Online Appendix

A. The Life and Work of an Early Experimentalist

Paul Raymond Stevenson was born in Cedar Rapids, Iowa in 1886, and received a master's degree in education and psychology from the University of Cincinnati in 1914. After teaching mathematics and science for several years, he enrolled in the doctoral program at Teachers College, Columbia University in 1919. Though my efforts to acquire Stevenson's transcript have been unsuccessful, it is highly likely that he studied experimental methods with William McCall during his time at Columbia. Stevenson cites McCall in his dissertation as the source for his methodology, and McCall's book on experimental methods includes a reference to Stevenson's work as an example of his ideas put to good use.

It seems that Teachers College also provided Stevenson with the connection that would allow him to put his training into practice. In 1920, he was hired as a research assistant to B.R. Buckingham, a Teachers College alumnus who served for many years as the chief statistician for the New York City School System. Buckingham had, like William McCall, also worked extensively on the standardized measurement of student achievement. In 1918, Buckingham had taken a professorship at the University of Illinois, and struck an agreement with the Chicago Public Schools to undertake an experimental evaluation of class size. Upon his arrival in Chicago, Stevenson was placed in charge of the study, which became the subject of his doctoral dissertation (awarded by Teachers College in 1925) and led to two widely cited publications (Stevenson 1922, 1923).

In 1922, Stevenson followed Buckingham to Ohio State University and continued his research on class size. A solicitation for the participation of schools—published in the inaugural issue of the Bulletin of the Ohio State Bureau of Education Research— provides a clear exposition of his method. Participating schools would "(1) select two or more classes in a given subject which can be taught by one teacher; (2) give intelligence tests to the pupils; (3) divide the pupils into large and small sections so that the average and variability of the intelligence of the sections will be approximately the same; (4) have one teacher teach both a large and a small section; (5) measure the relative efficiency of

the large and small sections by means of standardized tests and school examinations."
These additional experiments led to two more publications (Stevenson (1925a, 1925b)).

In 1926, Stevenson joined a group of researchers supported by the National Association of Secondary School Principals for the purpose of studying class size on a large scale. However, Stevenson died of a sudden illness in January of 1927, and the new field experiments that were planned with this group (as reported in the NASSP Bulletin, March 1926) seem to have never been completed.

My analysis of Stevenson's experiments in elementary schools uses data on average test score gains at the classroom-semester cell level. These experiments included students in grades two, five, and seven, and students were tested in multiple subjects. For simplicity and power, I pool data across subjects and grades. Because the dependent variable is a classroom average, one might also weight the regression by the number of students that were not rotated during the year, but this has a negligible impact on the results and I present un-weighted estimates. Note that using classroom level gains as the dependent variable might overstate the benefits of a smaller class. If the positive effect of a small class (or the negative effect of a large class) fades out over the following semester, then measuring second semester gains using achievement at the end of the first semester as a baseline will tend to overstate these effects. Unfortunately, Stevenson does not report other measures of achievement by classroom-semester.

The coefficient estimate quoted in the main text is displayed in Column 1 of Table A1 (0.010 with a standard error of 0.055) is quoted in the main text. In Column 2, I add fixed effects for each study-semester-grade-subject cell to control for differences across tests, time effects, etc. Note that, given Stevenson's experimental design, the indicator for large class and these fixed effects should be orthogonal. In practice, Stevenson's tables are missing data for a few class-subject cells, producing a correlation of -.0026 between the large class indicator and semester. Regardless, the point estimate remains essentially the same (0.009) and the standard error falls to 0.046 standard deviations. In Column 3, I add teacher fixed effects to the specification. Again, due to Stevenson's rotation method, each teacher provides instruction to both a small and large class, so that correlation between the large class indicator and teaching quality is ruled out by design. The point estimate is therefore identical, and the standard error falls slightly to 0.044.

The difference in size between large and small classes taught by the same teacher varied within Stevenson's sample. I therefore also report results that use class size as a covariate instead of a large class indicator. Because Stevenson's empirical design does not force all large and small classes to be of the same size, there may be correlation between class size and other factors, making the additional fixed effects potentially more useful. To make the coefficients more easily interpretable, I express class size in units of 10. The results without controls (Column 4) indicate that an increase of 10 students was positively related to test score growth, with a coefficient of 0.017 standard deviations and a standard error of 0.029. The additional controls (Column 5) change things slightly, moving the coefficient and standard error to 0.018 and 0.027, respectively.

