Éducation comparée / nouvelle série, n°5, pp. 17-34

Making Educational Accountability System Work for Equity

Learning from the U.S. Experiences and Studies

Jaekyung Lee
State University of New York
at Buffalo,
États-Unis

RÉSUMÉ

Cet article propose une synthèse critique de recherches sur l'impact des tests externes à « enjeux forts » et l'impact des politiques d'imputabilité fondées sur ces tests. A partir de l'expérience nord-américaine, nous pouvons identifier ce qui marche et ce qui ne marche pas pour améliorer la réussite des élèves et réduire les inégalités de réussite, notamment entre communautés ethniques. Une récente méta-analyse synthétisant des travaux antérieurs relatifs aux effets des politiques d'imputabilité sur la réussite des élèves en lecture et en mathématiques ne fait apparaître que des effets modestes sur les scores moyens des élèves et une absence d'effets sur l'écarts entre groupes ethniques, de même qu'une grande variabilité d'impacts et, au final, des interrogations sur cet effet. Les politiques antérieures orientées strictement sur la mesure et la sanction des performances ont créé l'illusion de progrès scolaires et encouragé une inflation des tests, mais n'ont pas pu traiter correctement les différences entre écoles en termes de conditions de scolarisation, en particulier en ce qui concerne les ressources à leur disposition. En construisant un modèle fondé sur le potentiel plutôt que sur des déficiences, cette revue de la littérature identifie des dimensions d'intervention critiques vis-à-vis des élèves et des écoles. La discussion suggère des recommandations en termes de politiques à développer pour la prochaine génération de politiques d'imputabilité.

MOTS CLÉS

Politiques d'imputabilité, évaluation externe, équité, inégalités, No child left behind

ABSTRACT

This article provides a critical synthesis of prior research on the impact of high-stakes testing and test-driven external accountability policy. Based on the U.S. experiences and studies, we learn what works or does not work for improving achievement and closing the gap for poor minority students. The meta-analysis of prior research on test-driven external accountability policy effects on student achievement in reading and math finds only modest effects on the average academic achievement and no effects on the racial achievement gap, while it also reveals variability and uncertainty of the evidence. The past performance-guarantee policy approach often generated an illusion of academic progress by giving incentives for test score inflation, but failed to address inadequacies and disparities in terms of schooling conditions, resources and capacity. Building on a model of strength as opposed to a model of deficiency, the review identifies critical areas of need for low-performing disadvantaged minority students and effective school practices. The discussion provides policy recommendations for the next generation of educational accountability systems.

KEYWORDS

accountability, high-stakes testing, equity, achievement gap, NCLB

Introduction

The origin of public high-stakes testing dates as far back as the 2nd century B.C. in China. Since then, modern external examinations have grown as instruments of control over educational systems in both Eastern and Western countries (Eckstein & Noah, 1993). While the national history of high-stakes testing is relatively short in the United States (which lacks a single national examination system), individual states have had their own statewide assessment systems based on different curriculum and performance standards. The U.S. school system suffers from the lack of common rigorous educational standards and inequalities of school and teacher qualities. In contrast, East Asian countries (e.g., Japan, Korea, Singapore) with rigorous national standards and testing systems have suffered from overstandardization, excessive competition, and private tutoring problems. Educational accountability movement in the U,S. during the past three decades or so may be seen as an attempt to benchmark

the East Asian model of academic success (Lee, 2001). The movement may have been successful in that the policy promoted higher standards and greater attention to the achievement gap problem with the requirement of disaggregation of achievement data by student subgroups. However, it was not so successful in that it failed to upgrade the culture and capacity of disadvantaged low-achieving schools that need the most help. The challenge is to create an American version of a standards-based education system that combines common national standards for rigor and equity with local autonomy for innovation and creativity.

