

QUT Digital Repository:  
<http://eprints.qut.edu.au/>



Luke, Allan and Woods, Annette F. (2008) Accountability as testing: Are there lessons about assessment and outcomes to be learnt from No Child Left Behind? . *Literacy Learning: The Middle Years* 16(3).

© Copyright 2008 Australian Literacy Educators' Association

## **Accountability as testing: Are there lessons about assessment and outcomes to be learnt from *No Child Left Behind*?**

**Allan Luke & Annette Woods** | Queensland University of Technology

*An earlier version of this paper appeared in the Professional Magazine 22 (Nov 2007) published by the Queensland Teachers' Union, Brisbane*

The fact that debate has continued over literacy teaching for the past three years since the 2005 release of the National Inquiry into the Teaching of Reading (National Inquiry into the Teaching of Reading, 2005), and that recent rearticulations of the Report's findings by its lead author (See for example Milburn, 2008) continue to take headline space, reminds us that literacy education remains a contentious policy and pedagogic issue for communities, schools, systems, teachers and students – and for politicians. During the past three to four year period we've all watched the latest *literacy crisis* played out in the pages of our newspapers and television current affairs shows. This *crisis* has predictably led to policy and curriculum initiatives offering simplistic solutions to the latest perceived problems. Under the last Conservative Federal Government, these Australian media and policy responses paralleled the debates in the United States over the No Child Left Behind Act (United States of America, 2001). So in a context where accountability is being narrowly framed as testing, and literacy likewise framed as basic decoding skills, are there lessons to be learnt for Australian teachers and policy makers in the No Child Left Behind legislation?

Research dispelling the success of NCLB has been available since its inception, but more recently the official reports have been calling the policy decisions implemented as part of this legislation into question, and political support for both NCLB and the *Reading First* program is beginning to waver. In this short paper we first lay out a brief introduction to the NCLB legislation and its policy effects. We document the official results and the critiques. We then suggest some lessons that Australian policy-makers and educators must consider as the decision about how best to promote a high quality / high equity system for all Australian school children is made in the new political context. We aim to offer a scientific and bibliographical resource for teachers who wish to engage with the debate.

### **What is *No Child Left Behind*?**

The No Child Left Behind Act (United States of America, 2001) (NCLB) began from a focus on improved quality of early literacy acquisition. It was launched as a bipartisan

move for US schooling to address criticisms of lagging literacy standards and a growing achievement gap for students from diverse backgrounds. A significant assumption behind the policy is: that a program of teaching basic literacy code breaking skills and regular testing can generate more equitable results and better achievement by students from cultural and linguistic minority and lower socio-economic backgrounds. NCLB is based on four basic principles: increased accountability; increased flexibility and local control; expanded options for parents; and teaching methods based on a “gold standard evidence base” ([www.ed.gov/programs/readingfirst/legislation.html](http://www.ed.gov/programs/readingfirst/legislation.html)). The legislation sets out a regime of accountability as standardised testing and public reporting, with a drive toward consistency across system and school contexts. What will count as evidence-based pedagogy is set by mandating scripted, highly prescriptive reading curriculum programs and methods, approved by the Federal government. We have chosen here to not enter the debate related to the ongoing conflict-of-interest controversies about the US Federal government selection of programs for funding (Grunwald, 2006) which continues to rage in the US and that is a matter of ongoing contention in Congressional hearings.

Within NCLB, compliance in relation to teaching methods and testing measures is controlled through sanctions and incentives provided for districts, schools, and teachers. Parental school choice is set up as a way to let the market drive the system: provide the facts about individual school performance and the market will decide. The ‘facts’ in this case are provided through published league tables of school test performance. There is, of course, the assumption that access to choice can be established as a level playing field for all, with voucher systems in place in some states for families from traditionally disadvantaged communities. To those who have been following the Australian policy debates – these propositions would be familiar.