Table A1
Effects of Class Size in Elementary School, Data from Stevenson (1923, 1925a)

	(1)	(2)	(3)	(4)	(5)
Large Class	0.010	0.009	0.009		
	(0.055)	(0.046)	(0.045)		
Class Size				0.017	0.018
				(0.029)	(0.027)
R^2	0.00	0.35	0.46	0.00	0.46
Fixed effects included:					
Study-Subject-Grade-Semester		√	√		√
Teacher/Classroom			√		√

Note: The unit of observation is classroom level average test score gain in a semester, and each regression has 774 observations. Subjects include arithmetic, reading comprehension, reading rate, language arts, spelling, grammar, and analysis. Second graders were not given a test of language in Stevenson (1923) and were not given a test of reading comprehension in Stevenson (1925a), fifth and seventh graders were not tested in spelling in Stevenson (1923). Across both studies, there are 32, 26, and 20 classes in, respectively, the second, fifth and seventh grades.

B. Studies of Class Size Prior to 1940

Field Experiments

Almack, J.C. (1922). *The adaptation of the school building to a program of educational efficiency.* Doctoral Dissertation, Stanford University.

Averill, L.A. & Mueller, A.D. (1925). Size of class and reading efficiency. *The Elementary School Journal*, 25(9), 682-691.

Brown, A.E. (1932). The effectiveness of large classes at the college level: An experimental study involving the size variable and size-procedure variable. *University of Iowa Studies in Education,* 7, 1-66.

Cunningham, M.S. (1931). An experiment in class size in B9 Algebra. *California Quarterly of Secondary Education,* 7, 19-33.

Dalthrop, C. J. (1934). An experiment with a large and small class in English composition. *High School Teacher*, 10, 51.

Davis, C.O. (1923). The size of classes and the teaching load in the high schools accredited by the north central association. *The School Review,* 31(6), 412-429.

Davis, E. & Goldzien, M. (1930). A study of class size in junior high school history. *The School Review, 38(5),* 360-367.

Eastburn, L.A. (1936). The relative efficiency of instruction in large and small classes on three ability levels. Doctoral Dissertation, Stanford University.

Eastburn, L. A., (1937). Report of class size investigations in the Phoenix Union High School, 1933-34 to 1935-36. *Journal of Educational Research*, 31(2), 107.

Ewan, S.N. (1935). *The relation of class size and selected teaching methods to pupil achievement.* Doctoral Dissertation, University of Pennsylvania.

Hand, H.C. & Smith, J.W. (1934). Effectiveness of instruction in a class group of one hundred pupils. *The School Review*, 42(10), 751-754.

Haertter, L.D. (1928). An experiment of the efficiency of instruction in large and small classes in plane geometry. *Educational Administration and Supervision,* 14(8), 580-590.

Holy, T.C. (1932). Effect of total teaching load on the efficiency of instruction in high school English. *National Education Association Proceedings,* 365-366.

Hudelson, E. (1928). *Class size at the college level.* Minneapolis: The University of
    Minnesota Press.

    Part of this study was a doctoral dissertation:
    Von Borgersrode, F. (1929). *The relation of class size to the effectiveness of
        college instruction as expressed through a direct measure of student
        achievement*. Doctoral Dissertation, University of Minnesota.
    The study was also summarized in:
    Hudelson, E. (1930). Are classes too large? *The Journal of Higher Education,*
        1(8), 436-439.

Jensen, M.B. (1930). The influence of class size upon pupil accomplishment in high
    school algebra. *Journal of Educational Research*, 21(5), 337-356.

Judd, C.H. (1927). Report of the consultative committee. *Bulletin: Department of
    Secondary School Principals*, 25, 49-61.

Metzner, A.B & Berry, C. (1926). Size of class for mentally retarded children. *Training
    School Bulletin*, (23), 241-251.

Smith, D.V. (1925). *Class size in high school English*. Minneapolis: The University of
    Minnesota Press.

Stevenson, P. R. (1923). *Smaller classes or larger; relation of size of class size to the
    efficiency of teaching.* Bloomington, IL: Public School Publishing Company.