The No Child Left Behind Act of 2001 (NCLB; Pub.L. 107-110, enacted January 8, 2002) has changed the traditionally decentralized system of public education with new standards and testing mandates for all states. It is broadly aimed at achieving student proficiency in reading and mathematics through annual testing of all students in grades 3-8 and one high school grade across all states. NCLB relies on high-stakes testing to ensure that schools make adequate yearly progress (AYP) toward the goal of 100% student proficiency in these subjects by 2014. The closing of achievement gaps between identifiable subgroups of American students is a major thrust of this legislation. Schools are required, under threat of strict sanctions, to raise achievement each year in math and reading and to eliminate the achievement gap by race, ethnicity, poverty, language, and special education status.

Can the nation meet the dual challenges of improving both academic excellence and equity at the same time? The goal is two-fold: not simply narrowing the relative gap, but also raising the overall level of achievement. This change involves raising performance standards from a basic to proficient level and targeting interventions for those needing improvement. This adds a new dimension of gap relative to adequate performance standard, thereby removing the possibility of lowering the achievement of the higher performing group in order to close the relative gap. Recent analyses of NAEP achievement trends before and after the NCLB suggest that the NCLB did not yet have significant impact on improving average reading and math achievement and narrowing the achievement gaps among racial and

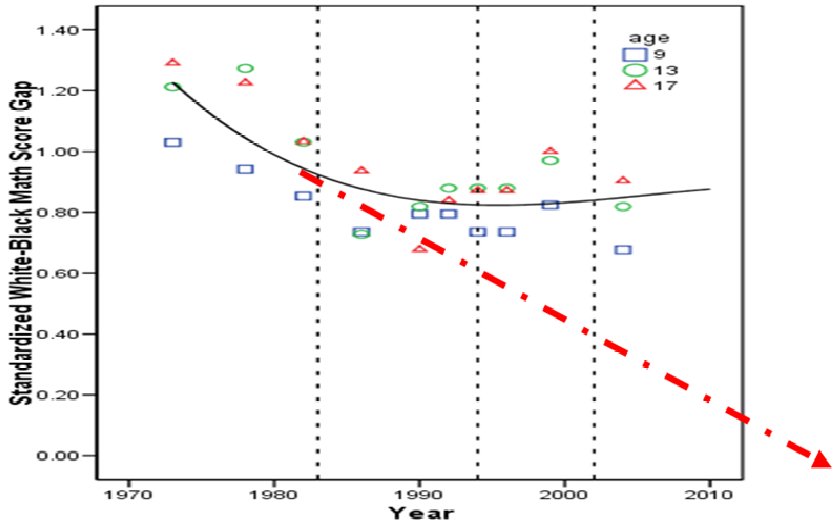
social groups (Lee, 2007). However, there continues to be intense debates about whether external, test-driven accountability policy enhances or hinders academic excellence and equity. In light of these concerns, this article gives a critical synthesis of prior research on educational accountability models, policies, and test results in the U.S.

Educational Policy Shifts and Trends of Achievement Gaps

Achievement gaps loom large between students in the U.S. and other industrial countries. In the midst of an international brain race, school reform movements during the past two decades were driven partly by international math and science test results (Baker, 2003; Lee, 2001). Since A Nation at Risk (NCEE, 1983) lamented mediocre performance by comparing the U.S. with Japan and Germany (industrial competitors of the U.S. at that time). Setting the national goal of being first in the world in mathematics achievement, Goals 2000 program was enacted into law in 1994, pushing for national standards and assessments (Ravitch, 1995). At the same time, the Improving America's Schools Act of 1994 (IASA) reinforced the performance-based accountability movement. The culmination of this policy shift is seen in the most recent federal educational policy initiative, the No Child Left Behind Act of 2001 (NCLB).

