

The Effect of Testing on Achievement: Meta-Analyses and Research Summary, 1910–2010

Source List, Outcomes, and References for Qualitative Studies

The text of this study is currently under review by a scholarly journal. The study summarizes the research literature on the effect of testing on student achievement, which comprises several hundred studies conducted from the early 20th century to the present day. Only qualitative studies, however, are included here (N = 244).

Qualitative studies overwhelmingly find testing's effect on student achievement to be positive: ninety-three percent of the studies analyzed reported positive effects, whereas only seven percent reported mixed effects, negative effects, or no change.

Author	Year	Method	Participants	Location	Scale	Findings	Rigor
Consultative Committee on Examinations	1910	research review	students	UK	classroom	Positive	High
Woody, Clifford	1917	case study	schools	UT	classroom	Positive inferred	Low
Gray, William S.	1918	experiment or pre-post comparison	students, teachers	IL	classroom	Positive	Medium
Brooks, Samuel S.	1922	interview	teachers, students	NH	classroom	Positive	Low
White, H. B.	1932	experiment or pre-post comparison	college students		classroom	Positive	High
Kulp, D. H., II	1934	experiment or pre-post comparison	college students		classroom	Positive	High
Messenger	1934	experiment or pre-post comparison	teachers	IA	large-scale	Positive inferred	High

Author	Year	Method	Participants	Location	Scale	Findings	Rigor
Scott, I.O.	1934	experiment or pre-post comparison	students	CO	classroom	Positive	High
Boucher, Chauncey Samuel	1935	case study	students, instructors	IL	classroom	Positive	Low
Brereton, J. L.	1944	case study	schools	England	large-scale	Positive	Medium
Stuit, D.B. (Ed)	1947	interview, observation, records/ document review	teachers	US	teacher	Positive	High
Wood, Ray G.	1953	survey	Ohio graduates	OH	large-scale	Positive	Medium
Feldhusen, John F.	1964	survey, interview	students	U.S.	classroom	Positive	High
Estes, Gary D., Colvin, Lloyd W., & Goodwin, Coleen	1976	case study	students	AZ	large-scale	Positive	High
Foss, Olive	1977	interview, survey	faculty	UK	classroom	Positive	High
Solberg, W.	1977	case study	students	Netherlands	large-scale	Positive	Medium
Enochs, James C.	1978	case study	students	CA (Modesto)	large-scale	Positive	Medium
Findley, Jim	1978	case study	teachers, administrators	NE	large-scale	Positive	Medium
Fisher, Thomas H.	1978	case study	students	FL	large-scale	Positive	High
Neill, S. B.	1978	case study	principal	AK	large-scale	Positive	Low
Brookover, W.B., & Lezotte, L.W.	1979	survey, interview, records/document review	teachers	MI	large-scale	Positive	High
Down, A. Graham	1979	case study	students	VA, CO	large-scale	Positive	Low
Gorth, William Phillip, & Perkins, Marcy R.	1979	case study	students	IN	large-scale	Positive	Medium
Jones, Randall L.	1979	case study	students	UT	classroom	Positive	Low
Ogden, J.	1979	experiment or pre-post comparison	students	TX (Austin)	large-scale	Positive	High
Rentz, R.R.	1979	case study, survey, interview	college faculty	GA	large-scale	Positive inferred	High
Venesky, R.L., & Winfield, L.F.	1979	case study, interview, observation, records/ document review	schools	DE	classroom	Positive	High
Cypress, Edward J.	1980	case study	teachers, students	NY	classroom	Positive	Medium

Author	Year	Method	Participants	Location	Scale	Findings	Rigor
Fisher, Thomas H.	1980	case study	students	FL	large-scale	Positive	Medium
Ogle, Donna, & Fritts, James	1981	case study	teachers	IL (Skokie)	classroom	Positive	High
Popham, W. James & Rankin, Stuart C.	1981	case study	teachers	MI (Detroit)	large-scale	Positive inferred	High
Schlavin, Sheila A.	1981	experiment or pre-post comparison	New York schools	NY	large-scale	Positive	High
Brunton, M.L.	1982	experiment or pre-post comparison	students	OR	large-scale	Positive	High
Alexander, Cordelia R.	1983	case study	students	TX	large-scale	Positive	Medium
Gipps, Caroline, Steadman, Stephen, Blackston, Tessa, and Stierer, Barry	1983	case study	administrators, teachers	England	large-scale	Positive	High
Brooke, Nigel & Oxenham, John	1984	case study	students	Ghana, Mexico	large-scale	Positive	High
Natriello, Dornbusch	1984	observation	students in 38 classrooms		classroom	Positive	High
Stevens, Floraline I.	1984	experiment or pre-post comparison	students	CA (LA)	large-scale	Positive	High
Corcoran, Thomas B.	1985	research review	schools	USA	large-scale	Positive	Medium
McClain, C. J., & Krueger, D. W.	1985	case study	schools	MO	large-scale	Positive inferred	Medium
Resnick & Resnick	1985	observation, interview, research review	teachers	England & Wales	large-scale	Positive	High
Robb, Donald W.	1985	case study	staff	CA	classroom	Positive	Low
Robb, Donald W.	1985	case study	staff	OH	classroom	Positive	Low
Robb, Donald W.	1985	case study	staff	MO	classroom	Positive	Low
Smith, William J.	1985	case study	teachers	NY	classroom	Positive	Medium
Losak, J.	1986	experiment or pre-post comparison	schools	FL	large-scale	Positive	Medium
Koffler, Stephen L.	1987	not specified	schools	NJ	large-scale	Positive	Medium
Hughes, A.	1988	case study	students	Turkey	large-scale	Positive	High
Pennycuik, D.	1988	case study	students	UK	large-scale	Positive	Medium
Pennycuik, D., & Murphy, R.	1988	case study	students	England & Wales	large-scale	Positive	High
Somerset, Anthony	1988	case study	schools	Kenya	large-scale	Positive	Low
Perrin, Micheline	1989	survey	students	Switzerland	large-scale	Positive	Medium

Author	Year	Method	Participants	Location	Scale	Findings	Rigor
Warwick, Donald P., Reimers, Fernando, & McGinn, Noel	1989	interview, survey	teachers	Pakistan	classroom	Positive	Medium
Anderson, John O., et al.	1990	case study	students, teachers	British Columbia	large-scale	Positive	Medium
Heyneman, Stephen P., & Ranson, Angela	1990	case study	countries	various	large-scale	Positive	Medium
Johnstone, W.	1990	experiment or pre-post comparison	schools	TX	large-scale	Positive inferred	Medium
Lerner, B.	1990	case study	students	NJ	large-scale	Positive	Medium
Ligon, Glynn, et al.	1990	case study	schools	TX	large-scale	Positive	Low
Singh, Jasbir Sarjit, Marimuthu, T., & Mukherjee, Hena	1990	case study	students	Malaysia	large-scale	Positive	High
Ferrara, Steven, Willhoft, Joseph, Seburn, Carolyn, Slaughter, Frank, & Stevensen, Jose	1991	interview	teachers, administrators	MD	large-scale	Positive inferred	High
Grisay, A.	1991	case study	teachers	Belgium	large-scale	Positive	High
Moore, W.P.	1991	experiment or pre-post comparison	teachers	KS	large-scale	Positive	High
Willoughby, T. L., & Bixby, A. R.	1991	case study	schools	United States	large-scale	Positive inferred	Medium
Brown, D. F.	1992	interview	teachers, principals	TN, IL, NY	large-scale	Positive inferred	Medium
Plazak, Tomasz & Mazur, Zygmunt	1992	interview	teachers	Poland	large-scale	Positive	High
Whetton, Chris	1992	case study	teachers, students	England, Wales	large-scale	Positive inferred	High
Bullard, P., & Taylor, B. O.	1993	interview	teachers	NY	large-scale	Positive	High
Ekstein, Max A., & Noah, Harold J.	1993	case study	students	8 countries	large-scale	Positive	High
Shohamy, Elana	1993	interview	teachers, students	Israel	large-scale	No change	Medium
Shohamy, Elana	1993	interview	teachers, students	Israel	large-scale	Positive	Medium
United States General Accounting Office	1993	interview	administrators, teachers, employers	Canada	large-scale	Positive	High
Wall, Dianne & Alderson, J. Charles	1993	interview, observation	teachers	Sri Lanka	large-scale	No change	High
Bentz, Susan K.	1994	interview	teachers	IL	teacher	Positive inferred	Medium

