Status: **Academic Book Chapter**

Full Reference:

Ellis, S (2010) 'The use of evidence in language and literacy teaching' in K.Hall, U.Goswami, C.Harrison, S.Ellis and J. Soler (eds) *Interdisciplinary Perspectives on Learning to Read*. London: Routledge Chapter 14 p. 193-207

ISBN: 9780415561242

CHAPTER FOURTEEN

THE USE OF EVIDENCE IN LANGUAGE AND LITERACY TEACHING

SUE ELLIS, *Department of Childhood and Primary Studies, University of Strathclyde*

Literacy education deserves evidence-based decisions. Low literacy costs the British economy between £1.73bn and £2.05bn per year (KPMG 2006) and the social and emotional costs are equally high. Yet we know that children who struggle with literacy can make fast progress when the instructional content and pedagogy closely match their needs. This chapter describes some of the paradigms and problems associated with the use of evidence in language and literacy education with examples from specific interventions and programmes. It raises issues about how literacy teachers are professionalized to attend to evidence, both the evidence in front of them and the research evidence 'out there'. It argues that to support teachers in using evidence effectively, we need to frame the research evidence about effective content, pedagogy and learning in ways that recognize the power and limitations of different evidence paradigms. To ensure that all children make fast progress, we need to appreciate how teachers develop broad and diagnostic understandings of literacy and literacy learning and of how policy and curriculum frameworks impact on the classroom decisions they make.

The problem of evidence-based education

For one group of researchers, solving the problem of evidence–based literacy education is straightforward: 'Use what works... Focus on research-proven programmes and practices' (Slavin 2008). The suggestion is that the 'evidence' issue in education boils down to two problems: firstly, the paucity of properly evaluated programmes and secondly, the problem of ensuring that teachers, school managers and policy makers know how to select for implementation only those programmes that have demonstrated a high impact. (Torgerson & Torgerson 2001; Chalmers, 2003; Slavin 2008; Tymms et al 2008; Chambers 2008). This view of evidence–based literacy education gained impetus in the

USA from that government's *Strategic Plan for Education 2002-2007* in which education research was criticized for lacking rigor and lacking the cumulative studies that generate knowledge about procedures that can be transferred across different classroom contexts, schools and education systems: '…unlike medicine, agriculture and industrial production, the field of education operates largely on the basis of ideology and professional consensus. As such, it is subject to fads and is incapable of the cumulative progress that follows from the application of the scientific method…' (US Department of Education 2002: 59).

The solution proposed in the *US Strategic Plan* was to set out a model for education research and evaluations that would ensure the quality of evidence. Although the initial proposal was for a complex and not necessarily hierarchical model, government grants were made available only to states, districts and schools that used "proven methods" of reading instruction and the federal *No Child Left Behind Act* mandated that districts use "scientifically proven" instructional methods. In practice, the various research designs rapidly became ranked with Randomized Controlled Trials (RCT) at the top, followed by controlled cohort studies, case study series, individual case studies and, at the bottom, professional observation (Eisenhart and Towne 2003).

The Randomized Controlled Trial (RCT) was ranked highest because it is replicable. A well-designed RCT identifies one variable for investigation (for example, the educational programme being implemented) and controls for others that may influence the outcome. It requires random allocation of pupils, strict compliance measures, blind assessment so that assessors do not know which children are in which group and it determines the levels for a significant response independently at the start of the trial. In education, meeting the conditions for a full RCT is problematic - children cannot be randomly allocated to schools and classes, or randomly allocated to follow different programmes within a class – and many education researchers use cluster randomization at the level of class or school rather than the individual pupil (e.g. Tymms et al 2008).

Problem One – Compliance

The rhetoric of the scientific RCT approach is seductive, particularly its focus on sorting out the literacy programmes that are effective from those that are not. It offers a vision of certainty in which teachers and policy makers can get clear-cut advice about what to do based on hard-nosed evidence that it will work. In practice, however, successful RCTs tend to lose their impact when rolled out to wider groups of schools (Datnow et al 2002). One reason for this is that when researchers conduct evaluation studies, the intervention or programme has a high profile and research procedures ensure that compliance measures are met. On rollout, the programme has a lower profile, it must compete with other programmes for curriculum space and teacher attention and it is more likely to be adapted or implemented in ways that compromise its design.