Recent criticism and debate around the effectiveness of the Act and its associated policy began to gain voice in 2007 when the media first reported that the scheduled reauthorisation of the Act originally scheduled for 2007, was likely to be delayed until 2009. More recently the release by the U.S. Department of Education of a report on the impact of *Reading First* - the \$1 billion-a-year reading program that has been a pillar of the Bush administration's education plan as epitomised by NCLB – reported that while *Reading First* had significantly increased average instructional time spent on five essential components of reading instruction (phonemic awareness, phonics, vocabulary, fluency and comprehension) in classrooms, that there had been no

change in the reading comprehension scores of students since the implementation of the *Reading First* initiatives (Institute of Educational Sciences, 2008). The key criticisms of the legislation have been extensively detailed in the literature (see as examples: Allen et al., 2007; Coles, 2003; Encisco, Katz, Kiefer, Price-Dennis, & Wilson, 2007; Lee, 2006; Luke & Woods, 2008; Nichols, Glass, & Berliner, 2005; Pennington, 2007). Early critiques were based on claims that only a very limited selection of research was taken into account in the National Reading Panel (2000) findings that formed the basic foundation for the NCLB initiatives. Yatvin, Weaver and Garan (2003) claim that the Reading Panel Report is misleading when it claims that 100,000 instances of research were considered. In fact this number represents the estimated number of research papers and reports published in the 30 year period considered, with a much smaller number of research papers actually being judged to fit the 'gold standard' of evidenced-based research set up a priori to the research review. The effect of this narrow definition of evidence-based research was the omission of much of the long standing educational research base – including all correlation studies and those that related to single case sites. In reality very small numbers of studies from the possible thousands were included in the recommendations and findings of the Panel. As an example the claims about the effectiveness of phonic instruction in early reading were made on the basis of just 38 studies, with key contradictory research excluded from the review.

Further, the National Reading Panel and subsequent NCLB policy have been critiqued for: discrepancies between the findings of sub group reports and the final reported findings; the omission of findings related to English language learners (ELLs) and the invisibility of special needs issues; incomparability of results across different populations being ignored as findings of a variety of research studies were conflated; and errors in the predictive validity of early reading achievement assessments (see as examples of this large research base Allington, 2002; Coles, 2003; Garan, 2002; Yatvin, Weaver & Garan, 2003). In a broader critique, Gee (2000) linked the panel's approach to science with its failure to engage with well-documented new cultures, technologies, and practices of literacy.

NCLB is premised upon the assumption that early intervention and acquisition of "alphabets" has longitudinal effects. In his reanalysis of the empirical research bases on early intervention, Paris (2005) makes an empirical and theoretical case that the popular push back to phonic programs as reading instruction has misconstrued the longitudinal and developmental effects of early achievement of the "constrained skills" of alphabets. He argues that achievement of the "unconstrained

skills” of comprehension—including vocabulary knowledge, inference, and critical analysis—are much stronger and robust predictors of later academic achievement. The same case is put strongly by Calfee (2003), whose analysis of California state test data noted that lower socio-economic, linguistic, cultural minority, and other at risk students experience significant problems in the transition to secondary school, even where early intervention programs have been put in place. There is a range of plausible explanations for this phenomenon, with traditional comprehension research suggesting the key role of vocabulary knowledge in sustainable upper primary and middle years reading achievement (for an overview see Alvermann, 2002).

### **The official results and first round of reanalysis**

Three years after its inception, the first official results of the policy initiative’s effectiveness were published (Centre on Education Policy (CEP), 2003). In what are effectively state self-reports, 73% of states and 72% of districts reported improvement on state tests over the first three years of NCLB. Further, 21 states reported a narrowing of the Hispanic achievement gap and 18 states similarly reported a narrowing of the African American achievement gap. These figures suggest improvement, although not of the scale and mass necessary to reach targets of 100% of students at grade proficiency by 2014. Issues related to equitable distribution of trained teachers across all schools, a lack of support for English Language Learning (ELL) and special needs students, a narrowing of the curriculum and issues of content and construct validity and confidence interval problems with state testing and reporting on some counts, were also reported. The CEP and other organisations have also noted the problems with state self-reporting. Some states struggle to establish reliable and valid testing and reporting systems, and lack reliable data that would enable meaningful longitudinal comparisons of systemic interventions.

In a major study for the Harvard Civil Rights Project, Lee (2006) used the US National Assessment of Educational Progress NAEP national testing data to reanalyze states’ performance claims. Lee’s findings show no significant positive impacts on NAEP reading or mathematics achievement since the inception of NCLB, with flat or slight declines in reading achievement. Despite some positive transient improvement in Grade 4 mathematics after NCLB, these positive effects diminished and achievement returned to pre-NCLB rates after Grade 4. Lee found that there were no signs that the achievement gaps for at risk groups were diminishing. In itself this finding is troubling, but perhaps of more broad concern are the plateau effects that can be seen in the results of those states that moved toward test-based

accountability systems in the 1990s (e.g. Texas, North Carolina, and Florida). Lee's analysis concludes that state achievement tests tend to significantly inflate proficiency levels and underestimate the ongoing racial and social achievement gaps across all states: "The higher the stakes of state assessments, the greater the discrepancies between NAEP and state assessment" results (p. 11). It is unsurprising that cases of state-level test score misrepresentation and, in instances, outright fraud are currently under investigation in six states, with media reports of numerous other cases.