Author	Year	Method	Participants	Location	Scale	Findings	Rigor
Bishop, John	1994	case study	countries	France, Holland, England, Scotland, US	large-scale	Positive	Medium
Matthews, Joan	1994	records/ document review	students	TX	large-scale	Positive	High
Bottoms, Gene, & Mikos, P.	1995	case study	students, teachers, administrators	SREB states	classroom	Positive	Medium
Prais, S.	1995	case study	countries	France, U.K., Germany	large-scale	Positive	Medium
Resnick, Nolan, & Resnick	1995	case study, records/ document review	curricula	France, Netherlands	large-scale	Positive	High
Waters, T., Burger, D., & Burger, S.	1995	case study	students	CO	classroom	Positive	Medium
Aguilera, Raymond V., & Hendricks, Joen M.	1996	case study	students	TX	large-scale	Positive	High
Anthony, Booker T.	1996	case study	teachers	NC	large-scale	Positive	Medium
Boylan, H, et al.	1996	survey, interview, records/document review	students	TX	large-scale	Positive	High
Khatti, Nidhi, Reeve, A.L., Kane, M.B., & Adamson, R.J.	1996	case study	teachers	various states	classroom	Positive	Medium
Poje, Daniel J.	1996	case study	schools	TN	large-scale	Positive	Medium
Robertson, S.N., & Simpson, C.A.	1996	case study	students	VA	large-scale	Positive	High
Shohamy, Elana; Donitsa-Schmidt, Smadar; & Ferman, Irit	1996	interview	students, teachers, inspectors	Israel	large-scale	Positive	High
Shohamy, Elana; Donitsa-Schmidt, Smadar; & Ferman, Irit	1996	interview, survey, records/document review	students, teachers, inspectors	Israel	large-scale	Positive	High
Van Stewart, Arthur	1996	case study	schools	KY	large-scale	Positive	Medium
Watanabe, Yoshinori	1996	interview, observation	yobiko teachers	Japan	classroom	Positive inferred	High
Andrews, S. & Fullilove, J.	1997	experiment or pre-post comparison	students	Hong Kong	large-scale	Positive	High
Beardon, D.	1997	case study	teachers	TX (Dallas)	classroom	Positive	Medium

Author	Year	Method	Participants	Location	Scale	Findings	Rigor
Cheng, Liying	1997	survey, observation, interview	teachers, students	Hong Kong	large-scale	Positive	High
Designs for Change	1997	interview, records/ document review	teachers, administrators, students	IL	large-scale	Positive	High
Florida Office of Program Policy Analysis	1997	survey, case study, records/ document review	principals	FL	classroom	Positive	High
Fox, J.	1997	interview	administrators	AL	large-scale	Positive	Low
Hurtgen, James R.	1997	case study	schools	NY	large-scale	Positive	Low
Manzo, K.	1997	interview	students	NC	large-scale	Positive	Medium
Miles, W. R., Bishop, Collins, Fink, Gardner, Grant, Hussain, et al.	1997	case study, interview	teachers	NY (Newport Junction)	large-scale	Positive	Medium
Nolet, McLaughlin	1997	experiment or pre-post comparison	students		large-scale	Positive	Low
Powell, Arthur G.	1997	research review	schools	US	large-scale	Positive	Medium
Southern Regional Education Board	1997	case study	one high school	NC	large-scale	Positive	Low
Stevenson, H. W., Lee, S., Carton, S., Evans, M., mezziane, S., Moriyoshi, N., & Schmidt, I.	1997	interview, records/ document reviews	parents, teachers, students	Japan	large-scale	Positive inferred	High
Stevenson, H. W., Lee, S., Carton, S., Evans, M., mezziane, S., Moriyoshi, N., & Schmidt, I.	1997	interview, records/ document reviews	parents, teachers, students	England	large-scale	Positive	High
Stevenson, H. W., Lee, S., Carton, S., Evans, M., mezziane, S., Moriyoshi, N., & Schmidt, I.	1997	interview, records/ document reviews	parents, teachers, students	France	large-scale	Positive	High
Williford, A. Michael	1997	case study	schools	OH	large-scale	Positive	Low
Argetsinger, Amy	1998	interview	teachers	MA	large-scale	Positive	Low
Chudowsky, Naomi, & Behuniak, Peter	1998	focus group	teachers	CT	large-scale	Positive	Medium
Grissmer, Flanagan	1998	records/ document review	students	TX, NC	large-scale	Positive	High
Johnson, Joseph F., Jr.	1998	case study	schools	TX	large-scale	Positive	High

Author	Year	Method	Participants	Location	Scale	Findings	Rigor
Johnson, Joseph F., Jr.	1998	records/ document review	students	TX	large-scale	Positive	Medium
Milwaukee Public Schools	1998	case study	administrators, teachers	WI	large-scale	Positive inferred	Medium
Trelfa, Douglas	1998	case study	teachers, students	Japan	large-scale	Positive	High
Berendt, Peter R. & Koski, Barry	1999	interview	principal, reading specialist	NY	large-scale	Positive	Medium
Clayton, Mark	1999	interview	teachers, principals, students	MA	large-scale	Positive	Low
Fuchs, Lynn; Fuchs, Douglas; Karns, Kathy; Hamlett, Carol L.; & Katzaroff, Michelle	1999	survey	teachers, students	Southeast	classroom	Positive	Low
Leithwood, K., Edge, Karen, & Jantzi, Doris	1999	case study	teachers, administrators	Scotland	large-scale	Positive	Low
Ragland, Mary A., Asera, Rose, Johnson, Joseph F., Jr.	1999	case study	schools	TX	large-scale	Positive	High
Schleisman, Jane	1999	interview	principals, counselors, teachers, district level employees	MN	large-scale	Positive inferred	High
Schmoker, M., & Marzano, R. J.	1999	research review	schools	range	large-scale	Positive	Low
Schmoker, Mike	1999	records/ document review	teachers, students	CO	large-scale	Positive	Medium
Steigemeier, Lois A.	1999	interview	teachers	WA	large-scale	Positive	Medium
Taylor, B., Pearson, P.D., Clark, K.F., & Walpole, S.	1999	experiment or pre-post comparison	teachers	UK	classroom	Positive	High
Zmuda, Allison & Tomaino, Mary	1999	interview	students, teachers	CT	classroom	Positive	Low
Benning, Victoria & Mathews, Jay	2000	interview	school administrators	VA	large-scale	Positive	Low
Blum, Robert E.	2000	interview, records/document review	administrators, teachers	OR	large-scale	Positive	Medium
Bradley, Ann	2000	interview	faculty	TX	large-scale	Positive	Low
Duggan, Terri, & Holmes, Madelyn	2000	case study	students	TX	large-scale	Positive	Low
Earl, Lorna, & Torrance, Nancy	2000	survey	schools	Canada	large-scale	Positive	Medium

Author	Year	Method	Participants	Location	Scale	Findings	Rigor
Fontana, J.	2000	records/ document review	schools	NY	large-scale	Positive	High
Gipps, Caroline	2000	observation, interview, survey	teachers	England	large-scale	No change	Medium
Grant, S. G.	2000	case study	teachers	NY	large-scale	Mixed	Medium
Hogan, K	2000	case study	teachers	TX	large-scale	Positive	Low
Hubler, Eric	2000	records/ document review	administrators	CO	large-scale	Positive	Low
Hurwitz, Nina & Hurwitz, Sol	2000	interview	administrators, teachers	TX	large-scale	Positive	Low
Hurwitz, Nina & Hurwitz, Sol	2000	interview	administrators, teachers	IL	large-scale	Positive	Low
Hurwitz, Nina & Hurwitz, Sol	2000	interview	administrators, teachers	NY	large-scale	Negative	Low
Janey, Clifford B.	2000	case study	schools	NY	large-scale	Positive	Low
Kelleher, J.	2000	case study	students	MA	large-scale	Positive	Medium
Mathews, Jay	2000	interview	schools	CT	large-scale	Positive	Low
Parker, E. T.	2000	interview	students	not specified	classroom	Positive	Medium
Reeves, Douglas B.	2000	case study	students	MO	large-scale	Positive	Medium
Skrla, Linda, Scheurich, James Joseph, & Johnson, Joseph F., Jr.	2000	case study	schools	TX	large-scale	Positive	Low
Skrla, Linda, Scheurich, James Joseph, & Johnson, Joseph F., Jr.	2000	case study	schools	TX	large-scale	Positive	Low
Strozeski, Michael W.	2000	records/ document review	students	TX	large-scale	Positive	Medium
van Dam, P. R. L.	2000	records/ document review	schools	Netherlands	large-scale	Positive	Low
Yussufu, Ahmed & Angaka, Johnstone A.	2000	case study	teachers	Kenya	large-scale	Positive	Medium
Anderson, Gerald E.	2001	case study	teachers, students	TX	large-scale	Positive	Medium
Carnoy, Martin; Loeb, Susanna; & Smith, Tiffany L.	2001	records/ document review	schools	TX	large-scale	Positive	High
Cawelti, Gordon & Protheroe, Nancy	2001	interview, records/ document review, observation	administrators, teachers	TX	large-scale	Positive	High

