

# **Experimental Estimates of the Impacts of Class Size on Test Scores: Robustness and Heterogeneity\***

Weili Ding  
Queen's University

Steven F. Lehrer  
Queen's University and NBER

*June 2011*

## **Abstract**

Proponents of class size reductions draw heavily on the results from Project STAR to support their initiatives. Adding to the political appeal of these initiative are reports that minority and economically disadvantaged students received the largest benefits from smaller classes. We extend this research in two directions. First, to address correlated outcomes from the same class size treatment, we account for the over-rejection of the Null hypotheses by using multiple inference procedures. Second, we conduct a more detailed examination of the heterogeneous impacts of class size reductions on measures of cognitive and noncognitive achievement using more flexible models. We find that students with higher test scores received greater benefits from class size reductions. Furthermore, we present evidence that the main effects of the small class treatment are robust to corrections for the multiple hypotheses being tested. However, these same corrections lead the differential impacts of smaller classes by race and freelunch status to become statistically insignificant.

*JEL codes: I20 , C21, C12 and I24.*

*Keywords: class size; multiple inference; unconditional quantile regression; treatment effect heterogeneity; test score gaps; and education experiment*

---

\* We would like to thank Caroline Hoxby, Richard Murnane, an anonymous referee and seminar participants at the 2010 International Workshop on Applied Economics of Education, XIX Meeting of the Economics of Education Association, Simon Fraser University and the 2004 AEA annual meetings session on Education for the Disadvantaged for helpful comments and suggestions. We are grateful to Alan Krueger for generously providing a subset of the data used in the study. This paper is a revised version of the second chapter of Weili Ding's University of Pittsburgh 2002 thesis. Lehrer wishes to thank SSHRC for research support. We are responsible for all errors.