The National Assessment of Educational Progress (NAEP), the so-called nation's report card, provides information on the state of students' academic achievement in core academic subjects. For example, there is a documented achievement gap in mathematics between White students and minority students in the U.S., particularly Black and Hispanic students. For example, an average Black high school 12th grader's standardized math test score can be as low as that of an average White 8th grader. While the racial and social achievement gaps were greatly narrowed during the 1970s and 1980s, there were still significant setbacks in national progress toward closing achievement gaps in the 1990s (Lee, 2002). While the task of closing the gap could have been accomplished without the setback, it

is not clear whether and how past enormous gains can be repeated within the next generation or so (Peterson, 2006). Figure 1 shows the trend of the Black-White gap in NAEP math as marked by reference lines of key policy events: a Nation at Risk of 1983, Goals 2000 and IASA of 1994, and NCLB of 2001. The racial gap stopped narrowing further after key national educational events that influenced educational accountability movement with a policy shift from equity



to excellence. Similar trends are found for socioeconomic achievement gaps.

Figure 1. NAEP Long-term Trends in National Black-White Reading Score Gaps at age 9, 13, 17,in standard deviation units (solid black line shows the actual trend and broken red line show the expected trend)

In the 1960s and 70s, education and social policies worked for narrowing the racial gap by guaranteeing a minimally adequate level of achievement for minorities through compensatory education, minimum competency testing, school desegregation, equalization of school funding, the War on Poverty, and affirmative action[1]. While

[1] The "War on Poverty" legislation, first introduced by U.S. President Lyndon B. Johnson in 1964, expands the government's role in social welfare and education

the focus of education policy shifted from equity to excellence since the 1980s, it might have had some chilling effects on the earlier efforts of gap reduction. Further, social and educational conditions for minority students in high poverty urban schools have deteriorated with school resegregation by race and social class as well as with the influx of immigrant student populations with limited English proficiency. Despite this reversal, it must be noted that there were several limitations of the past racial achievement gap reductions. First, racial gaps narrowed when White students did not improve and only Black and Hispanic groups improved in their achievement scores. Second, the narrowing of the racial achievement gap occurred mostly at the basic skills level. Third, the narrowing of the racial achievement gap occurred mostly in preschool or primary school years (with less progress in middle or high school).

Did NCLB make any difference in the achievement gap trends ? After NCLB, some states have claimed progress in closing the achievement gap. However, Lee's study (2010) reveals some states' "trick or treat" strategies under NCLB. States that chose the "trick" path with an illusion strategy made gains on their own state assessments but not on NAEP. While NCLB makes state participation in the NAEP assessment mandatory, the federal government still allows individual states to base their school accountability decisions on state assessment results. Consequently, some states have lowered their student performance standards in order to raise student proficiency rates under NCLB. In contrast, other states that chose the "treat" path with full implementation of the federal policy have not yet brought gains on either the state assessment or NAEP. If we continue the current policy course under NCLB, academic proficiency is unlikely to improve significantly on NAEP, and it is possible that

---

programs. For example, Project Head Start was designed to help end poverty by providing preschool children from low-income families with a program that would meet emotional, social, health, nutritional, and psychological needs. Affirmative action refers to policies that take race/ethnicity, gender, social class into consideration for promoting equal opportunities of employment and education.

the state assessment will continue to give a false impression of progress (Lee, 2006).

Contested Models of Accountability

Input-guarantee vs. Performance-guarantee

Accountability often has multiple meanings and purposes, and there are several models of educational accountability (see Darling-Hammond, 1989; Adams & Kirst, 1999; Linn, 2003; Harris & Herrington, 2006). The issue of who holds whom accountable and for what purpose has been contentious in the history of educational accountability (Dorn, 1998). Despite the historical debate, an accountability model that is performance-driven, test-driven, measurable, and statistical in nature came to dominate current policy and practice. The logic of performance-driven accountability policy appears to draw upon rationalistic and behavioristic views of human behavior. By positing that holding schools, teachers, and students accountable for academic performance, with incentives provided (i.e., rewards and sanctions), will inform, motivate, and reorient the behavior of schooling agents toward the goal (Benveniste, 1985; Wise, 1979). However, studies challenge simple, rationalistic approaches to performance-driven accountability policy and identify contingencies for policy success.