Author	Year	Method	Participants	Location	Scale	Findings	Rigor
Cawelti, Gordon & Protheroe, Nancy	2001	interview, records/ document review, observation	administrators, teachers	ID	large-scale	Positive	High
Cawelti, Gordon & Protheroe, Nancy	2001	interview, records/ document review, observation	administrators, teachers	TX	large-scale	Positive	High
Cawelti, Gordon & Protheroe, Nancy	2001	interview, records/ document review, observation	administrators, teachers	WV	large-scale	Positive	High
Cawelti, Gordon & Protheroe, Nancy	2001	interview, records/ document review, observation	administrators, teachers	TX	large-scale	Positive	High
Cawelti, Gordon & Protheroe, Nancy	2001	interview, records/ document review, observation	administrators, teachers	CA	large-scale	Positive	High
Clubine, Betsy, Knight, Dorothy L., Schneider, Cynthia L., & Smith, Pamela A.	2001	case study	administrators, teachers, counselors, students, parents	TX	large-scale	Positive	High
Garcia, Joseph & Rothman, Robert	2001	case study	schools	range	large-scale	Positive	Low
Hansen, Philip J.	2001	case study	students	IL	large-scale	Positive	High
Klinger, Don	2001	case study	schools	Canada	large-scale	Positive inferred	High
McGrath, J., Ashyby, N., Winters, K, Kickbush, P.	2001	case study	schools	VA	large-scale	Positive	Low
Milanowski, Anthony T. & Heneman, Herbert G., III	2001	interview, survey	teachers	Midwest	teacher	Positive inferred	Low
Monk, D. H., Sipple, J. W., & Killeen, K.	2001	interview	administrators, teachers	NY	large-scale	Mixed	High
Nelson, K.	2001	case study	teachers	MI	classroom	Positive	Medium
Phelps, Richard P.	2001	records/ document review	students		large-scale	Positive	High
Reid, K. S.	2001	interview	teachers, students	FL	large-scale	Positive	Low
Roderick, M., & Engel, M.	2001	interview	students	IL	large-scale	Positive	High
Bottoms, Gene	2002	experiment or pre- post comparison	students	SREB states		Positive	High

Author	Year	Method	Participants	Location	Scale	Findings	Rigor
Bradby, Denise, & Dykman, Ann	2002	experiment or pre-post comparison	students	SREB states	classroom	Mixed	High
Council of Chief State School Officers	2002	case study	teachers, principals	TX	large-scale	Positive	Medium
Council of Chief State School Officers	2002	case study	teachers, principals	TX	large-scale	Positive	Medium
Council of Chief State School Officers	2002	case study	teachers, principals	TX	large-scale	Positive	Medium
Council of Chief State School Officers	2002	case study	teachers, principals	TX	large-scale	Positive	Medium
Council of Chief State School Officers	2002	case study	teachers, principals	TX	large-scale	Positive	Medium
Fletcher, Michael A.	2002	case study	administrators	NY	large-scale	Positive	Low
Schafer, William D., Hultgren, Francine H., Hawley, Willis D., Abrams, Adnrew L., Seubert, Carole C., & Mazzoni, Susan	2002	case study	schools	MD	large-scale	Positive	High
Singh, Judy, & McMillan, James H.	2002	interview, focus group	teachers, principals	VA	large-scale	Positive	Medium
Stephens, Donnya	2002	interview	teachers	TX	large-scale	Positive inferred	Low
Wideman, R.	2002	case study	teachers	Ontario, Canada	large-scale	Positive	Medium
Wideman, Ron	2002	interview	teachers	Canada	large-scale	Positive inferred	Medium
Wright, Wayne E.	2002	interview	teachers	CA	large-scale	No change	Low
Brookhart, Susan M., & Bronowicz, Diane L.	2003	case study, interview	students		classroom	Positive	Medium
Brozo, W. G., & Hargis, C.	2003	case study	teachers	TN	classroom	Positive	Medium
Churchill, A.	2003	research review	schools	MA	large-scale	Positive	Medium
Flores, B. B. & Clark, E. R.	2003	journals	teachers	TX	large-scale	Positive	High
Stefanou, Candice, & Parkes, Jay	2003	experiment or pre-post comparison	science students	not specified	classroom	Positive	Low
Stone, Clement A., & Lane, Suzanne	2003	records/ document review	teachers, students	MD	large-scale	Positive	High

Author	Year	Method	Participants	Location	Scale	Findings	Rigor
Wang, Aubrey H., Coleman, Ashaki, B. Coley, Richard J., & Phelps, Richard P.	2003	records/ document review	unspecified	Singapore, Australia, England, Hong Kong, Japan, Korea, Holland	teacher	Positive	High
Burrows, C.	2004	interview	teachers	Australia	large-scale	Positive	Medium
Driscoll, D.	2004	experiment or pre-post comparison	schools	MA	large-scale	Positive	Medium
Driscoll, D.	2004	experiment or pre-post comparison	teachers	MA	teacher	Positive	Medium
Ferman, I.	2004	survey, interview, records/ document review	students	Israel	large-scale	Positive	High
Foster, David, & Noyce, Pendred	2004	case study	teachers	CA	large-scale	Positive	Medium
O'Day, Jennifer, Bitter, Catherine, Kirst, Mike, Carnoy, Martin, Woody, Elizabeth, Buttles, Melissa, Fuller, Bruce, & Ruenzel, David	2004	interview	teachers, principals, external evaluators, district staff	CA	large-scale	Positive	High
Qi, L.	2004	interview	test constructors, teachers, English inspectors	China	large-scale	No change	High
Snooks, Margaret K.	2004	observation	college students	TX	large-scale	Positive	Medium
University of Massachusetts, Donahue Institute	2004	interview	teachers, administrators, parents	MA	large-scale	Positive	High
Achievement Alliance	2005	case study	teachers, administrators	MN	large-scale	Positive	High
Achievement Alliance	2005	case study	teachers, administrators	NY	large-scale	Positive	High
Achievement Alliance	2005	case study	teachers, administrators	DE	large-scale	Positive	High
Achievement Alliance	2005	case study	teachers, administrators	WA	large-scale	Positive	High
Achievement Alliance	2005	case study	teachers	ID	large-scale	Positive	High
Achievement Alliance	2005	case study	teachers, administrators	AK	large-scale	Positive	High
Achievement Alliance	2005	case study	teachers, administrators	MA	large-scale	Positive	High
Achievement Alliance	2005	case study	teachers, principal	MA	large-scale	Positive	High

Author	Year	Method	Participants	Location	Scale	Findings	Rigor
Holland, D., Gross, B., & Anderson, J.	2005	interview	teachers	range	large-scale	Positive inferred	High
Rossi, Peter & McCulloch, Rob	2005	experiment or pre-post comparison	students	IL	large-scale	Positive	Low
Wise, Laress L., et al.	2005	interview	administrators, teachers	CA	large-scale	Positive	High
Achievement Alliance	2006	case study	teachers, administrators	GA	large-scale	Positive	High
Achievement Alliance	2006	case study	teachers, administrators	DE	large-scale	Positive	High
Achievement Alliance	2006	case study	teachers, administrators	PA	large-scale	Positive	High
Achievement Alliance	2006	case study	teachers, administrators	NY	large-scale	Positive	High
Achievement Alliance	2006	case study	teachers, administrators	AL	large-scale	Positive	High
Center for Public Education	2006	case study	administrators	IN	large-scale	Positive	Low
Center for the Future of Arizona, Mottison Institute for Public Policy	2006	case study	students	AZ	classroom	Positive	High
Center for the Future of Arizona, Mottison Institute for Public Policy	2006	case study	students	AZ	classroom	Positive	High
Center for the Future of Arizona, Mottison Institute for Public Policy	2006	case study	students	AZ	classroom	Positive	High
Center for the Future of Arizona, Mottison Institute for Public Policy	2006	case study	students	AZ	classroom	Positive	High
Faulkner, Shawn A., Cook, Christopher M.	2006	survey	middle school personnel	Kentucky	large-scale	Mixed	High
Morrison Institute for Public Policy, Arizona State University, Center for the Future of Arizona	2006	interview, survey	schools	AZ	large-scale	Positive	High
Wikstrom, Christina	2006	research review	schools	Sweden	large-scale	Negative	High
Yeh, Stuart S.	2006	interview	teachers and administrators	MN	large-scale	Positive	High