    A somewhat shorter version of these results are reported in:
    Stevenson, P. R. (1922). Relation of size of class to school efficiency. *University
    of Illinois Bulletin*, 14(45).

Stevenson, P.R. (1925a). Class-size in the Elementary School. *Bureau of Educational
    Research Monographs, No.3, 2(10).*

Stevenson, P.R. (1925b). More evidence concerning small and large classes.
    Educational Research Bulletin, 4(11), 231-233.

Stevenson, P.R. (1925). More Evidence Concerning Small and Large Classes.
    Educational Research Bulletin, 4(11), 231-233.

Tope, R.E.; Groom, E.; and Beeson, M.F. (1924). Size of class and school efficiency.
    Journal of Educational Research, 2, 126-132.

Wasson, W.H. (1929). *A controlled experiment in the size of classes.* Master's thesis,
    University of Chicago.

Whitney, F.L. & Willey, G. (1932). Advantages of small classes. *School Executives Magazine: The American Educational Digest*, 51, 504-506.

Analysis of this field experiment is also presented in:
Whitney, F.L., (1931). "The Trinity-Pueblo Class Size Experiment in the Primary School," Department of Educational Research Study #113, Colorado State Teachers College.

Observational Studies (Matching)

Bloomfield, L.S. (1930). Studies in class size: Class size in senior American History. *National Association of Secondary School Principals Bulletin,* 14, 6-9.

Edmonson, J. B. and Mulder, F.J. (1924). Size of class as a factor in university instruction. *Journal of Educational Research*, 9(1), 1-12.

Ewan, S.N. (1932). Class size. *National Association of Secondary School Principals Bulletin,* 16, 124-129.

Kirk, J.R. (1929). A study of class size, teaching efficiency, and student achievement. *Phi Delta Kappan*, 12, 59-61.

McGuire, M. & Hudelson, E. (1930). Learning to teach larger classes. *Educational Supervision and Administration,* 15, 34-38.

Remmers, H.H. (1931) *An experiment on the relative effectiveness of the lecture-quiz and recitation methods of class instruction in elementary psychology.* Unpublished Report, Purdue University.

Remmers, H.H., Hadley, L, & Long, J.K. (1932). Learning, effort, and attitude as affected by class size in beginning college engineering mathematics. *Bulletin of Purdue University,* 32(9).

Taylor, H.R. (1929). An experiment with independent study. In Douglas, N.R. (Ed.) *Controlled experimentation in the study of methods of college teaching.* (pp. 300-312). Eugene, Oregon: University of Oregon Publication, Education Series.

Observational Studies (Correlational)

Bates, D.A. (1928). The relation of the size of class to the efficiency of teaching. *NASSP Bulletin*, 13(22), 44.

Boyer, P.A. (1914). Class size and school progress. *Psychological Clinic*, 8, 82-90.

Breed, F.S., and McCarthy, G.D. (1916). Class size and efficiency of teaching. *School and Society,* 4, 965-971.

Brooks, L.W. (1930). The large class versus the small class by the laboratory method. *National Association of Secondary School Principals Bulletin,* 14, 11-13.

Cornman, O.P. (1909). Size of classes and school progress. *The Psychological Clinic*, 3, 206-212.

Dawe, H.C. (1934). The influence of size of kindergarten group upon performance. *Child Development*, 5(4), 295-303.

Harlan, C.L. (1916). Size of class as a factor in schoolroom efficiency. *Educational Administration and Supervision,* 1, 195-214.

Holland, B.G. (1928). The effect of class size on scholastic acquirement in educational psychology. *School and Society*, 27, 668-670.

Miller, P.S. (1929). A quantitative investigation of the efficiency of instruction in high school Physics. *Journal of Educational Research*, 19, 119-127.

Moss, F.A., Loman, W., & Hunt, T. (1929). Impersonal measurement of teaching. *Educational Record*, 10(1), 40-50.

Rice, J.M. (1918). *Scientific management in education.* New York: Hinds, Noble & Eldredge.

Wetzel, W.A (1930). Teaching technique and size of class. *School Life*, 15, 181-182.

Whitney, F.L. (1929). The study of class size in the primary grades. *Yearbook of the Department of Classroom Teachers,* 4, 95-98.