State policy approaches to accountability tend to favor either a primary emphasis on input guarantees as a mechanism for achieving equal access to learning or a primary emphasis on performance guarantees (Elmore & Fuhrman, 1995). The advocates of opportunity-to-learn (OTL) standards argue that every student must have equal access to high-quality learning by specifying key inputs (per-pupil spending, textbooks, teacher training, and the like) in the form of binding standards. The achievement gaps among racial and socioeconomic groups were found to be relatively small in states with more equitable distribution of school resources and classroom OTL (Wong & Lee, 1998). In contrast, the critics of OTL standards argue

that input standards of any kind create constraints on how educators respond to the learning problems of students. While the critics argue that holding schools and students accountable for performance creates incentives for schools to find out which practices work most effectively, there is no guarantee that performance measures automatically lead to increased equality in the distribution of student learning (O'Day & Smith, 1993).

Indeed, the two approaches are not mutually exclusive, and combining them may be the most successful path. Figure 2 illustrates a model of educational accountability with such combination of input-guarantee and performance-guarantee approaches. While the input-guarantee policy provides resources and opportunities for capacity-building, the performance-guarantee policy provides incentives in return for better results. Can test-driven external accountability policy improve both incentives and capacity? This question can be divided into two components. Do you want to do it? Can you do it? High-stakes testing policy often attempts to address the first question but not the second question. The policy is grounded in the theory that high-stakes testing based on common measurable standards for what students should know and be able to do will motivate the improvement of student achievement outcomes. Holding schools and students accountable for results may motivate them to improve their efforts for teaching and learning. Without adequate resources and capacity, however, high-stakes testing can bring perverse incentives to inflate or manipulate test results (e.g., through teaching to the test, coaching, exclusion or retention of low-performing students). Further, the impact of external accountability policy depends on how it interacts with internal accountability at the school and classroom levels; school administrators and teachers' collective professional responsibility for student learning as well as parents and community members' support for school improvement.