Author	Year	Method	Participants	Location	Scale	Findings	Rigor
Achievement Alliance	2007	case study	teachers, administrators	CA	large-scale	Positive	High
Achievement Alliance	2007	case study	teachers, administrators	NY	large-scale	Positive	High
Achievement Alliance	2007	case study	teachers, principal	KS	large-scale	Positive	High
Ganesh, Annapurna	2007	interview, observation	teachers	AZ	large-scale	Mixed	Low
Hayward, E. Louise	2007	case study	teachers	Scotland	large-scale	No change	High
Hayward, Geoff, & McNicholl, Jane	2007	case study	schools	England	large-scale	Positive	Medium
James, David, & Simmons, Jonathan	2007	case study	students, staff	England	large-scale	Positive	High
Le Floch, Kerstin Carlson, et al.	2007	records/ document review	schools	US	large-scale	Positive	High
Ryan, K.E., Ryan, A.M., Arbuthnot, K., & Samuels, M.	2007	interview	students	Midwest	large-scale	Positive	High
Wall, Dianne & Horak, Tania	2007	interview, observation	10 teachers, 21 students, 8 directors	Central and Eastern Europe	classroom	No change	High
Bisoux, Tricia	2008	interview	faculty	range	classroom	Positive inferred	Low
Lips, Dan, & Ladner, Matthew	2008	case study	schools	FL	large-scale	Positive	Low
Opfer, V. Darleen; Henry, Gary T.; & Mashvurn, Andrew J.	2008	survey	teachers	range	large-scale	Positive	High
Prapphal, K	2008	case study	schools	Thailand	large-scale	Positive inferred	Low
Sasaki, Miyuki	2008	research review	schools	Japan	large-scale	Negative	Medium
Steiny, Julia	2008	records/ document review	middle school math teachers	RI	classroom	Positive inferred	Low
Torres, Mario S., Zellner, Luana, Erlandson, David	2008	survey	principals	TX	large-scale	Positive	Medium
Willey-Rendon, Ruby	2008	case study	teachers	TX	large-scale	Positive	Low
Zimmerman, Barry J. & Dibenedetto, Maria K.	2008	interview	teachers, students	TN	large-scale	Positive	Medium
Heyneman, Stephen P.	1987, 1988	case study	countries	various	large-scale	Positive	Medium
Goldberg, Gail & Roswell, B.S.	1999-2000	case study, survey	teachers	MD	large-scale	Positive	High
Accountability in Action		case study	administrators, teachers		large-scale	Positive inferred	Medium

Author	Year	Method	Participants	Location	Scale	Findings	Rigor
Frederiksen, Norman		case study	students	MD	large-scale	Positive	Medium
WestEd		case study	schools	CA	large-scale	No change	Medium

References

- Achievement Alliance. (2005). It's being done: Dayton's Bluff, St. Paul, Minnesota. *The Alliance Alert*, 1(4).
- Achievement Alliance. (2005). It's being done: East Millsboro Elementary, Delaware. *The Alliance Alert*, 2(2).
- Achievement Alliance. (2005). It's being done: Elmont Memorial Junior-Senior High School, Nassau County, New York. *The Alliance Alert*, 1(10).
- Achievement Alliance. (2005). It's being done: Frankford Elementary, Delaware. *The Alliance Alert*, 1(10).
- Achievement Alliance. (2005). It's being done: Granger High School, Washington. *The Alliance Alert*, 1(10).
- Achievement Alliance. (2005). It's being done: Lapwai Elementary. *The Alliance Alert*, 1(6).
- Achievement Alliance. (2005). It's being done: Oakland Heights Elementary, Russellville, Arkansas. *The Alliance Alert*, 1(5).
- Achievement Alliance. (2005). It's being done: Port Chester Middle School, New York. *The Alliance Alert*, 2(5).
- Achievement Alliance. (2005). It's being done: Rock Hall Elementary, Maryland. *The Alliance Alert*, 1(1).
- Achievement Alliance. (2005). It's being done: University Park, Worcester. *The Alliance Alert*, 1(6).
- Achievement Alliance.(2006). It's being done: Capitol View Elementary, Atlanta, Georgia. *The Alliance Alert*, 2(6).
- Achievement Alliance. (2006). It's being done: M. Hall Stanton Elementary School, Philadelphia, Pennsylvania. *The Alliance Alert*, 2(1).
- Achievement Alliance. (2006). It's being done: West Jasper Elementary School, Alabama. *The Alliance Alert*, 2(3).
- Achievement Alliance. (2007). It's being done: Imperial High School. *The Alliance Alert*, 3(1).
- Achievement Alliance. (2007). It's being done: P.S./M.S. 124, Osmond A. Church School, New York. *The Alliance Alert*, 3(2).
- Achievement Alliance. (2007). It's being done: Ware Elementary School, Fort Riley, Junction City, Kansas. *The Alliance Alert*, 3(3).
- Aguilera, R.V., & Hendricks, J.M. (1996). Increasing standardized achievement scores in a high risk school district. *Curriculum Report*, 26(1). Reston, VA: National Association of Secondary School Principals.

- Alderson, J.C., & Hamp-Lyons, L. (1996). TOEFL Preparation Courses: A Study of Washback. *Language Testing, 13*(3), 280-297.
- Alexander, C.R. (1983). A case study: Testing in the Dallas Independent School District. In W.E. Hathaway (Ed.), *Testing in the schools*. San Francisco, CA: Jossey-Bass.
- Anderson, G.E. (2001). Brazosport Independent School District: Implementation of the Quality agenda to ensure excellence and equity for all students. *Education Reform Success Stories*. Amherst, MA: National Evaluation Systems.
- Anderson, J.O., & et al. (1990). *The Impact of Provincial Examinations on Education in British Columbia: General Report*.
- Andrews, S., & Fullilove, J. (1997). The elusiveness of washback: Investigating the impact of a new oral exam on students' spoken language performance. Paper presented at the International Language in Education Conference, University of Hong Kong.
- Anthony, B.T. (1996). Assessing writing through common examinations and student portfolios. In T.W. Banta, J.P. Lund, K.E. Black & F.W. Oblander (Eds.), *Assessment in practice: Putting principles to work on college campuses*. San Francisco, CA: Jossey-Bass.
- Argetsinger, A. (1998, December 9). Maryland students boost test scores. *Washington Post*, p. B1,
- Barksdale-Ladd, M.A., & Thomas, K.F. (2000). What's at state in high-stakes testing: Teachers and parents speak out. *Journal of Teacher Education, 51*(5), 384-397.
- Beardon, D. (1997). An overview of the elementary mathematics program 1996-97. Dallas, TX: Dallas Public Schools.
- Benning, V., & Mathews, J. (2000). Statewide scores up on most VA tests. *Washington Post*.
- Bentz, S.K. (1994). The impact of certification testing on teacher education. *Continuing discussions in teacher certification testing*. Amherst, MA: National Evaluation Systems.
- Berendt, P.R., & Koski, B. (1999). No Shortcuts to Success. *Educational Leadership, 56*(6), 45-47.
- Bishop, J.H. (1994). *Impacts of school organization and signaling on incentives to learn in France, The Netherlands, England, Scotland, and the United States*. Ithaca, NY: Cornell University, New York State School of Industrial and Labor Relations, Center for Advanced Human Resource Studies.
- Bisoux, T. (2008). Measures of success. *BizEd*.
- Blum, R.E. (2000). Standards-based reform: Can it make a difference for students? *Peabody Journal of Education, 75*(4), 90-113.
- Bottoms, G. (2002). *Raising the Achievement of Low-Performing Students: What High Schools Can Do*. Atlanta, GA: Southern Regional Education Board.
- Bottoms, G., & Mikos, P. (1995). *Seven most-improved "High Schools that Work" sites raise achievement in reading, mathematics, and science: A report on improving student learning*.

Atlanta, GA: Southern Regional Education Board.

Boucher, C.S. (1935). *The Chicago college plan*. Chicago, IL: University of Chicago.

Boylan, H., Bonham, B., Abraham, A., Anderson, J., Morante, E., Ramirez, G., et al. (1996). *An evaluation of the Texas Academic Skills Program*. Austin, TX: Texas Higher Education Coordinating Board.

Bradby, D., & Dykman, A. (2003). *Effects of "High Schools that Work" practices on Student Achievement (Research Brief)*. Atlanta, GA: Southern Regional Education Board.

Bradley, A. (2000, October 4). Put to the test. *Education Week*.

Brereton, J.L. (1944). *The case for examinations: An account of their paces in education with some proposals for their reform*. London: Cambridge University Press.

Brooke, N., & Oxenham, J. (1984). *The influence of certification and selection on teaching and learning*. In J. Oxenham (Ed.), *Education versus qualifications? A study of relationships between education, selection for employment and the productivity of labor*. London, UK: George Allen & Unwin.