### **Phonics and the standardisation of teaching**

There are common themes between NCLB and our own recent responses to improving literacy standards and outcomes in Australia. What is distinctive about this current debate, across both contexts, has been the rise of a 'gold standard' of evidence-based research as the major criterion for deciding what will be considered 'valid' as evidence of success in literacy teaching. The gold standard for NCLB, set within the foundation reports of the National Reading Panel (2000) and affiliated studies (Snow, Burns & Griffin, 1998), is based on a number of propositions.

However, it begins from what we term the *phonics hypothesis* which states that there is scientific evidence that literacy achievement can be improved through systematic curricular approaches to pedagogy that emphasise 'alphabets' or phonics. In a major and award-winning reanalysis of reading research, Paris (2005) claims that "most of the scientific evidence about reading skills and reading development, particularly relating to decoding skills, is based on inadequate theories, measures and interpretations" (p. 201). He also encourages a broadening of our understandings of what 'valid' reading assessments might be and claims that this requires new theories of reading skills and development to achieve. He argues, "alphabets" are of importance and are *necessary but not sufficient* for sustained reading gains.

In classic binary logics, once the *phonics hypothesis* is tabled and presented as truth, the inverse of this proposition is taken as fact also. That is that other 'methods' that are not phonic based, and that have not been verified by the 'gold standard' of randomized field trials actually *contribute* to current patterns of early literacy achievement and failure generally, and more specifically to the underperformance of minority and lower socio-economic groups. The Australian report released in 2005 makes precisely such a claim (National Inquiry into the Teaching of Reading, 2005). While the *phonics hypothesis* is defended with examples of research which reach the

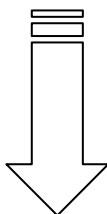
'gold standard', none of the studies cited in any of the reports or panels offers comparable quality evidence that other methods contribute to failure, and yet this has been the uptake of the Report's findings.

The next assumption of NCLB is that the 'right method' to improve literacy outcomes overall, and minority and at risk student performance more specifically, can best be achieved through what we term the *standardised curriculum hypothesis*. This hypothesis states that standardized curriculum programs that script, monitor, and benchmark teachers' everyday teaching can be implemented across schools, communities, and student cohorts to achieve a better and more uniform spread of the optimal 'method' for teaching literacy.

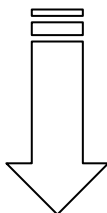
Again binary logics work here. Once the *standardised curriculum hypothesis* is tabled and presented as truth, the inverse of this proposition is taken as fact also. That is that teachers working with professional autonomy and making curriculum, pedagogy and assessment decisions within the local privacy of their own schools and classrooms are the root cause of literacy failure in our schools. This hypothesis seems to have been the basis for the barrage of teacher bashing that we experienced in the media throughout 2007. Part of the complexity of this hypothesis and its implications is that the very nature of the local ecology of the classroom means that it does not answer to the 'lab' results of the 'gold standard' of research in any clear, causal or implicated fashion. The classroom is not a lab.

These two assumptions —one about the efficacy of a phonics approach and another about the efficacy of standardising 'methods' as a way to 'teacher-proof' the curriculum – lead to a particular reform response and limit others. The argument looks something like this:

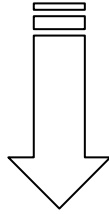
That current teacher methods for teaching literacy are unscientific and flawed (problem);



that government identification and selection of a scientifically verified approach to early literacy training that emphasises phonics (policy),



implemented through an accountability system based on state standardised testing (policy),



will lead to test score gains, particularly of those children from historically underperforming groups (outcome).

So the problem is seen to be teacher failure to implement a scientifically verified method, and the answer to this becomes gaining control of methods used in the classroom – making the classroom a lab – and controlling the behaviour, talk and responses of those within it. There is little recognition of the host of contributing factors identified in ethnographic, case-based, and quantitative literacy research. Factors like home/school transitions and access; the variable impacts of community cultural and linguistic background; the effects of poverty; the increasing incidence of special needs (e.g. Gregory, 2000); the increased diversity of students within our schools; and the impacts of differential school resourcing (OECD, 2005) and internal tracking structures of schools (e.g. Oakes, 1985) as examples. Because this corpus of work is multidisciplinary and does not focus principally on pedagogic method as dependent variable, it was ruled out of the scientific “gold standard” of the national reports, and the findings of this large field of research have consequently not been accounted for in these official debates.

