 0.09 | +4 | > 0.05 |
| **Domain average for mathematics achievement, individual gain vs. comparison** | | | | | | **+0.03** | **+1** | **Not statistically significant** |
| **Mathematics achievement, team gain vs. comparison** | | | | | | | | |
| *ISAT/ITBS* | K–8 schools | 59 homerooms/ 903 students | 0.04 (nr) | –0.07 (nr) | 0.11 | 0.04 | +2 | > 0.05 |
| *ThinkLink Predictive Assessment* | K–8 schools | 60 homerooms/ 981 students | –0.02 (nr) | –0.15 (nr) | 0.13 | 0.09 | +4 | > 0.05 |
| **Domain average for mathematics achievement, team gain vs. comparison** | | | | | | **+0.07** | **+3** | **Not statistically significant** |

**Table Notes:** Positive results for mean difference, effect size, and improvement index favor the intervention group; negative results favor the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the change (measured in standard deviations) in an average student's outcome that can be expected if the student is exposed to the intervention. 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 exposed to 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. The study is characterized as demonstrating positive impacts of both individual and team loss aversion on students' mathematics achievement because at least one finding in each of those domains was statistically significant and positive and none were statistically significant and negative. The study is characterized as demonstrating indeterminate effects of individual and team gain on students' mathematics achievement because no impacts were statistically significant in those domains.

**Study Notes:** The means presented in this table are regression-adjusted, provided by the author at the request of WWC. The authors controlled for student characteristics, including race, gender, age, free-lunch status, English proficiency, special education status, and school and grade fixed effects. No corrections for clustering or multiple comparisons were needed. The effect sizes and *p*-values presented here were reported in the original study, in column 5 of Table 3 (ThinkLink Predictive Assessment outcomes) and column 5 of Table 4 (ISAT/ITBS).

## 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 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 single study review protocol, version 2.0. The WWC rating applies only to the results that were eligible under this topic area and met WWC standards without reservations or met WWC standards with reservations, and not necessarily to all results presented in the study.

[2] Some outcomes were included in the study but are not included in this review. These include: teacher survey results (excluded because they examine attitudes about the program and not effectiveness of the program) and reading results on ThinkLink, ISAT, and ITBS tests (excluded because they were presented in an appendix and the authors did not describe them as major outcomes).

## Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, May). WWC review of the report: *Enhancing the efficacy of teacher incentives through loss aversion: A field experiment*. 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.