Two Scottish studies of children with severe and persistent language difficulties illustrate these problems (Boyle et al 2007; McCartney, Ellis and Boyle 2009). Like many other countries, Scotland's policy of social inclusion is premised on a socio-cultural model of

learning support. Mainstream schools are seen to offer rich learning environments for social and language development and children with severe and persistent language difficulties are encouraged to attend them. Speech and language therapists (SLTs) advise education staff about how to address the needs of individual pupils but the schools have the legal responsibility for ensuring that pupils' educational needs are met.

The first study, a full RCT on 161 language-impaired children in mainstream schools, showed both that individual and group language intervention could make a significant difference to these children's expressive language. The intervention activities were delivered directly by SLTs or by assistants (SLTAs) who were given a small amount of training and specific language targets by the SLT and activities and advice from a manual developed for the project. Additional support in the form of meetings, phone calls and written communications was also available. Standardized tests (CELF-3 UK, Semel et al 2000) before and immediately after the intervention showed that all intervention groups made significantly more progress than a control group given 'usual therapy' (Boyle et al 2007). It did not matter whether the intervention activities were delivered in a group or individually or by the SLT or the SLTA.

These results seemed to indicate that a positive intervention was possible and the training and the manual developed for this project could help teachers and classroom assistants work with SLTs to support language impaired children in school. It would help the schools meet their legal responsibilities and deliver an intervention of proven impact. Delivery by school staff offered further potential benefits because teachers and classroom assistants have wide opportunities to harness more general classroom activities to meet the children's needs.

A cohort study of 38 children in 19 schools was devised (McCartney, Ellis and Boyle 2009). The children for this study met the same inclusion criteria as for the RCT and it used the same consultancy and training model. The SLT provided specific language targets for each child and helped school staff (class teachers, learning support teachers and classroom assistants) with identify suitable activities and advice from the manual. The teachers and head teachers agreed to implement the activities on the same schedule as the RCT and to log each contact session. The children were pre tested and then tested immediately after the intervention using the same standardized test as the RCT.

However, analysis of the test scores for this cohort intervention showed no significant impact on the children's expressive or receptive language. Analysis of the teachers' logs detailing the sessions children had been given indicated that, whereas the RCT provided 45 contact sessions over 15 weeks (an average of three contacts per week), the cohort study averaged only one or two contacts per week. Teachers generally planned for the agreed three activities a week, but not all were delivered. The logs identified a range of issues that prevented the activities from occurring, including difficulties in planning, in managing time and in accessing and managing the support staff allocated to work with the children.

This cohort study yielded important information about the operational issues that teachers face in schools, information that did not emerge from the RCT. It showed that the involvement of class teachers did not compensate for the lack of dedicated sessions. Also that the common-sense principle whereby school staff assumed that part-delivery of the intervention had some worth (on the basis that 'half a cake is better than no cake') was not well-founded; one or two sessions a week had no significant impact on the children's language and unless schools could ensure three sessions per week, the programme was probably not good use of time.

In terms of evidence-based practice, the RCT showed that, with the right support, children with severe and persistent language difficulties can make progress. The cohort study showed that evidence about how a programme actually operates in the real world is vital. The data highlights the need for a model of intervention that is more complex than simply specifying the content and operational parameters of a programme. An effective intervention model must address several levels of policy implementation within the school, employing planning and record keeping systems that incorporate a range of prompts and checks to ensure that school managers and classroom practitioners prioritize and deliver the intervention activities.

However, designing for compliance in this way risks producing a 'top-heavy' system in which the internal focus is on monitoring to ensure delivery and the external focus is on a package of activities. It is an empirical question whether working in this way would enhance teachers' understanding or their ability to recognize and use evidence diagnostically to better match core teaching content to children's needs. The approach assumes that intervention programmes operate as 'sealed units' in which the context of implementation has no impact and it could position school staff as peripheral, complicating factors to be controlled or circumvented whenever possible.

Problem Two – The Context of Implementation

Intervention programmes do not operate as sealed units and the context of implementation affects their impact even when delivery and compliance are closely monitored. The evidence from the evaluation of Reading Recovery in Northern Ireland shows this (Munn and Ellis 2005).