With teachers becoming the ‘problem’, and government control becoming the solution, it is not teachers, students, or communities who act to address the problem and improve outcomes. Instead, the focus is upon the role of government in selecting, sanctioning and implementing the proper ‘method’ to teach and the ‘essential’ content to test. While governments and systems are of course important players in these decisions and should remain so, teachers as professionals need to be key players too.

### **Balancing prescription and professionalism**

So while we wait to see what the approach of our new government in Australia will be to national consistency, we can comment on the approach taken under our last

government as a way to question what the next move should or could be. As described above, the argument to date has proposed that current content and methods are the issue and controlling or standardising what happens in classrooms has become the answer. So the dual policy fix is: fix and mandate new (or old) content (change the “prescription”); enforce this through increased accountability pressure, incentives and disincentives for teachers (change the “professionalism”). This reflects the US NCLB policy approach. Nichols, Glass and Berliner’s (2005) major study explains that increases in “accountability pressure” ratings such as those proposed by this ‘fix’ have not led to improved quality or equity in national testing outcomes in the United States. This is a consideration for Australian policy makers.

Schleicher (2009 in press) refers to this as “uninformed prescription” that is linked to “uninformed professionalism”. Uninformed prescription entails strong centralised accountability without the resources or the opportunities for building strong knowledge-based and evidence-based teacher professionalism. Schleicher stresses the need for an approach that sets the conditions for local teacher professionalism, where teachers are supported to use professional knowledge and evidence to make informed and relevant decisions about teaching, learning and assessment. In other words, “informed prescription” depends upon teachers’ professional capacity to locally interpret, adapt, and adjust curriculum, pedagogy and assessment within contexts, that is it depends on informed professionalism.

Our experience of high quality / high equity systems such as Ontario and Finland, where the bar is set high and the achievement gap is narrow, is that the prescription is enforced by testing which is diagnostic and functional rather than high stakes, and that there is an expectation that teachers will exercise informed and autonomous professionalism. This expectation is supported by strong system messages about standards, equity and outcomes, alignment between pre-service and in-service training, and local evidence-based decision making which informs and works in a reciprocal relationship with evidence-based decision making at a system and State level. High quality / high equity systems tend to strike a balance of prescription and professionalism on the ‘informed’ axis, there is a balance of systemic standard setting and accountability with well-resourced, local school leadership with a strong focus on building teacher capacity for curriculum, pedagogy and assessment.

In contrast, the uninformed prescription model, reinforced by testing for purposes of surveillance and quality control, mandates that teachers reproduce existing, mandated programs and approaches. Its most extreme form is in commodified curriculum packages, “teacher-proof” or “scripted” instruction, where the system

attempts to 'micromanage' teacher/student interaction in the interests of quality assurance and accountability through curriculum prescription, a host of "collateral" effects that include narrowing of the curriculum, teaching to the test, teacher deskilling and attrition, documented test score fraud and manipulation at the state and school level – with no visible sustainable effects at improving equity outcomes (Nichols & Berliner, 2007; Paris, 2005).

### **Lessons for the Australian system**

So as Australia moves to new deliberations on the form and content of a National Curriculum, and place of State curriculum and syllabus documents in our public and private systems what should we learn from seven years of NCLB in the United States?

To begin with, *increased accountability does not necessarily lead to more equitable outcomes for all*. A basic proposition of NCLB is that accountability as standardised testing is an optimal means of encouraging literacy gains especially amongst students from minority and lower-socioeconomic backgrounds. However research has suggested that this proposition is not valid. Amrein and Berliner (2003) found that there were no consistent US state-by-state effects on the National Assessment of Educational Progress (NAEP) test results. In this large scale quantitative comparison of state results, there was no evidence that children's literacy outcomes were improving as a result of new testing regimes of NCLB. In a reanalysis of state test scores, Rosenshine (2003) also found no consistent effects demonstrated in relation to improved outcomes or standards. While there is evidence that supports the claim that grade 8 mathematics results have improved as a result of testing (Braun, 2004; Carnoy & Loeb, 2002), there are also claims that the some results are being inflated by either exclusion of lower achieving students (Amrein & Berliner, 2003) or as a result of rising adolescent drop-out rates (Heubert & Hauser 1999). Added to this, the most recent official report states there has been no improvement in comprehension scores of students' reading (Institute of Educational Sciences, 2008) since the implementation of *Reading First* and the testing measures of NCLB.