Brookhart, S.M., & Bronowicz, D.L. (2003). "I Don't Like Writing. It Makes My Fingers Hurt": students talk about their classroom assessments. *Assessment in Education: Principles, Policy & Practice*, 10(2), 221.

Brookover, W.B., & Lezotte, L.W. (1979). *Changes in school characteristics coincident with changes in student achievement*. East Lansing, MI: Michigan State University, Institute for Research on Teaching.

Brooks, S.S. (1922). *Reactions of teachers and pupils to standardized tests*. East Swanzey, NH: Winchester School District.

Brown, D.F. (1992, April). Altering curricula through state testing: Perceptions of teachers and principals. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Brozo, W.G., & Hargis, C. (2003). Using low-stakes reading assessment. *Educational Leadership*, 61(3), 60-64.

Brunton, M.L. (1982, March). Is competency testing accomplishing any breakthrough in achievement? Paper presented at the annual meeting of the Association for Supervision and Curriculum Development, Anaheim, CA.

Bullard, P., & Taylor, B.O. (1993). *Making School Reform Happen*. Needham Heights, MA: Allyn & Bacon.

Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian Adult Migrant English Program. In L. Cheng & Y. Watanabe (Eds.), *Washback in language testing: Research contexts and methods* (pp. 113-128). Mahwah, NJ: Lawrence Erlbaum Associates.

- Carnoy, M., Loeb, S., & Smith, T.L. (2001). *Do higher state test scores in Texas make for better high school outcomes?* Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.
- Cawelti, G., & Protheroe, N. (2001). *High school achievement: How six school districts changed into high-performance systems.* Arlington, VA: Educational Research Service.
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education, 11*(1).
- Chudowsky, N., & Behuniak, P. (1998). Using Focus Groups to Examine the Consequential Aspect of validity. *Educational Measurement: Issues and Practice, 17*(4), 28-38.
- Churchill, A. (2003). Conclusions: The impact of education reform after ten years. *Education reform: Ten years after the Massachusetts Education Reform Act of 1993* (pp. 34-36). Washington, DC: Center for Educational Policy.
- Clayton, M. (1999, April 6). *Do high-stakes tests change a school? Yes.* Retrieved February 13, 2001, from <http://www.csmonitor.com>
- Clubine, B., Knight, D.L., Schneider, C.L., & Smith, P.A. (2001). *Opening Doors: Promising Lessons from Five Texas High Schools:* For full text: <http://www.utdanacenter.org>.
- Consultative Committee on Examinations* (1910). Report. London, UK: Author.
- Corcoran, T.B. (1985). *Competency testing and at-risk youth.* Philadelphia, PA: Research for Better Schools.
- Council of Chief State School Officers (2002). *Expecting Success: A study of five high performing, high poverty schools.* Washington, DC: Author.
- Cypress, E.J. (1980). Making reading achievement tests work for the inner-city student. In C.B. Stalford (Ed.), *Testing and evaluation in schools: Practitioners' views* (pp. 27-32). Washington, DC: U.S. Department of Education.
- Dawson, K.S., & Dawson, R.E. (1985). *Minimum competency testing and local schools.* Unpublished manuscript.
- Designs for Change (1997). *Chicago elementary schools with a seven-year trend of improved reading achievement: What makes these schools stand out?* Chicago, IL: Author.
- Down, A.G. (1979, April). Implications of minimum-competency testing for minority students. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.
- Duggan, T.E., & Holmes, M.E. (2000). *Closing the Gap: A Report on the Wingspread Conference "Beyond the Standards Horse Race: Implementation, Assessment, and Accountability-The Keys to Improving Student Achievement"* (Racine, Wisconsin, November 2-4, 1999). Special Report.
- Earl, L., & Torrance, N. (2000). Embedding accountability and improvement into large-scale

- assessment: What difference does it make? *Peabody Journal of Education*, 75(4), 114-141.
- Eckstein, M.A., & Noah, H.J. (1993). *Secondary school examinations: International perspectives on policies and practice*. New Haven, CT: Yale University Press.
- Enochs, J.C. (1978). Modesto, California: A return to the four Rs. *Phi Delta Kappan*, 609-610.
- Estes, G.D., Colvin, L.W., & Goodwin, C. (1976, April). A Criterion-Referenced Basic Skills Assessment Program in a Large City School System. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Faulkner, S.A., & Cook, C.M. (2006). Testing vs. Teaching: The Perceived Impact of Assessment Demands on Middle Grades Instructional Practices. *Research in Middle Level Education Online*, 29(7), 1-13.
- Feldhusen, J.F. (1964, February). Student perceptions of frequent quizzes and post-mortem discussion of tests. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. In L. Cheng & Y. Watanabe (Eds.), *Washback in language testing: Research contexts and methods* (pp. 199-210). Mahwah, NJ: Lawrence Erlbaum Associates.
- Ferrara, S., Willhoft, J., Seburn, C., Slaughter, F., & Stevenson, J. (1991). Local assessments designed to parallel statewide minimum competency tests: Benefits and drawbacks. In R.G. O'Sullivan & R.E. Stake (Eds.), *Advances in program evaluation: Effects of mandated assessment on teaching* (pp. 41-74). Greenwich, CT: JAI Press.
- Findley, J. (1978). Westside's minimum competency graduation requirements: A program that works. *Phi Delta Kappan*, 614-618.
- Firestone, W.A., & Mayrowetz, D. (2000). Rethinking 'high stakes': lessons from the United States and England and Wales. *Teachers College Record*, 102(4), 724-749.
- Firestone, W.A., Monfils, L., & Schorr, R.Y. (2004). Test preparation in New Jersey: inquiry-oriented and didactic responses. *Assessment in Education: Principles, Policy, & Practice*, 11(1), 67-88.
- Fisher, T.H. (1978). Florida's approach to competency testing. *Phi Delta Kappan*, 59, 599-602.
- Fisher, T.H. (1980). Florida competency testing program. In R.M. Jaeger & C.K. Tittle (Eds.), *Minimum competency achievement testing: Motives, models, measures, and consequences*. Berkeley, CA: McCutchan Publishing Corporation.
- Fletcher, M.A. (2002, January 2). After school's test success comes worry. *Washington Post*.
- Flores, B.B., & Clark, E.R. (2003). Texas voices speak out about high-stakes testing: Preservice teachers, teachers, and students. *Current Issues in Education*, 6(3).
- Fontana, J. (2000). New York's test-driven standards. In A.A. Glatthorn & J. Fontana (Eds.), *Coping with standards, tests, and accountability: Voices from the classroom*. Washington,

DC: NEA Teaching and Learning Division.

- Foss, O. (1977). A new approach: Vocational foundation courses and examinations Criteria for awarding school leaving certificates: An international discussion (pp. 191-209). Based on the Proceedings of the 1977 Conference the International Association for Educational Assessment held at the Kenyatta Conference Center, Nairobi.
- Foster, D., & Noyce, P. (2004). The Mathematics Assessment Collaborative: Performance Testing to Improve Instruction. *Phi Delta Kappan*, 85(5), 367-374.
- Fox, J. (1997). Alabama ranking system lifts test scores, spirits. *Education Daily*, 30, 2-3.
- Fox, J., & Cheng, L. (2007). Did we take the same test? Differing accounts of the Ontario Secondary School Literacy Test by first and second language test-takers. *Assessment in Education*, 14(1), 9-26.
- Fredericksen, N. (n.d.). *Information for use in school accountability*.
- Fuchs, L., Fuchs, D., Karns, K., Hamlett, C.L., & Katzaroff, M. (1999). Mathematics performance assessment in the classroom: Effects on teacher planning and student problem solving. *American Educational Research Journal*, 35(3), 609-645.
- Garcia, J., & Rothman, R. (2001). *Three paths, one destination: Standards-based reform in Maryland, Massachusetts, and Texas*. Washington, DC: Achieve.
- Gilmore, A. (2005). *The impact of PIRLS (2001) and TIMSS (2003) in low- and middle-income countries*.
- Gipps, C. (2000). Findings from large scale assessment in England, in session. *Foreign Language Teaching and Research*, Vienna, Austria: IAEA.
- Gipps, C., Steadman, S., Blackstone, T., & Stierer, B. (1983). *Testing children: Standardised testing and local education authorities and schools*. London: Heinemann Educational Books, Ltd.
- Goldberg, G., & Roswell, B.S. (2000). From perception to practice: The impact of teachers' scoring experience on performance-based instruction and classroom assessment. *Educational Assessment*, 6(4), 257-290.
- Gorth, W.P., & Perkins, M.R. (1979). *Final comprehensive report*. Amherst, MA: National Evaluation Systems.
- Grant, S.G. (2000). Teachers and tests: Exploring teachers' perceptions of changes in the New York state testing program. *Education Policy Analysis Archives*, 8(14).
- Grisay, A. (1991, September 12-15). Improving assessment in primary schools: "APER" research reduces educational failure rates. Paper presented at the Assessment of pupil achievement: Motivation and school success, Liege, Belgium.
- Grissmer, D., & Flanagan, A. (1998). *Exploring rapid score gains in Texas and North Carolina*. Santa Monica, CA: RAND Corporation.