Reading Recovery was devised in New Zealand by Marie Clay in the 1970s as an intervention to impact on the lowest attaining pupils. Its research base is rooted not in RCT evidence, but in a substantial series of case studies into the mechanisms of learning to read. It is not a programme per se, but a tight framework for teaching in which specially trained teachers withdraw children for individual tuition. They use generic activities and highly structured observation and analysis techniques to provide instruction that is carefully tailored to the individual needs of each child and they coach struggling readers to use knowledge flexibly and develop reading behaviours that are self-sustaining and self-expanding (Clay 1991). Clay always claimed that Reading Recovery works regardless of its context of implementation and most studies have focused on it as a

stand-alone intervention (see for example, Brooks 2002; Gardner et al 1998; Shanahan and Barr 1995).

The operational parameters of Reading Recovery are designed to ensure high compliance and fidelity and to withstand the pressure to adapt exerted by the school contexts. Thus, Reading Recovery teachers are supported and monitored by Reading Recovery tutors who are in turn supported by trainers from the Reading Recovery National Network. These trainers are not employees of the school district and are accountable only to the International Reading Recovery Network. They operate a highly effective 're-direction system' to ensure that Reading Recovery is delivered as specified and resists adaptation and colonization by the various 'host' systems in which it operates.

Our evaluation study in Northern Ireland showed that even these strict compliance measures could not completely mitigate the effect of school context on the efficacy of Reading Recovery interventions (Munn and Ellis ibid). A wide range of schools had taken-on Reading Recovery and, although some had actively sought involvement, many were involved because they had a long 'tail' of underachievement or were linked to a secondary school with such a 'tail'. These schools wanted to address underachievement but not necessarily to change their core literacy curriculum, Reading Recovery was therefore operating in a wide range of literacy contexts some of which were highly attuned to the intervention's approach, some were indifferent to it and some were rather hostile. The evaluation study collected quantitative data from 114 Reading Recovery teachers on the nature and scope of their involvement with classroom teachers and on literacy practices in their school and data on 1552 children who had been through the system. This data included discontinuation status, entry and exit book levels, the number of lessons each child had received and the number of weeks spent in Reading Recovery.

Although Reading Recovery was effective in all contexts, it worked more quickly (on average requiring 10 fewer lessons) when classroom literacy teaching practices dovetailed closely with Reading Recovery methods. This made Reading Recovery more cost-effective and efficient in some contexts than others (Munn & Ellis 2001; 2005).

Problem Three: 'horses for courses'

There are further problems associated with prioritizing RCT knowledge about 'what works'. RCTs are excellent for building and interrogating theoretical knowledge and they provide robust evidence of a programme's impact once the variables associated with specific contexts or populations are ironed out. However, evidence of impact on a general population, whilst interesting, may be less useful to practitioners than knowing its impact on a particular type of pupil cohort. Schools are not perfectly located with randomized catchment areas but tend to represent skewed populations and highly localized implementation contexts. Evidence about the impact on particular pupil cohorts is crucial to understanding whether a programme will be effective.

Gathering this information requires a different type of trial, a controlled cohort study (Sackett et al 2000). Here, the characteristics of the participants or context are clearly

defined so that progress can be compared, either to cohorts and contexts with different characteristics, or to similar cohorts following different interventions or programmes. Robust cohort studies identify, document and measure any characteristics of the cohort or the context of implementation that could affect the outcome. Despite their ability to take account of different cohorts and contexts, they are considered less reliable than RCTs in the US ranking because unacknowledged factors associated with the non-randomized sample might generate misleading information.

However, a large body of evidence indicates with remarkably consistency that socio-economic status, gender and race are all closely related to how quickly and easily children learn to read and write. We know from ethnographic (e.g. Gregory 2008; Moss, 2007; Lareau 2003) and survey (e.g. McCoach et al 2006; D'Angiulli et al 2004; Topping et al 2003) research that children begin school with different amounts of literacy experience, different knowledge and skills and that they respond differently to the literacy education that schools offer. For example, children living in less-advantaged socio-economic circumstances are likely to begin school with a poorer phonological awareness and alphabetic knowledge, to have had less experience of books and digital technologies, to have poorer access to books, different understandings about the purposes and uses of literacy, a more limited knowledge of the world, poorer oral language and vocabulary skills, a less secure grasp of narrative and are more likely to have to have mothers and care-givers who are less-well educated and poorly positioned to support them in learning to read and write. (Zill and Resnick 2006). Also, Stanovich's work on 'Matthew effects' shows that a poor start to the most 'visible' aspect of reading - decoding words - can have lasting and compounding effects throughout a child's school career (Stanovich 1986). This would suggest the value of research paradigms that acknowledge different understandings and experiences.

