



RESEARCH PAPER

Improving secondary students' academic achievement through a focus on reform reliability

**Four- and nine-year findings from the
High Reliability Schools project**

Sam Stringfield
University of Louisville, USA

David Reynolds
School of Education, University of Plymouth, United Kingdom

Eugene C Schaffer
Department of Education, University of Maryland –
Baltimore County, USA

Welcome to CfBT Education Trust



CfBT Education Trust is a leading charity providing education services for public benefit in the UK and internationally. Established 40 years ago, CfBT Education Trust now has an annual turnover exceeding £100 million and employs more than 2,000 staff worldwide who support educational reform, teach, advise, research and train.

Since we were founded, we have worked in more than 40 countries around the world. Our work involves teacher and leadership training, curriculum design and school improvement services. The majority of staff provide services direct to learners in schools or through projects for excluded pupils, in young offender institutions and in advice and guidance for young people.

We have worked successfully to implement reform programmes for governments throughout the world. Current examples

include the UK Department for Children, Schools and Families (DCSF) Programme for Gifted and Talented Education and a nationwide teacher training programme for the Malaysian Ministry of Education.

Other government clients include the Brunei Ministry of Education, the Abu Dhabi Education Council, aid donors such as the European Union (EU), the Department for International Development (DfID), the World Bank, national agencies such as the Office for Standards in Education (Ofsted), and local authorities.

Surpluses generated by our operations are reinvested in educational research and development. Our new research programme – Evidence for Education – will improve educational practice on the ground and widen access to research in the UK and overseas.

Visit www.cfbt.com for more information.

The views and opinions expressed in this publication are those of the authors and do not necessarily represent the views of CfBT Education Trust.

© CfBT copyright September 2008

All rights reserved

A full version of this report is available on www.cfbt.com

Contents

Introduction	5
High reliability organisations	8
The Welsh context	10
The evolution of the HRS programme in Wales	11
Methods	14
Results	16
Reflective themes	19
Discussion	21
References	23

Acknowledgements

The authors wish to express their deep respect for and gratitude to the leaders and teachers of the Welsh HRS schools,

both for their implementation efforts and their willingness to participate in follow-up interviews.

About the authors

David Reynolds is Professor of Education at Plymouth University and has published widely on what makes good schools, and how to make schools good. He designed the Numeracy Strategy, has been on the Boards of BECTA and the TDA/TTA and was a part-time adviser to the (then) DfES from 1999 to 2006. He was awarded the CBE for services to education in 2002. He has authored 20 books and over 250 academic chapters and papers.

Sam Stringfield is the Academic Director of the Nystrand Center of Excellence in Education, a Distinguished University Scholar, and a Professor in the Departments of Teaching and Learning and Leadership, Foundations, and Human Resource Education at the University of Louisville. He was formerly a Principal Research Scientist at the Johns Hopkins University Center for Social Organization of Schools (CSOS). He served as co-director of the Systemic Supports for School Reform section of the Center for Research on the Education of Students Placed At Risk (CRESPAR). Stringfield is also co-director of the Program on Integrated Reform at the University of California at Santa Cruz' Center for Research on Education,

Diversity and Excellence (CREDE). He is founding editor of the *Journal of Education for Students Placed At Risk (JESPAR)*, and a member of the City of Baltimore's New Board of School Commissioners.

Eugene C. Schaffer received his doctorate from Temple University and served as member of the faculties at Valparaiso University, The University of North Carolina Charlotte, National Kaohsiung Normal University (Taiwan), and Tokyo University. Currently, he is Professor of Education at the University of Maryland Baltimore County. He has been Chair of the Department of Education at the University of Maryland Baltimore County and University of North Carolina Charlotte (UNCC). Dr. Schaffer teaches course work on the Study of Teaching and Teacher Leadership. His current research includes an international study of classroom behavior in third world schools, a study of home-school relationships, and an experimental study of at-risk schools. His research interests include classroom interaction, school effectiveness research and the study of students placed at-risk.

Abstract

The authors describe a reform effort in which characteristics derived from High Reliability Organisation research are used to shape whole school reform. Longitudinal analyses of outcome data from 12 Welsh secondary schools indicate that four years after the effort was initiated results at the sites were strongly positive. Additional quantitative and qualitative

data, gathered five years after the end of the intervention, indicate that the majority of the schools continued using the high reliability principles and continued making strong progress. Results are discussed in terms of the original High Reliability Schools model, sustainability, and systemic effects.