- Hansen, P. (2001). Chicago public schools: Improvement through accountability. *Education Reform Success Stories*. Northampton, MA: National Evaluation Systems.
- Hayward, E.L. (2007). Curriculum, Pedagogies and Assessment in Scotland: The Quest for Social Justice. "Ah Kent Yir Faither". *Assessment in Education: Principles, Policy & Practice*, 14(2), 251-268.
- Hayward, G., & McNicholl, J. (2007). Modular mayhem? A case study of the development of the A-level science curriculum in England. *Assessment in Education: Principles, Policy & Practice*, 14(3), 335-351.
- Heyneman, S. (1987). *Uses of examinations in developing countries: Selection, research, and education sector management*. Washington, DC: Seminar Paper No. 36, Economic Development Institute, The World Bank.
- Heyneman, S.P., & Ransom, A.W. (1990). Using examinations and testing to improve educational quality. *Educational Policy*, 4(3), 177-192.
- Hogan, K. (2000). Educational reform in Texas. In A. Glatthorn & J. Fontana (Eds.), *Coping with standards, tests, and accountability: Voices from the classroom*. Washington, DC: NEA Teaching and Learning Division.
- Holland, D., Gross, B., & Anderson, J. (2005, April). Subject matters: How accountability impact high school math and English departments. Paper presented at the annual conference of the American Association of Educational Research, Montreal, Canada.
- House, E., Rivers, W., & Stufflebeam, D. (1974). *An assessment of the Michigan Accountability System*. Lansing, MI: Michigan Department of Education.
- Hubler, E. (2000). How schools are preparing for CSAP. *Denver Post*.
- Hughes, A. (1988). Introducing a needs-based test of English language proficiency into an English-medium university in Turkey. In A. Hughes (Ed.), *Testing English for university study* (pp. 134-153). London: Modern English Publications.
- Hurtgen, J.R. (1997). Assessment of General Learning: State University of New York College at Fredonia. *New Directions for Higher Education*, 25(4), 59-69.
- James, D., & Simmons, J. (2007). Alternative Assessment for Learner Engagement in a Climate of Performativity: Lessons from an English Case Study. *Assessment in Education: Principles, Policy & Practice*, 14(3), 353-371.
- Janey, C.B. (2000, August 2). Pathways to high school success. *Education Week*.
- Johnson, J.F., Jr. (1998). Improving public schools in Texas. *Basic Education*, 43(2), 2-5.
- Johnson, J.F., Jr. (1998). The influence of a state accountability system on student achievement in Texas. *Virginia Journal of Social Policy & the Law*, 6(1).
- Johnstone, W. (1990, January 25-27). Local school district perspectives. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, Texas.

- Jones, R.L. (1979). Performance testing of second language proficiency. In E.J. Briere & F.B. Hinofotis (Eds.), *Concepts in language testing: Voices from the classroom*. Washington, DC: Teachers of English to Speakers of Other Languages.
- Kelleher, J. (2000). Developing rigorous standards in Massachusetts. In A. Glatthorn & J. Fontana (Eds.), *Coping with standards, tests, and accountability: Voices from the classroom*. Washington, DC: NEA Teaching and Learning Division.
- Khattari, N., et al. (1995). *Assessment of Student Performance. Volume I: Findings and Conclusions. Studies of Education Reform*. Washington, DC: U.S. Department of Education.
- Kiser, S.M. (2007). An evolving change in public schools: An assessment of teachers' and administrators' perceptions and classroom changes concerning high-stakes testing: PhD dissertation, East Tennessee State University.
- Klinger, D. (2001). Oops, that was a mistake: Examining the effects and implications of changing assessment policies. In P. de Broucker, & Sweetman, Arthur (Ed.), *Towards evidence-based policy for Canadian education* (pp. 333-346). Montreal, Canada: John Deutsch Institute for the Study of Economic Policy.
- Koffler, S.L. (1987). Assessing the impact of a state's decision to move from minimum competency testing toward higher level testing for graduation. *Educational Evaluation and Policy Analysis*, 9(4), 325-336.
- Kulp, D.H., II (1934). Weekly tests for graduate students? In C.C. Ross (Ed.), *Measurement in today's schools*. New York, NY: Prentice-Hall.
- Le Floch, K.C., Martinez, F., O'Day, J., Stecher, B., Taylor, J., & Cook, A. (2007). *State and Local Implementation of the "No Child Left Behind Act." Volume III-Accountability under "NCLB" Interim Report*. Washington, DC: US Department of Education.
- Leithwood, K., Edge, K., & Jantzi, D. (1999). *Educational accountability: The state of the art*. Gütersloh, Germany: Bertelsman Foundation Publishers.
- Lerner, B. (1990). Good news about American education. *Commentary*, 91(3).
- Ligon, G., et al. (1990, January 25-27). Statewide testing in Texas. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX.
- Lips, D., & Ladner, M. (2008). Demography defeated: Florida's K-12 reforms and their lesson for the nation. *Goldwater Institute Policy Report* (227).
- Losak, J. (1986, October 15-17). Mandated entry- and exit-level testing in the state of Florida: A brief history. Paper presented at the California State University of Conference on Student Outcomes Assessment, Pomona, CA.
- Luna, C., & Turner, C.L. (2001). The impact of the MCAS: Teachers talk about high stakes testing. *English Journal*, 91(1), 79-87.
- Madaus, G.F. (1981). Reactions to the 'Pittsburgh Papers'. *Phi Delta Kappan*, 62(9), 634-636.

- Magruder, J., McManis, M., & Young, C. (1997). The right idea at the right time: Development of a transformational assessment culture. In P. Gray & T.W. Banta (Eds.), *The campus-level impact of assessment: Progress, problems, and possibilities. New directions for higher education; No. 100*. San Francisco: Jossey-Bass.
- Manzo, K.K. (1997). High stakes: Test truths or consequences. *Education Week on the Web*, 1-2.
- Mathews, J. (2000, July 18). Connecticut's education success story: State getting results with tough standards and high salaries for teachers, rigorous annual tests for students. *Washington Post*, p. A11,
- Matthews, J. (1994). The effectiveness of TASP-induced remediation among Texas's tri-ethnic population. *Continuing discussions in teacher certification testing*. Amherst, MA: National Evaluation Systems.
- McClain, C.J., & Krueger, D.W. (1985). Using outcomes assessment: A case study in institutional change. *New Directions for Institutional Research (47)*, 33-46.
- McClellan, M.C. (1988). Testing and reform. *Practical Applications of Research*, 769-771.
- McDermott, K.A. (2003). Capacity to implement education reform. *Education reform: Ten years after the Massachusetts Education Reform Act of 1993* (pp. 31-33). Washington, DC: Center for Education Policy.
- McGrath, J., Ashby, N., Winters, K., & Kickbush, P. (2001). *Achieving high: A Virginia school raises expectations and proves every child can succeed*. Washington, DC: US Department of Education, Community Update and the Satellite Town Meeting.
- Messenger. (1934). Unpublished Dissertation, University of Iowa, Iowa City, IA.
- Milanowski, A., & Heneman, H.G., III (2001). Assessment of teacher reactions to a standards-based teacher evaluation system: A pilot study. *Journal of Personnel Evaluation in Education*, 15(3), 193-212.
- Miles, W.R., Bishop, Collins, Fink, Gardner, Grant, et al. (1997). *High standards for all in New York state results of ten case studies*. New York, NY: Boards of Cooperative Educational Services.
- Milwaukee Public Schools (1998). *Characteristics of effective schools*. Milwaukee, WI: Authors.
- Monk, D.H., Sipple, J.W., & Killeen, K. (2001). *Adoption and adaptation: New York state school district responses to state imposed learning and graduation requirements: An eight-year retrospective*. State College, PA: Penn State University.
- Moore, W.P. (1991). Relationships among teacher test performance pressures, perceived testing benefits, test preparation strategies, and student test performance. PhD dissertation, University of Kansas, Lawrence.
- Murnane, R.J., & Levy, F. (1998). Standards, information, and the demand for student achievement. *Economic Policy Review - Federal Reserve Bank of New York*, 117-124.