The implications of children's different skill levels for the literacy curriculum are the starting-point for the work of Carol Connor and her colleagues (Connor et al 2004; Connor et al 2005). In a series of cohort studies, they investigated the optimum balance between phonics and reading-for–meaning activities, and also between child-directed or teacher-directed learning opportunities. The optimum 'literacy learning mix' depended on children's pre-existing letter-word reading and vocabulary skills. Those starting school with above-average vocabulary scores made greatest gains when they spent more time engaged in meaning-focused and self-directed activities such as independent reading; children starting school with below-average vocabulary scores made greater gains when given teacher-directed phonics activities at the start and increased self-directed, reading-for-meaning, activities as the year progressed. However, the picture is complex: children requiring less phonics instruction in Grade 1, benefited from more teacher-directed phonics work in Grade 2. (Connor et al 2007a). To take the guesswork out of teaching phonics, the team developed algorithms that calculate the optimum amount of time that children with different literacy and language skill profiles should spend on each kind of instruction and learning activity to maximize progress.

This is an important shift in the view of evidence-based teaching. Rather than producing evidence about the 'best programme' regardless of the pupil cohort, Connor offers

teachers access to complex data to help them plan the most effective ways to implement their existing reading and phonics curriculum. It is a different vision of the teacher's role in driving literacy education. It does not focus on compliance to a programme designed, trialled and chosen by others, but uses technological advances in information handling to harness research evidence as part of the teacher's planning process. This positions teachers' everyday use of evidence in a positive and symbiotic relationship to the evidence produced by researchers and creates the possibility of pulling the 'evidence in the classroom' and the 'research evidence out there' into greater alignment. This work illustrates the power of a series of cohort studies to map the terrain and develop a complex model in which teachers and teaching are seen as part of the solution rather than a complicating factor in the problem.

How teachers are positioned in relation to evidence-based teaching matters. RCT programme evaluations inevitably tend to see any contextual analysis, including teachers' professional judgment, as a threat to compliance procedures. They offer little incentive to take a diagnostic view of learning and no clear mechanisms for adapting programmes in the light of such evidence. Although some researchers express irritation with school cultures which, they feel, privilege professional judgment and encourage teachers to cling to ineffective practices, teachers are equally irritated by initiatives that, they feel, neither expect nor allow them to respond to the children in front of them but promote rigid curriculum frameworks with narrow models of teaching or literacy.

The evidence promoting programme adaptation

Yet there are strong imperatives for teachers to attend to the ecology of the whole curriculum rather than implementing discrete programmes. The language and literacy curriculum has to meet wide (and frequently changing) sets of goals. For example, children must learn to read, but they may also need to learn to work together or to use their reading skills in specifically creative, critical or entrepreneurial ways. A series of atomistic programmes is rarely an effective way to deliver diverse policy outcomes, and programmes are often adapted to take account of these wider goals.

The evidence on pupil engagement also highlights the importance of adapting programmes to link curricular areas and to contextualize and present activities in ways that pupils find relevant and interesting. Reading engagement matters because high engagement mitigates the worst effects of socio-economic status on reading attainment (Topping et al 2003). Teachers, who can do nothing to change the socio-economic factors that impinge on children's lives, can limit its impact on literacy attainment by actively promoting reading engagement. Moreover, avid readers develop richer vocabularies, better verbal reasoning skills, and wider general knowledge, which drives-up attainment across the curriculum (Cunningham and Stanovich 1998).