The assumption of NCLB is that incentives and sanctions on schools and teachers based on student performance on tests will have a tonic effect on overall standards and achievement. So the assumption is that teachers and students will work harder and produce better results when faced with incentives and discipline (Nichols, Glass & Berliner, 2005). The complex factors that can be shown empirically to mediate student achievement include issues of content and construct validity and test

preparation, the overreliance on single-shot assessment, and the dynamics of spatialized poverty and demographic change. Triangulated by extensive qualitative documentation on the unintended effects of NCLB (Nichols & Berliner, 2006), reanalysis of NAEP data suggests that “the relationship between high-stakes testing and its intended impact on achievement is mixed and inconclusive” (Nichols, Glass, & Berliner, 2005, p. 2).

In fact it is likely that the collateral damage of these policy initiatives in the form of lowering retention rates and an increased achievement gap with service cuts to priority groups will have implications for the United States for many years. Nichols et al. (2005) measured and ranked states according to state-level testing pressure through a system of Accountability Pressure Ratings (APR). This rating system was used to query whether “the pressure of high-stakes testing increases achievement” (p. 3). Simply, the positive link claimed by many states between the introduction of high-stakes testing through NCLB and improved student achievement is tenuous, with no gains being demonstrable on reading achievement in years 4 or 8. There was a negative correlation between increased accountability pressure and retention to senior and college years. That is, systemic increases in high stakes testing have an apparent link with increased drop-out rates.

A second lesson is that *the mandating of content and method as a means to control what happens in the classroom and thus ‘teacher-proof’ the curriculum does not lead to high quality / high equity systems, but rather a system based in uninformed prescription and uninformed professionalism.* This has two themes. The first is the fact that it has been a long standing axiom of the school curriculum that policy does not control, in any absolute sense, the curriculum-in-use. The classroom is not a lab and as such written policy is recontextualised and remediated through numerous iterations before and as it is enacted by teachers and students within classrooms. Second, there is a risk of long term, collateral damage in de-skilling our teacher workforce. Large scale pedagogy and school reform research (see for example Luke, Freebody, Lau, & Gopinathan, 2005; Education Queensland. 2001) has demonstrated that ‘good’ teachers combine, meld and weave (Luke, Cazden, Lin, & Freebody, 2004) pedagogical approach, content and method in ways that have local logic based on contextual knowledge.

Based on the NCLB experiment, we can state that the combination of increased testing, standardised programs, increased accountability and incentives/sanctions for schools, districts and states who do not reach targets has not been a success. There have been extensive local ‘collateral effects’ documented (e.g. Nichols et al, 2005).

These include: test-preparation sessions, school and district-level test administration and test-score manipulation, loss of experienced teachers, inadequate funding for professional development, lack of support for English language learners and students with special needs, increased teacher utilisation of packaged materials without 'scientific' backing, and, significantly, a narrowing of the overall curriculum.

So the combination of increased testing, standardised programs, increased accountability and incentives/sanctions for schools, districts and states who do not reach targets has not been a success. Surely these are telling lessons for policy makers in Australia; telling lessons for teachers as they consider their place in the current Australian debates.

## References

- Allen, J., Altwerger, B., Edelsky, C., Larson, J., Rios-Aguilar, C., Shannon, P., et al. (2007). Taking a stand on NCLB. *Language Arts, 84*(5), 456-464.
- Allington, R. (2002). *Big Brother and the National Reading Curriculum: How ideology trumped evidence*. Portsmouth, NH: Heinemann.
- Alvermann, D. (2002). Effective literacy instruction for adolescents. *Journal of Literacy Research, 34*, 189-208.
- Amrein, A., & Berliner, D. (2003). The effects of high-stakes testing on student motivation and learning. *Educational Leadership, 60*(5), 32-41.
- Braun, H. (2004). Reconsidering the impact of high-stakes testing. *Education Policy Analysis Archives, 12*(1). Retrieved September 11, 2007, from <http://epaa.asu.edu/epaa/v12n1/>
- Calfee, R. (2002, August). *An overview of reading achievement in California*. Paper presented at the Conference of University of California Literacy Educators and Researchers, Berkeley.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis, 24*, 305-331.
- Centre on Education Policy (2003). *From the Capital to the Classroom: Year 3 of the No Child Left Behind Act*. Washington, DC: Centre for Educational Policy.
- Coles, G. (2003). *Reading the Naked Truth: Literacy, Legislation and Lies*. Portsmouth, NH: Heinemann.