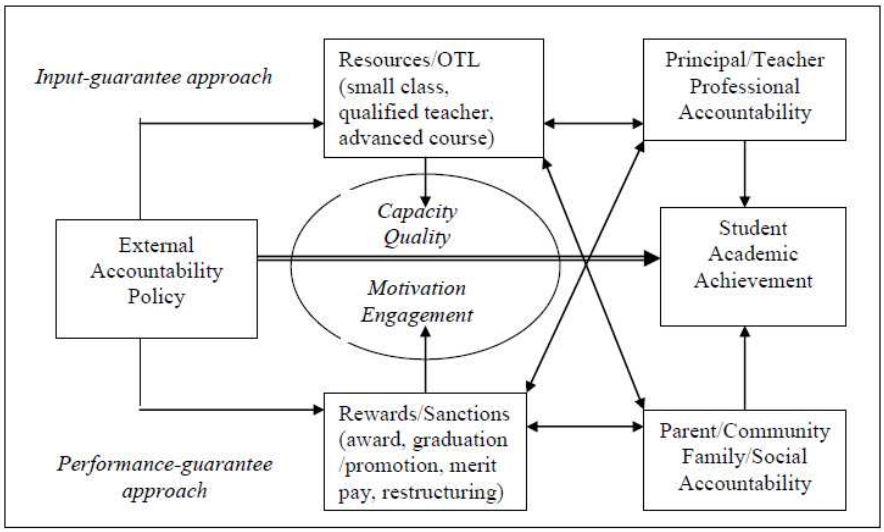


Figure 2. An Integrated Model of Educational Accountability Policy

Despite the expectations of shared responsibility forimproving academic achievement, concerns about exclusive focus on educators and/or students and ignorance of other responsible parties including policymakers were raised (Elmore, 2005; Linn, 2003; Porter & Chester, 2002). Studies show that assessment and accountability policies imposed changes in schools with little to no support and this unfunded mandate has limited impact on achievement gaps (Lee & Wong, 2004). The case that drew the most attention in the U.S. prior to NCLB was Texas, where the evidence on the effects of high-stakes testing on equity was mixed and often contradictory (Haney, 2000; Skrla, Scheurich, Johnson, & Koschoreck, 2004; Valencia, Vanezuela, Sloan, & Foley, 2004). While NCLB builds upon the alleged success of first generation accountability states such as Texas and North Carolina (that had adopted test-based accountability systems prior to NCLB), there is highly mixed evidence on high-stakes accountability policy effects. A recent meta-analysis (76 effect size estimates from 14 selected studies) by Lee (2008) shows a modestly positive effect on average with highly mixed evidence, but no positive effect on narrowing the racial achievement gap.

What Works for Disadvantaged Minorities

Shift from Model of Deficiency to Model of Strength

The achievement gap among student groups were shown to emerge prior to their school entry and widen over the course of schooling, generating further opportunity gap for high school graduation and college entrance. Although family risk factors (e.g., poverty, limited English proficiency) push down the growth curve, schools must protect disadvantaged minority kids from the risk of academic failure, and promote their academic resilience (Borman & Overman, 2002; Finn & Rock, 1997; Wang, Haertel & Walberg, 1994). Particularly, early childhood education programs such as intensive high-quality preschool contribute to sustainable academic growth and educational attainmentfor disadvantaged children (Campbell et al., 2001; Fusaro, 1997; Gormley et al., 2005; Lee, Brooks-Gunn, Schnur & Liaw, 1990; Ramey & Ramey, 1998; Reynolds et al., 2002). Adequate funding for qualified teachers and small classes are also crucial in closing the gap during early childhood education years (Kane & Staiger, 2005; Krueger, 1999).

There are some stories of high-poverty urban or rural schools that achieve more with fewer resources. These schools were remarkably successful in improving academic achievement of low-income minority students (Barth et al., 1999; Carter, 2000; Mathews, 2009). The key to academic success for disadvantaged minority students is to balance academic press with social support. Externally set standards and tests may serve as extrinsic motivations for academic press, but that their ultimate effects on academic performance depend on school capacity and social support (Newmann, King, & Rigdon, 1997; Lee & Smith, 1999). A study on the impact of high-stakes testing in Chicago suggests that the additional resources provided to at-risk students, such as reduced class sizes, after-school tutoring, and summer school programs, enhances achievement gains (Roderick, Bryk, Jacob, Easton,

& Allensworth, 1999). Further, high standards and high-stakes testing without adequate support often have negative consequences such as disengagement and dropout problems. As traditional communal institutions like families and neighborhoods have become less stable and supportive, it becomes more critical for schools to provide more personalized support and caring environment for at risk students with increased out-of-school time support (e.g., after-school and summer school wraparound services). Studies also show that a supportive school is necessary but not sufficient for improving learning. Small schools are often designed to address this problem by affording more individualized and caring learning environment and social support. A recent study shows that Chicago small school conversion experiment provided a more personalized and supportive school context for students but this structural change did not lead to instructional and academic improvement (Kahne et al., 2008). In brief, small schools can help develop a more safe, orderly, caring, and supportive community of learning. This environment is necessary for academic improvement, but not sufficient.

High standards cannot succeed without increased time and effort for learning. High-performing East Asian countries (e.g., Korea, Japan, and Taiwan) have significantly longer school years (about more than 40 school days) than the United States. A successful example of extended time and effort model in the U.S. can be found in a San Francisco-based network of charter schools called the Knowledge is Power Program (KIPP) that operates on a schedule, approximately 60% longer than the conventional 180 six-hour day schedule.[2] A study of KIPP schools also emphasizes collective responsibility for student learning (Woodworth et al., 2008, p. 79): "Ultimate accountability for results is also part of this equation, one that school leaders sign on to when they take the job. But it is not the external accountability that drives day-to-day operations; it is internal or professional accountability."