In a meta-analysis of the research evidence on reading engagement, Guthrie and Humenick (2004) show that, to produce engaged readers, the literacy curriculum must promote *curricular coherence* (so that pupils see the links between subjects and tasks),

strategy teaching (so that pupils know how to apply the skills and knowledge gained in one task on others), *intrinsic purposes* (so that tasks that are meaningful and have outcomes that the pupils believe are worthwhile), *choice* (so that pupils can influence the learning tasks, their timing, their sequence their outcomes and materials) and *collaboration* (so that pupils' learning is social, which aids their persistence and perseverance with challenging tasks). These elements are not about individual programmes, but are about how teachers create coherent links between programmes, how they link literacy teaching to other areas of the curriculum, and how they contextualize literacy tasks to make them interesting and intrinsically motivating for pupils. This evidence implies that the impact of a programme is not solely affected by its content and design but by how it is adapted to complement other programmes and to dovetail with the core concerns and interests of the pupils. It sits in almost direct opposition to the compliance-measures required by RCTs. The challenge is to employ programmes in ways that respect the ecology of the curriculum, the policy goals of the school, the evidence from research and the learning needs of the pupils.

Understanding how teachers use evidence

Perhaps the key issue is to develop a better understanding of how teachers use the evidence of research and of pupil performance in this complex process, and of the ways that staff development, school systems and curriculum frameworks can prompt better use of this evidence.

Although modern schools have more data about pupil performance than ever before, it does not always make a positive difference to what teachers actually do. Some research indicates that this is due to factors such as lack of time or difficulties in accessing the information in a usable form (Wayman and Stringfield 2006). A number of studies indicate that teachers make fairly accurate judgments about language and literacy attainment (e.g. Williams 2006) but less accurate diagnostic judgments (Cabell et al 2009). The accuracy may depend on the specific aspect of language and literacy that is being considered. Nation and Angell (2006) indicate that teachers overlook reading comprehension difficulties, especially when the pupil has strong decoding skills, but that they do notice decoding difficulties. Screening procedures that could help teachers identify pupils with problems often lack the sensitivity to accurately identify pupils with problems when used alone and, although using two screening tests can increase their sensitivity, their specificity (i.e. their accuracy in *not* identifying those who do not have problems) and their predictive power are too poor to form a reliable basis for intervention (Fletcher et al 2001).

The few studies that have been done on effective evidence-use by teachers highlight the need for schools and local authorities to have a deliberate agenda for getting teachers to think about the formal data that schools keep on pupils and to agree common understandings of how it can be used in relation to their professional judgment and diagnostic analyses (Zhao & Frank, 2003; Wayman 2005). They suggest that teachers are enthusiastic about evidence when it provides useful information for their classroom practice and that establishing a rationale for the use of particular types of evidence,

modeling such use, and structuring time for teachers to learn about how to use it can be a helpful way forward. Any debate will be complicated by the cross-disciplinary nature of literacy learning research. Psychology, linguistics, sociology, literature and philosophy all directly and indirectly inform classroom pedagogy and each values different types of evidence and different discourses around that evidence.

We also know that teachers differ in their ability to plan and organize their work; in an RCT of their phonics algorithm software, Connor et al found that some teachers struggled to manage the group planning and organization elements recommended by the software. Teachers also differed in how frequently they accessed the data to plan their lesson content and delivery. Those who were most diligent and organized in using the software had classes with greater reading gains (Connor et al 2007b).

Pedagogical expertise apart, ethnographic studies indicate that teachers have different views about what learning to be literate involves, what constitutes high-quality evidence and what is appropriate use of that evidence. Some teachers focus on evidence of the child's understanding and skills but others focus on what it tells them about the learning processes, pupil engagement, the suitability of course content or the impact of their teaching. These differences are found to exist in the classroom and at every level of policy implementation in the education system (Coburn and Talbert 2006).

Coburn (2001; 2003) and Stein and Coburn (2008) show that some staff development approaches prompt more effective use of evidence about pupil learning than others. The most powerful specify 'big ideas' about teaching content and provide brief tutorials where necessary, but importantly clarify the purpose of the learning activities. Less effective strategies focus on instructional routes through programme material and tend to generate discussions about pupil throughput or the programme's management and organization. Effective strategies produced conversations that focused on teaching and learning and prompted teachers to discuss teaching in the context of specific pupils, their responses to lessons, their understanding and attitudes and the evidence of their learning. This led to new understandings of the teaching content, new insights into the evidence of pupils' learning and a better grasp of the teaching issues. (Stein and Coburn 2008).