- Nassif, P. (1992). Aligning assessment and instruction: Can teacher testing result in better teaching? *Current topics: Teacher certification testing*. Amherst, MA: National Evaluation Systems.
- Natriello, G., & Dornbusch, S.M. (1984). *Teacher evaluative standards and student effort*. New York, NY: Longman.
- Neill, S.B. (1978). *The competency movement: Problems and solutions*. Sacramento, CA: Education News Service.
- Nelson, K. (2001). Assessing student competence in the visual arts. In C.A. Palomba & T.W. Banta (Eds.), *Assessing student competence in accredited disciplines* (pp. 177-216). Sterling, VA: Stylus.
- Nolet, V., & McLaughlin, M. (1997). Using CBM to explore a consequential basis for the validity of a statewide performance assessment. *Diagnostique*, 22(3), 147-163.
- Nuttall, D.L., & Stobart, G. (1994). National curriculum assessment in the UK. *Educational Measurement: Issues and Practice*, 13(2), 24-27.
- O'Day, J., Bitter, C., Kirst, M., Carnoy, M., Woody, E., Buttles, M., et al. (2004). *Assessing California's Accountability System: Successes, Challenges, and Opportunities for Improvement. Policy Brief 04-2*. Berkeley, CA: Policy Analysis for California Education.
- Ogden, J. (1979, April). High school competency graduation requirements: Do they result in better graduates? Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Ogle, D., & Fritts, J. (1981). Criterion-referenced reading assessment valuable for process as well as for data. *Phi Delta Kappan*, 62(9), 640-641.
- Palomba, C.A. (1997). Assessment at Ball State University. *New Directions for Higher Education*, 25(4), 31-45.
- Parker, E.T. (2000). Unexpected Benefits of Testing. *Performance Improvement*, 39(9), 40-44.
- Passman, R. (2001). Experiences with Student-Centered Teaching and Learning in High-Stakes Assessment Environments. *Education*, 122(1).
- Pennycook, D. (1988). The development, use and impact of graded tests. In R. Murphy & H. Torrance (Eds.), *The changing face of educational assessment*. Milton Keynes, UK: Open University Press.
- Pennycook, D., & Murphy, R. (1988). *The impact of graded tests*. London: Falmer Press.
- Perrin, M. (1989). Summative evaluation and pupil motivation. In P. Wilson (Ed.), *Assessment of pupil achievement: Motivation and school success*. Genève, Switzerland: University of Geneva.
- Phelps, R.P. (2001). Benchmarking to the world's best in mathematics: Quality control in curriculum and instruction among the top performers in the TIMSS. *Evaluation Review*,

25(4), 391-439.

- Plazak, T., & Mazur, Z. (1992). University entrance in Poland. In P. Black (Ed.), *Physics examinations for university entrance: An international study. Science and technology education. Document series, No. 45*. Paris: UNESCO.
- Poje, D.J. (1996). Student Motivation and Standardized Testing for Institutional Assessment. In T.W. Banta, J.P. Lund, K.E. Black & F.W. Oberlander (Eds.), *Assessment in practice: Putting principles to work on college campuses* (pp. 179-182). San Francisco, CA: Jossey-Bass.
- Popham, W.J., & Rankin, S.C. (1981). Minimum competency tests spur instructional improvement. *Phi Delta Kappan*, 62(9), 637-639.
- Powell, A.G. (1997). Student incentives and the College Board system. *American Educator*, 21(3), 11-17.
- Prais, S. (1995). *Productivity, education and training. Vol. II*. London: National Institute for Economic and Social Research.
- Prapphal, K. (2008). Issues and trends in language testing and assessment in Thailand. *Language Testing*, 25(1), 127-143.
- Pronaratna, B. (1976). *Examination reforms in Sri Lanka. Experiments and innovations in education. No. 24. International Bureau of Migration Series*. Asian Centre of Educational Innovation for Development (Bangkok), Paris: UNESCO.
- Qi, L. (2004). Has a high-stakes test produced the intended changes? In L. Cheng & Y. Watanabe (Eds.), *Washback in language testing: Research contexts and methods* (pp. 171-190). Mahwah, NJ: Lawrence Erlbaum Associates.
- Ragland, M.A., Asera, R., & Johnson, J.F., Jr. (1999). Urgency, Responsibility, Efficacy: Preliminary Findings of a Study of High-Performing Texas School Districts: Web site: <http://www.starcenter.org/services/main.htm#product>.
- Ramanathan, H. (2008). Testing of English in India: A developing concept. *Language Testing*, 25(1), 111-126.
- Reeves, D.B. (2000). Standards Are Not Enough: Essential Transformations for School Success. *NASSP Bulletin*, 84(620), 5-19.
- Reeves, D.B. (2004). The 90/90/90 schools: A case study. *Accountability in action: A blueprint for learning organizations*. Denver, CO: Advanced Learning Press.
- Reid, K.S. (2001, April 11). From worst to first. Education Week.
- Rentz, R.R. (1979). Testing and the college degree. In W.B. Schrader (Ed.), *Measurement and educational policy: New directions for testing and measurement*. San Francisco: Jossey-Bass.
- Resnick, D.P., & Resnick, L.B. (1985). Standards, curriculum, and performance: A historical and comparative perspective. *Educational Researcher*, 14(4), 5-20.

- Resnick, L.B., Nolan, K.J., & Resnick, D.P. (1995). Benchmarking Education Standards. *Educational Evaluation and Policy Analysis*, 17(4), 438-461.
- Robb, D.W. (1985). Strategies for implementing successful mastery learning programs: Case studies. In J. Hsia (Ed.), *Improving Student Achievement Through Mastery Learning Programs*. San Francisco, CA: Jossey-Bass Publishers.
- Robertson, S.N., & Simpson, C.A. (1996). In T.W. Banta, J.P. Lund, K.E. Black & F.W. Oberlander (Eds.), *General education discipline evaluation process for the community college. Assessment in practice: Putting principles to work on college campuses* (pp. 190-194). San Francisco: Jossey-Bass.
- Roderick, M., & Engel, M. (2001). The Grasshopper and the Ant: Motivational Responses of Low-Achieving Students to High-Stakes Testing. *Educational Evaluation and Policy Analysis*, 23(3), 197-227.
- Rossi, P., & McCulloch, R. (2005, May 27). Preliminary analyses of effects of non-disclosure. *Chicago Business Online*.
- Ryan, K.E., Ryan, A.M., Arbuthnot, K., & Samuels, M. (2007). Students' motivation for standardized math exams. *Educational Researcher*, 36(1), 5-13.
- Sasaki, M. (2008). The 150-year history of English language assessment in Japanese education. *Language Testing*, 25(1), 63-83.
- Schafer, W.D., Hultgren, F.H., Hawley, W.D., Abrams, A.L., Seubert, C.C., & Mazzoni, S. (2002). *Study of Higher-Success and Lower-Success Elementary Schools*. College Park, MD: School Improvement Program, University of Maryland.
- Schlawn, S.A. (1981, December). The New York State testing program in writing: Its influence on instruction. Paper presented at the International Conference on Language Problems and Public Policy, Cancun, Mexico.
- Schleisman, J. (1999, October). An in-depth investigation of one school district's responses to an externally-mandated, high-stakes testing program in Minnesota. Paper presented at the annual meeting of the University Council for Educational Administration, Minneapolis, MN.
- Schmoker, M. (1996). *Results: The key to continuous school improvement*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).
- Schmoker, M., & Marzano, R.J. (1999). Realizing the promise of standards-based education. *Educational Leadership*, 56(6).
- Scott, C. (2007). Stakeholder perceptions of test impact. *Assessment & Evaluation in Higher Education*, 14(1), 27-49.
- Scott, I.O. (1934). Unpublished Dissertation, University of Iowa, Iowa City, IA.
- Shohamy, E. (1993). *The power of tests: The impact of language tests on teaching and learning*. Longman: Harlow, England.

- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298-317.
- Singh, J., & McMillan, J.H. (2002, April). Staff development practices in schools demonstrating significant improvement on high-stakes tests. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Singh, J.S., Marimutha, T., & Mukjerjee, H. (1990). Learning motivation and work: A Malaysian perspective. In P. Broadfoot, R. Murphy & H. Torrance (Eds.), *Changing educational assessment: International perspectives and trends* (pp. 177-198). London: Routledge.
- Skrla, L., Scheurich, J.J., & Johnson, J.F., Jr. (2000). *Equity-driven achievement-focused school districts*. Austin, TX: The Charles A. Dana Center.
- Smith, W.J. (1985). Incorporating testing and retesting into the teaching plan. In J. Hsia (Ed.), *Improving Student Achievement Through Mastery Learning Programs*. San Francisco, CA: Jossey-Bass Publishers.
- Snooks, M.K. (2004). Using Practice Tests on a Regular Basis to Improve Student Learning. *New Directions for Teaching and Learning*, 2004(100), 109-113.
- Solberg, W. (1977). School leaving examinations: Why or why not?: The case for school leaving examinations: The Netherlands. In F.M. Ottobre (Ed.), *Criteria for awarding school leaving certificates: An international discussion* (pp. 37-46). Nairobi.
- Somerset, A. (1988). Examinations as an instrument to improve pedagogy. In S.P. Heyneman & I. Fagerlind (Eds.), *University examinations and standardized testing*. Washington, DC: The World Bank (Technical Paper, 78).
- Southern Regional Education Board (1997). Case Study: Hoke County High School, Raeford, North Carolina. Atlanta, GA: Author.
- Stancavage, F.B., Roeber, E.D., & Bohrnstedt, G.W. (1993). *Impact of the 1992 Trial State Assessment program: A followup study*. Washington, DC: The National Academy of Education.
- Stefanou, C., & Parkes, J. (2003). Effects of Classroom Assessment on Student Motivation in Fifth-Grade Science. *Journal of Educational Research*, 96(3), 152-162.
- Steiny, J. (2008, October 19, 2008). Self-evaluation helps Barrington teachers succeed. Retrieved October 20, 2008, from http://www.projo.com/education/juliasteiny/content/se_educationwatch19_10-19-08_QHBUANJ_v6.22b1a0d.html
- Stephens, D. (2002). *Impact of standards of African Americans in Texas: Practitioners' critical*. San Francisco, CA: Caddo Gap Press.
- Stevens, F.I. (1984). *The effects of testing on teaching and curriculum in a large urban school district: ERIC/TM Report 86*, ERIC Clearinghouse on Tests, Measurement, and Evaluation.
- Stevenson, H.W., & Lee, S., et al. (1997). *International comparisons of entrance and exit*

examinations: Japan, United Kingdom, France, and Germany. U.S. Department of Education, Office of Educational Research and Improvement.

- Stiegemeier, L.A. (1999). *Organizing for Success: A Study about Mathematics Assessment Results in Washington State*. Washington, DC: Eisenhower Professional Development Program.
- Stone, C.A., & Lane, S. (2003). Consequences of a State Accountability Program: Examining Relationships between School Performance Gains and Teacher, Student, and School Variables. *Applied Measurement in Education*, 16(1), 1-26.
- Strozeski, M.W. (2000, April 25-27). Alignment of curriculum and instruction to state standards and assessments: A visit to the real world. Paper presented at the National Council on Measurement in Education, New Orleans, LA.
- Stuit, D.B. (1947). *Personnel research and test development in the Bureau of Naval Personnel*. Princeton, NJ: Princeton University Press.
- Taylor, B., Pearson, P.D., Clark, K.F., & Walpole, S. (1999). *Beating the odds in teaching all children to read. Report No. 2-006*. Ann Arbor, MI: Center for the Improvement of Early Reading.
- The Center for Public Education. (2006, August 27). Accountability plan spurs achievement gains for Indiana district. Retrieved September 28, 2008, from http://www.centerforpubliceducation.org/site/c.kjXJ5MPlwE/b.1504677/k.F19F/Accountability_plan_spurs_achievement_gains_for_Indiana_district.htm
- Torrance, H. (2007). Assessment as Learning? How the Use of Explicit Learning Objectives, Assessment Criteria and Feedback in Post-Secondary Education and Training Can Come to Dominate Learning. *Assessment & Evaluation in Higher Education*, 14(3), 281-294.
- Trelfa, D. (1998). The development and implementation of education standards in Japan, Chapter 2, *The Educational System in Japan: Case Study Findings*. U.S. Department of Education, Office of Educational Research and Improvement, National Institute on Student Achievement, Curriculum, and Assessment.
- University of Massachusetts, Donahue Institute. (2004). *A study of MCAS achievement and promising practices in urban special education: A cross-case analysis of promising practices in selected Massachusetts urban public high schools*. Hadley, MA: Author.
- van Dam, P.R.L. (2000). The effects of testing on primary education in the Netherlands: The pupil monitoring system. The effects and related problems of large scale testing in educational assessment. *Foreign Language Teaching and Research*, IAEA.
- Van Stewart, A. (1996). Improving professional student performance on National Board Examinations through effective administrative intervention. In T.W. Banta, J.P. Lund, K.E. Black & F.W. Oblander (Eds.), *Assessment in practice: Putting principles to work on college campuses* (pp. 124-129). San Francisco, CA: Jossey-Bass.
- Venesky, R.L., & Winfield, L.F. (1979). *Schools that succeed beyond expectations in teaching* (No.

1, *Technical Report*): University of Delaware Studies on Education.

- Waits, M.J., Campbell, H.E., Gau, R., Jacobs, E., Rex, T., & Hess, R.K. (2006). *Why Some Schools with Latino Children Beat the Odds...and Others Don't*. Tempe, AZ: Arizona State University, Morrison Institute for Public Policy.
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory (No. 22)*. New York, NY: University of Cambridge.
- Wall, D., & Alderson, J.C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing, 10*, 41-69.
- Wall, D., & Horak, T. (2007). Using baseline studies in the investigation of test impact. *Assessment & Evaluation in Higher Education, 14*(1), 99-116.
- Wang, A.H., Coleman, A.B., Coley, R.J., & Phelps, R.P. (2003). *Preparing teachers around the world*. Princeton, NJ: Educational Testing Service.
- Warwick, D.P., Reimers, F., & McGinn, N. (1989). *Teacher characteristics and student achievement in math and science (No. 5)*. Cambridge, MA: Harvard Institute for International Development.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing, 13*(3), 318-333.
- Waters, T., Burger, D., & Burger, S. (1995). Moving up before moving on. *Educational Leadership, 52*(6), 35-40.
- WestEd. (1999) *Impact of standards-based accountability systems: Evaluation of California's Standards Based Accountability System*. San Francisco, CA: WestEd/MAP.
- Whetton, C. (1992). *The assessment system: Purposes and constraints*. Berkshire, United Kingdom: National Foundation for Education Research.
- White, H.B. (1941). Testing as an aid to learning. In C.C. Ross (Ed.), *Measurement in today's schools* (Vol. 345-346). New York, NY: Prentice-Hall.
- Wideman, R. (2002). Using action research and provincial test results to improve student learning. *International Journal for Leadership in Learning, 6*(20).
- Willey-Rendon, R. (2008). Reading instruction in a high-stakes world: A comparative case study of three fifth-grade teachers. Unpublished Dissertation, Texas Technical University, Houston, TX.
- Williford, A.M. (1997). Ohio University's multidimensional institutional impact and assessment plan. Unpublished Dissertation, Ohio University, Athens, OH.
- Wise, L.L., et al (2005). *Independent evaluation of the California High School Exit Examination (CAHSEE): 2005 evaluation report*. Alexandria, VA: Human Resources Research Organization.
- Wood, R.G. (1953). A twenty-year pilot study of what has become of Ohio's superior high

school graduates. *Tenth Yearbook of the National Council on Measurement in Education*.

Woody, C. (1917). Tests and measures in the schoolroom and their value to the teachers. *School and Society*, 6(184), 61-66.

Woody, E.L., Buttles, M., Kafka, J., Park, S., & Russell, J. (n.d.). *Voices from the field*.

Wright, W.E. (2002). The effect of high stakes testing in inner-city elementary school: The curriculum, the teachers, and the English language learners. *Current Issues in Education*, 5(5).

Yang, X. (1991). Experiments on general high school completion tests in China. In A.J. Luijten (Ed.), *Issues in public examinations: A selection of the proceedings of the 1990 IAEA conference*.

Yeh, S.S. (2006). *Raising student achievement through rapid assessment and test reform*. New York, NY: Teachers College Press.

Yussufu, A., & Angaka, J.A. (2000). *National examinations and their effects on curriculum development and implementation in Kenya. The effects and related problems of large scale testing in educational assessment*. Foreign Language Teaching and Research, IAEA.

Zimmerman, B.J., & Dibenedetto, M.K. (2008). Mastery learning and assessment: Implications for students and teachers in an era of high-stakes testing. *Psychology in the Schools*, 45(3), 206-216.

Zmuda, A., & Tomaino, M. (1999). A contract for the high school classroom. *Educational Leadership*.

Citation: Phelps, R.P. (2011). The Effect of Testing on Achievement: Meta-Analyses and Research Summary, 1910–2010: Source List, Effect Sizes, and References for Qualitative Studies, *Nonpartisan Education Review / Resources*. Retrieved [date] from <http://www.npe.ednews.org/Review/Resources/QualitativeList.pdf>