2 Under the mottos of "No Short-Cuts" and "Hard Work", the KIPP model features: (1) A culture of high expectations for student academic performance and behavior; (2) extensive time and support for student learning; (3) A focus on tracking student progress and careful instructional planning; (4) A philosophy of continuous improvement and shared norms of professional accountability (Mathews, 2009).

Next Generation of Accountability System

Challenges and Recommendations

The critical review of studies and policy experiences in the U.S. gives some implications for the design and implementation of new educational accountability system. We need to redefine the concept of educational accountability from a broader, systemic and symmetric perspective. The central question for accountability decision-making may be posed: "Who is held accountable for what, how and why?" (Lee, 2007).

Regarding the issue of "whom" to target for accountability, it should involve all parties of education. Despite the expectations of shared responsibility or symmetric accountability for academic improvement, concerns about exclusive focus on educators and ignorance of other responsible parties have been raised. The target of NCLB accountability focuses on schools and there is an asymmetry of responsibility and pressure. Policymakers need to be held accountable for providing adequate support to build school capacity and implement enacted policies. For successful local adaptation and innovation, internal accountability must precede external accountability.

Regarding the issue of instruments ("how") for accountability, both inducement and capacity-building have been used as policy tools less frequently by states than mandates. The problem is that incentives and assistance are only rarely offered, and are also allocated in a way that benefits only those who are already highly motivated and well-prepared for teaching and learning. Accountability policy should involve efforts to improve the distribution of school funding and highly qualified teachers for disadvantaged schools. At the same time, however, we should address areas of inefficiencies in current school spending along with increased funding. More drastic public school

reform efforts such as restructuring and school choice (charter school, voucher) brought highly mixed results and cannot be a substitute for capacity-building and investment. An outstanding question is whether successful charter schools such as KIPP model can be transferrable to all public schools for scale-up .

Regarding the issue of “what” to measure and improve for accountability, we need to look at a complete picture of educational system and to expand narrow measures of school performance as measured by standardized test scores. While most states use proficiency rates as measures of school performance under NCLB, they do not capture value-added contributions of schools to improvement of student achievement. We should shift the focus of evaluation from status (how well school perform) to growth (how much schools improve). Further, we need to set much more realistic performance goals and timelines. It is necessary to guard against potential bias such as the inflation of proficiency levels and gains by adopting multiple measures and more than one performance standard (Proficient). Major concerns are that the state's emphasis on Proficiency as the "gold standard” overlooks the progress made by lower-performing students who meet or exceed the minimum competency level and has a chilling effect on teachers who have improved a significant number of low-achieving students below the Proficient level.

Regarding the issue of “why”, we need to revisit the goals of accountability and examine reasons for policy success or failure in achieving the goals. NCLB may have been successful in disseminating more standardized tests to schools, drawing their attention to test results on the achievement gap, and generating pressure for schools to fix the achievement gap. This resulted in narrowing the testing gap between the U.S. and other developed countries. However, the long-term success of the school accountability system does not depend on the sheer number of test administrations or the number of passing schools and students on standardized tests but on the genuine progress of student learning and successful transition into work or postsecondary education. The scope of an educational accountability system should go beyond traditional focus on K-12 (elementary and

secondary education) and expand to P-16 education for the seamless transition of students from pre-school through completion of the college degree. The ultimate test of NCLB is whether schools provide every single child with opportunities to realize their potential for full development regardless of any educationally irrelevant background characteristics such as race/ethnicity and social class, thus improving equity and closing achievement gaps.

Références bibliographiques

ADAMS, J. E., & KIRST, M. W. (1999). New demands and concepts for educational accountability: Striving for results in an era of excellence. In J. Murphy & K. S. Louis (Eds.), *Handbook of research on educational administration* (pp. 463–490). San Francisco: Jossey-Bass.