National policies and evidence-based education

Given this evidence and the evidence of diverse pupil populations, the wisdom of making blanket recommendations about the 'best' teaching programmes and pedagogies is questionable. Yet in countries with centralized literacy curricula, this is exactly how literacy policy often operates. A further complication comes because evidence is often understood and used differently by politicians, policy makers and practitioners. For example, the discussion of phonics teaching in England, Australia and the US recently focused on a rather polarized debate about the relative efficacy of synthetic or analytic phonics teaching. Two meta-analyses of the research yielded no clear advantage for either approach, simply that systematic phonics programmes were all more effective than non-systematic ones (NICHD 2000; Torgerson et al 2006) and the data is complex and

needs careful interpretation (Wyse and Goswami 2008). Despite this, a policy review in England cited successful initiatives in, amongst other places, Clackmannanshire, Scotland to recommend that all schools adopt discrete, systematic synthetic phonics programmes, (DfES 2006). The evidence is, at best, contradictory; whilst psychology tests showed an average three-and-a-half year gain in decoding words, national reading test attainment was disappointingly average in the two largest and most advantaged schools (which accounted for over half the pupils involved in the study) and there was marked variation amongst the other schools involved, with some doing exceptionally well but others rather poorly (Ellis 2007).

Moss and Huxford observe, 'Phonics in the policy context is not the same as phonics in the research context or phonics as a focus for a political campaign' (Moss and Huxford 2007: 74). In England, phonics was a lever for opposition politicians, pressure groups such as the *Reading Reform Foundation*, some phonics researchers and sections of the media to challenge the government's *National Literacy Strategy*. Attainment in early literacy had leveled out. The teaching-and-learning solution to this might require detailed and complex conversations about the characteristics and trajectories of particular pupil cohorts, the intricate nature of the early reading curriculum and how phonological development interacts with comprehension, phonics and wider issues of reading engagement. However, the political and policy imperatives required the exact opposite: a quick and clear position statement and an unambiguous plan of action to prevent further political fallout. Thus, new curriculum advice was issued for schools and the Prime Minister's Strategy Unit suggested 'The best response may be for the government to take a top-down approach and require the adoption of best practice' (PMSU 2006: 58). The government established a committee to establish which commercial phonics schemes met the new curriculum requirements (Brooks 2006) and requested details of the phonics content of university initial teacher education courses.

Such policy actions form the backdrop for teachers' work in schools, and set the tone and agenda for educational debate in the media. They help define how teachers and everyone involved in the education system thinks about literacy teaching. They do not focus teachers on the evidence in front of them or on the research evidence out there but exert considerable pressure on them to deliver programmes and comply with the latest government policy. It is a rather depressing thought that centralized curriculum decisions will *always* be under wider policy and political pressure to deliver clear, definitive answers and, consequently, will always distract the teaching profession from more nuanced and complex debates about teaching and learning, and about how to use evidence more effectively to teach literacy in particular contexts with particular children and teachers.

Conclusion

Issues undoubtedly exist surrounding teachers' use of evidence. Researchers and those who fund and use education research have a responsibility to ensure that designs that generate contextual information are not undersold. Policymakers, managers and teachers themselves, have a responsibility to promote discussions that deepen diagnostic

understandings and prompt effective action. To make real progress in these tasks however, we need wider discussion of how teachers are professionalized to attend to the research evidence and to the evidence in front of them, and a better understanding of what impacts on teachers' use of evidence and of how wider policy and curriculum structures support or undermine this process.

References

Brooks, G. (2002) 'What works for children with literacy difficulties'? The effectiveness of intervention schemes, *Research Report No. 380, Department for Education and Skills. Norwich: HMSO.*

Brooks, G. (2006) Personal communication during discussion at ESRC seminar: Pedagogy and Curriculum in Early Reading- Multidisciplinary perspectives, *Nov. 2006, The Open University: Milton Keynes.*

Boyle, J., McCartney, E., Forbes J., O'Hare, A. (2007) 'A randomized controlled trial and economic evaluation of direct versus indirect and individual versus groups modes of speech and language therapy for children with primary language impairment', *Health Technology Assessment*, 11, 25: 1-158.

Cabell, S.Q., Justice, L.M., Zucker, T.A. & Kilday, C.R. (2009) 'Validity of teacher report for assessing the emergent literacy skills of at-risk preschoolers', Language Speech and Hearing Services in Schools, *40, 2: 161-173.*

Gardner, J. Sutherland, A & Meenan-Strain, C. (1998) Reading Recovery in Northern Ireland: the first two years, *Belfast: The Blackstaff Press.*

Chalmers, I. (2003) 'Trying to do more good than harm in policy and practice: The role of rigorous, transparent, up-to-date, replicable evaluations', *Annals of the American Academy of Political and Social Science*, 589: 22-40.