- Education Queensland. (2001). *School reform longitudinal study: Final report*. Brisbane, Australia: School of Education, University of Queensland.
- Encisco, P., Katz, L., Kiefer, B., Price-Dennis, D., & Wilson, M. (2007). NCLB day by day... *Language Arts*, 84(5), 408-409.
- Garan, E. (2002). *Resisting Reading Mandates: How to Triumph with the Truth*. Portsmouth, NH: Heinemann.
- Gee, J.P. (2000). The limits of reframing: A response to Professor Snow. *Journal of Literacy Research* 32(1), 121-128.
- Gregory, E. (2000). *City literacies*. London: Routledge.
- Grunwald, M. (2006). The education issue, *Washington Post* (1/10/06), Retrieved 30<sup>th</sup> October, 2007, from <http://umdrive.memphis.edu/lmcgllvr/public/THE%20EDUCATION%20ISSUE%20-%20washingtonpost.com.pdf>
- Heubert, J. P., & Hauser, R. M. (Eds). (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, D.C.: National Academy Press.
- Institute of Educational Sciences (National Centre for Educational Evaluation and Regional Assistance). (2008). *Reading First Impact Study*. Washington, DC: U.S. Department of Education.
- Lee, J. (2006). *Tracking Achievement Gaps and Assessing the Impact of NCLB on the Gaps: An in-depth look into national and state reading and maths outcome trends*. Cambridge, MA: Harvard University Civil Rights Project.
- Luke, A., Cazden, C., Lin, A. & Freebody, P. (2004). *The Singapore classroom coding system*. Technical Report. Singapore: Centre for Research in Pedagogy and Practice.
- Luke, A., Freebody, P., Lau, S., & Gopinathan, S. (2005). Toward research-based innovation and reform: Singapore schooling in transition. *Asia Pacific Journal of Education*, 25, 1, 5-28.
- Luke, A., & Woods, A. (2008). Large-scale policy and adolescence. In L. Christenbury, R. Bomer, & P. Smagorinsky (Eds). *Handbook of Adolescent Literacy Research* New York: Guilford Press.
- Milburn, C. (2008). Literacy chairman sees red on uni 'claptrap'. *The Age* (4-8-08). Retrieved 4<sup>th</sup> August, 2008, from <http://www.theage.com.au/national/literacy-chairman-sees-red-on-uni-claptrap-20080803-3pee.html> on August 4, 2008.

- National Inquiry into the Teaching of Literacy, (2005). *Teaching Reading*. The Commonwealth of Australia: Australia.
- National Reading Panel (2000). *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. Washington, DC: National Institute for Child Health & Human Development.
- Nichols, S., & Berliner, D. (2006). *Collateral damage*. Cambridge, MA: Harvard Education Press.
- Nichols, S., Glass, G., & Berliner, D. (2005). *High stakes testing and student achievement: Problems for the No Child Left Behind Act*. Tempe, AZ: Educational Policy Research Unit.
- Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.
- Organisation for Economic Co-operation and Development. (2005). *School factors relating to quality and equity: Results from PISA 2000*. Paris: Author. Retrieved September 11, 2007, from <http://www.pisa.oecd.org/dataoecd/15/20/34668095.pdf>
- Paris, S. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, 40, 2, pp. 184-202.
- Pennington, J. (2007). Re-viewing NCLB through the figured worlds of policy and teaching: Creating a space for teacher agency and improvisation. *Language Arts*, 84(5), 465-474.
- Rosenshine, B. (2003). High-stakes testing: Another analysis. *Education Policy Analysis Archives*, 11, 4. Retrieved May, 2007 from <http://epaa.asu.edu/epaa/v11/N24/>
- Schleicher, A. (2009 in press). Seeing school systems through the prism of PISA. In - Luke, A., Woods, A., & Weir, K. *Curriculum, syllabus design and equity: A primer and a model* NY: Routledge.
- Snow, Burns & Griffin (1998). *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.
- United States of America. "No Child Left Behind Act of 2001", 20 USC 6301 (2001) & (2002).

Yatvin, J., Weaver, C., & Garan, E. (2003). Reading first: Cautions and recommendations. *Language Arts*, 81, 1, pp. 28-33.