BAKER, D. P (2003). Should we be more like them ? Reflections on causes of cross-national high school achievement differences and implications for American educational reform policy. In D. Ravitch (Ed.), *Brookings papers on education policy* (pp.309-325). Washington DC: Brookings Institution.

BARTH, P. et al. (1999). *Dispelling the myth : High-poverty schools exceeding expectations*. Washington, DC : Education Trust.

BENVENISTE, G. (1985). *The design of school accountability systems. Educational Evaluation and Policy Analysis*, 7(3), 261-279.

CAMPBELL, F. A., PUNGELLO, E. P., MILLER-JOHNSON, S., BURCHINAL, M., & RAMEY, C. T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology*, 37(2), 231-242.

CARTER, S. C. (2000). *No excuses : Lessons from 21 high-performing high-poverty schools*. Washington, DC: The Heritage Foundation.

DARLING-HAMMOND, L. (1989). Accountability for professional practice. *Teachers College Record*, 91, 59–80.

DORN, S. (1998). The political legacy of school accountability systems. *Education Policy Analysis Archives*, 6(1). Retrieved November 5, 2003, from http://epaa.asu.edu/epaa/v6n1

ECKSTEIN, M. A., & NOAH, H. J. (1993). *Secondary school examinations : International perspectives on policies and practices*. New Haven, CT : Yale University Press.

ELMORE, R. F. (2005). Agency, reciprocity, and accountability in democratic education. In S. Fuhrman & M. Lazerson (Eds.), *The Public Schools*. New York: Oxford University Press.

ELMORE, R. F., & FUHRMAN, S. (1995). Opportunity-to-learn standards and the state role in education. *Teachers College Record*, 96, 432–457.

FINN, J. D., & ROCK, D. A. (1997). Academic success among students at risk for school failure. *Journal of Applied Psychology*, 82, 221-234.

FUSARO, J. (1997). The effect of full-day kindergarten on student achievement : A meta-analysis. *Child Study Journal*, 27 (4), 269-277.

GORMLEY, W.T., GAYER, T., PHILLIPS, D., & DAWSON, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology*, 41(6), 872-884.

HANEY. W. (2000). The myth of the Texas miracle in education. *Educational Policy Analysis Archives*. Retrieved March 3, 2001, from http://epaa.asu.edu/epaa/v8n41

HARRIS, D. N., & HERRINGTON, C. D. (2006). Accountability, Standards, and the Growing Achievement Gap : Lessons from the Past Half-Century. *American Journal of Education*, 112, 209–238.

KAHNE, J. E., SPORTE, S. E., DE LA TORRE, M., & EASTON, J. Q. (2008). Small high schools on a larger scale: The impact of school conversions in Chicago. *Educational Evaluation and Policy Analysis*, 30(3), 281-315.

KANE, T. J., & STAIGER, D. O. (2005). *Using imperfect information to identify effective teachers*. Unpublished Paper. Los Angeles: School of Public Affairs, University of California–Los Angeles.

KRUEGER, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics* 114 (2): 497–532.

LEE, J. (2001). School reform initiatives as balancing acts : Policy variation and educational convergence among Japan, Korea, England and the United States. *Educational Policy Analysis Archives*.

LEE, J. (2006a). *Tracking achievement gaps and assessing the impact of NCLB on the gaps : An in-depth look into national and state reading and math outcome trends*. Cambridge, MA : The Civil Rights Project at Harvard University.

LEE, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity ? *Educational Researcher*, 31(1), 3–12.

LEE, J. (2007). *The Testing Gap : Scientific Trials of Test-driven School Accountability Systems for Excellence and Equity*. Charlotte, NC: Information Age Publishing.

LEE, J. (2008). Is Test-driven External Accountability Effective? Synthesizing the Evidence from Cross-State Causal-Comparative and Correlational Studies. *Review of Educational Research*, 78(3), 608-644.