Chambers, B. (2008) 'A response to Tymms, Merrell & Coe' *The Psychology of Education Review*, 32, 2, September 2008: 9.

Clay, M. (1991) Becoming Literate: the construction of inner control, *Aukland, NZ: Heinemann.*

Coburn, C (2001) 'Collective sense-making about reading: how teachers mediate reading policy in their professional communities', *Educational Evaluation and Policy Analysis*, 23, *2: 45-70.*

Coburn, C. (2003) 'Rethinking Scale: Moving Beyond Numbers to Deep and Lasting Change', *Educational Researcher*, 32, 6: 3-12

Coburn, C. E. & Talbert, J. E. (2006). Conceptions of evidence-based practice in school districts: Mapping the terrain. American Journal of Education, 112(4), 469-495.

Connor, C. M., Morrison, F. J., & Katch, E. L. (2004) 'Beyond the reading wars: The effect of classroom instruction by child interactions on early reading', *Scientific Studies of Reading*, 8, *4*: 305-336.

Connor, C. M., Son, S., Hindman, A., & Morrison, F. J. (2005) 'Teacher qualifications, classroom practices, family characteristics and preschool experience: Complex effects on first graders' vocabulary and early reading outcomes', *Journal of School Psychology, 43*: 343-375.

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007b) 'The Early Years: Algorithm-guided individualized reading instruction', *Science, 315,* 5811: 464-465.

Connor, C. M., Morrison, F. J., & Underwood, P. (2007a) 'A second chance in second grade? The independent and cumulative impact of first and second grade reading instruction and students' letter-word reading skill growth', *Scientific Studies of Reading,* 11, 3: 199-223.

Cunningham, A.E. & Stanovich, K.E. (1998) 'What Reading does for the mind', *American Educator 22,* 1&2: 8-15.

D'Angiulli, A., Siegel, L. S., & Hertzman, C. (2004) 'Schooling, socioeconomic context, and literacy development', *Educational Psychology*, 24: 867-883.

Datnow, A., Hubbard, L and Mehan, H. (2002) *Extending Educational Reform: From One School to Many*, New York: Routledge Falmer.

DfES (2006) *Independent Review of the Teaching of Early Reading: Final Report*. (The Rose Review), London: DfES.

Eisenhart, M. & Towne, L. (2003) 'Contestation and Change in National Policy on "Scientifically Based" Education Research', *Educational Researcher,* 32, 7: 31-38 .

Ellis, S. (2007) Policy and research: Lessons from the Clackmannanshire Synthetic Phonics Initiative Journal of Early Childhood Literacy. 7 (3): 281-297

Fletcher, J., Tannock, R. & Bishop, D.V.M. (2001) 'Utility of Brief Teacher Rating Scales to Identify Children with Educational Problems: Experience with an Australian Sample', *Australian Journal of Psycholog,* 53, 2: 63-71.

Gregory, E (2008) *Learning to read in a new language: Making sense of words and worlds*, London: Sage.

Guthrie, J.T. & Humenick (2004) 'Motivating students to read: Evidence for classroom practices that increase reading motivation', in P. McCardle and V. Chhabra (eds) *The Voice of Evidence in Reading Research* New York: Erlbaum. Chapter 14: 29-355.

International Reading Association (IRA) (2005) 'News from Reading Today Daily: Why Synthetic Phonics was a Natural Winner', *Reading Today Daily Online.* [Online] Available at: http://blog.reading.org/archives/000794.html Accessed March 15th 2007

KPMG Foundation (2006) *The Long Term Costs Of Literacy Difficulties*. December 2000 [Online] Available at http://kpmg.co.uk/about/foundation/index.cfm Accessed April 30th 2009.

Lareau, A. (2003) *Unequal childhoods,* Berkeley: University of California Press.

McCartney, E., Ellis, S. & Boyle, J (2009) 'The mainstream primary classroom as a language-learning environment for children with severe and persistent language impairment – implications of recent language intervention research', *Journal of Research in Special Educational Needs,* 9, 2: 80-90.

McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006) 'Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school', *Journal of Educational Psychology*, 98, 1: 14-28.

Moss, G. & Huxford, L. (2007) 'Exploring Literacy Policy-Making From the Inside Out', in Lesley Saunders (ed) *Exploring the Relationship between Educational Research and Education Policy-Making*. London: Routledge Falmer.

Moss, G (2007) *Literacy and Gender,* Abingdon: Routledge.

MRC (Medical Research Council) (2000) *A framework for development and evaluation*

of RCTs for complex interventions to improve health, London: MRC.

Munn, P. & Ellis, S. (2005) 'Interactions Between School Systems and Reading Recovery Programmes: Evidence From Northern Ireland', *The Curriculum Journal*, 16, 3: 341-362.

Nation, K. & Angell, P. (2006) 'Learning to read and learning to comprehend', *London Review of Education*, 4, 1: 77-87.

NICHD (National Institute of Child Health and Human Development) (2000) *Report of the National Reading Panel. Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. NIH Publication Number 00-4754, Washington DC: US Government Printing Office.

Prime Minister's Strategy Unit (PMSU) (2006) *The UK Government's approach to public sector reform: A discussion paper*, London: Cabinet Office.

Sackett, D. L., Straus, S. E., Scott Richardson, W., Rosenberg, W & Haynes,R.B. (2000) *Evidence-based medicine*, Edinburgh: Churchill Livingstone.

Shanahan, T. & Barr, R. (1995) '*Reading Recovery: An independent evaluation of the effects of an early instructional intervention for at-risk learners'*, Reading Research Quarterly, *20, 4: 958-995.*

Slavin, R.E. (2008) 'What works? Issues in synthesizing educational programme evaluations', *Educational Researcher*, 37, 1: 5–14.

Stanovich, KE (1986) 'Matthew Effects in Reading: Some consequences of individual differences in the acquisition of literacy', *Reading Research Quarterly*, 21, 4: 360-407.

Stein, M. & Coburn, C (2008) 'Architectures for learning: a comparative analysis of two urban school districts', *American Journal of Education*, 114, 4: 583-626.

Topping,K.,Valtin, R., Roller, C., Brozo, W. & Dionisio, ML (2003) *Policy and Practice Implications of the Program for International Student Assessment (PISA) 2000: Report of the International Reading Association PISA Task Force*, Newark, NJ: IRA.

Torgerson, C.J., Brooks, G. & Hall, J. (2006) 'A Systematic Review of the Research Literature on the Use of Phonics in the Teaching of Reading and Spelling*', DfES Research Report 711*, London: DfES.

Torgerson, C.J. & Torgerson, D.J. (2001) 'The need for randomized controlled trials in educational research', *British Journal of Educational Studies*, 49, 3: 316–329.

Tymms, P., Merrell, C. & Coe, R. (2008) 'Educational Policies and Randomized Controlled Trials', *The Psychology of Education Review*, 32, 2: 3-8.

US Department of Education (2002). Strategic goal four: Transform education into an evidence-based field. Washington, DC: US Department of Education. Retrieved from http://www.ed.gov/about/reports/strat/plan2002-07/plan.pdf, 28th October, 2009

Wayman, J. (2005) 'Involving teachers in data-driven decision making: using computer data systems to support teacher inquiry and reflection', *Journal of Education for Students Placed at Risk*, 10, 3: 295-308.

Wayman, J.C. and Stringfield, S. (2006) 'Data Use for School Improvement: School Practices and Research Perspectives.' American Journal of Education, 112, 4: 463-8

Williams, C. (2006) 'Teacher judgments of the language skills of children in the early years of schooling', *Child Language Teaching and Therapy*, 22, 2: 135-154.

Wyse, D. and Goswami, U. (2008) 'Synthetic phonics and the teaching of reading', *British Educational Research Journal*, 34, 6: 691-710.

Zhao, Y., & Frank, K. A. (2003). Factors affecting technology users in schools: An ecological perspective. American Educational Research Journal, 40: 807–840.

Zill, N & Resnick,G (2006) 'Emergent literacy of low-income children in Head start', in David K Dickinson and Susan B. Neuman (eds) *Handbook of Early Literacy Research Vol 2,* New York: Guildford Press. Chapter 25: 347-374.