LEE, J. (2010). Trick or Treat: New Ecology of Education Accountability System in the USA. *Journal of Education Policy*, 25(1), 73-93.

LEE, J., & WONG, K. K. (2004). The impact of accountability on racial and socioeconomic equity: Considering both school resources and achievement outcomes. *American Educational Research Journal*, 41(4).

LEE, V. E., BROOKS-GUNN, J., SCHNUR, E., & LIAW, F. R. (1990). Are Head Start effects sustained ? A longitudinal follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Child Development*, 61, 495-507.

LEE, V. E., & SMITH, J. B. (1999). Social support and achievement for young adolescents in Chicago : The role of school academic press. *American Educational Research Journal*, 36(4), 907–945.

LINN, R. L. (2003). Accountability : Responsibility and reasonable expectations. *Educational Researcher*, 32(7), 3–13.

MATHEWS, J. (2009). *Work Hard. Be Nice. How Two Inspired Teachers Created the Most Promising Schools in America*. Algonquin Books.

NEWMANN, F. M., KING, M. B., & RIGDON, M. (1997). Accountability and school performance: Implications from restructuring schools. HARVARD EDUCATIONAL REVIEW, 67(1), 41–74.

No Child Left Behind Act of 2001, Pub. L. No. 107-110.

O'DAY, J., & SMITH, M. (1993). Systemic reform and educational opportunity. In S. Fuhrman (Ed.), *Designing coherent educational policy* (Ch. 8). San Francisco: Jossey-Bass.

PETERSON, P. E. (Ed.). (2006). *Generational change: Closing the test score gap*. Lanham, MD: Rowman & Littlefield.

PORTER, A. & CHESTER, M. (2002). Building a high-quality assessment and accountability program : The Philadelphia example. In D. Ravitch (Ed.). *Brookings Papers on Education Policy*. (pp. 285-315). Washington DC: Brookings Institution.

RAMEY, C.T., & RAMEY, S.L (1998). Early intervention and early experience. *American Psychologist*, 53(2), 109-120.

RAVITCH, D. (1995). *National standards and American education*. Washington, DC: Brookings.

REYNOLDS, A. J., TEMPLE, J. A., ROBERTSON, D. L., & MANN, E. A. (2002). Age 21 cost-benefit analysis of the Title I Chicago Child-Parent Centers. *Educational Evaluation and policy Analysis*, 24(4), 267-303.

RODERICK, M., BRYK, A. S., JOCOB, B. A., EASTON, J. Q., & ALLENSWORTH, E. (1999). *Ending social promotion in Chicago*. Chicago : Consortium on Chicago School Research.

SKRLA, L., SCHEURICH, J. J., JOHNSON, J. F., & KOSCHORECK, J. W. (2004). Accountability for equity: Can state policy leverage social justice ? In L. Skrla & J. J. Scheurich (Eds.), *Educational equity and accountability: Paradigms, policies, and politics* (pp. 51–78). New York, NY: RoutledgeFalmer.

WANG, M. C., HAERTEL, G. D., & WALBERG, H. J. (1994). Educational resilience in inner-cities. In M. C. Wang & E. W. Gordon (Eds.*), Educational resilience in inner-city America: Challenges and prospects* (pp. 45-72). Hillsdale, NJ: Erlbaum.

WISE, A. E. (1979). *Legislated learning: The bureaucratization of the American classroom*. Berkeley, CA: University of California Press.

WONG, K. K., & LEE, J. (1998). Interstate variation in the achievement gap among racial and social groups : Considering the effects of school resources and classroom practices. In K. K. Wong (Ed.), *Advances in Educational Policy* (Vol. 4, pp. 119–144). Greenwich, CT: JAI Press.

WOODWORTH, K. R. et al. (2008). *San Francisco Bay Area KIPP Schools: A Study of Early Implementation and Achievement Final Report*. Menlo Park, CA: SRI International.