Introduction

“ One obvious strand of research in educational effectiveness concerns efforts to chart the effects of diverse school improvement efforts. ”

The first challenge of change is to ensure that it's desirable and the second challenge is to make it doable, then the biggest challenge of all is to make it durable and sustainable.

(Hargreaves & Fink, 2006, p. 2)

The last 30 years have been a time of great ferment in public education in all developed nations. Whether an observer begins in Great Britain, Taiwan, the Netherlands, or the US, the calls for school reform have been loud and continuous. In the educational environment, school improvement discussions, educational improvement 'programmes' and 'whole-school reform designs' are emerging within individual countries and travelling internationally with unprecedented frequencies and speed.

One obvious strand of research in educational effectiveness concerns efforts to chart the effects of diverse school improvement efforts. Over a 60-year period, a reasonably stable body of research has evolved. From the Eight Year Study (Aikin, 1942), the Follow-Through Planned Variation Study (Stallings & Kaskowitz, 1974), the RAND Change Agent Study (Berman & McLaughlin, 1977; McLaughlin, 1990), the DESS studies (Crandall, Loucks-Horsley, Bauchner, Schmidt, Eiseman, & Cox, 1982), and the Special Strategies Studies (Stringfield, *et al.*, 1997), to the New American Schools initiatives (Ross *et al.*, 2001) and studies of specific reforms such as Success For All/Roots and Wings (Slavin, *et al.*, 1996; Borman, Slavin, Cheung, Chamberlain, Madden, & Chambers (2005) and Core Knowledge (Datnow, Borman, & Stringfield, 2000), the strength of the research base has been slowly growing.

Six of the conclusions that can be drawn from this half-century-plus of large-scale studies of school reform follow. A first is that most, but not all, of the reforms can point retrospectively to one or more schools that have participated in the particular reform and improved some combination of faculty attitudes, student deportment, and student attendance. A few

reforms can point to positive results from several multi-site studies of achievement outcomes (Borman, *et al.*, 2002; Reynolds, Hopkins, Potter, & Chapman, 2001; What Works Clearinghouse, 2007).

A second is that all of the reforms that have 'scaled up' to significant numbers of schools, have examples of schools in which the reform has had no measurable effect and has been discontinued. In education, as in law and medicine, there are very few absolute truths and simple cures. While many reforms may contain some valid ideas at their core, to date, their implementation strategies have been less than reliable. In educational reform, as in research, reliability sets the upper boundary of validity.

A third general finding has been that, over diverse reform efforts and contexts, reforms have been substantially more likely to produce measured results if they focused on primary schools. This is not to suggest that all primary-based reforms have proven equally effective. They have not. However, secondary school reforms have consistently found the achievement of measurable, long-term results challenging (see Fullan, 2001; Cuban, 2001). Summarising a multi-year study of a well funded and highly publicised reform effort in Chicago (USA), Sebring, Bryk, and Easton (1995) reported consistent data indicating lower educator involvement in reform efforts and lower student achievement or achievement gain. Darling-Hammond, Hightower, Husbands, LaFlors, Young, and Christopher (2006) conclude, 'High schools have presented a perennial challenge to school reform efforts.' (p. 169). A recent review of British reform efforts makes clear that difficulty in obtaining measurable, multi-school student results in

secondary education has hardly been limited to the US (Reynolds, Hopkins, Potter, & Chapman, 2001). Earl, Torrance, and Southern (2006) conclude that:

*There are no examples anywhere of successful **whole district high school reform**. There are a few high schools, here and there, that have improved significantly, but none as a group.*

A fourth finding from previous research is that in virtually every case involving on-site observations of the reform implementation efforts, success has been greatest in schools where the design team and the local educators worked together to create the most efficacious interaction of the local realities with the reform design. Berman and McLaughlin (1977) referred to this as 'mutual adaptation'. More recently, others have described the process in a more active, engaging voice as the 'co-construction' of school reform (Datnow, Hubbard & Mehan, 1998; Datnow & Stringfield, 2000). The important point is that there have been virtually no sites described in any of the available studies in which a local school literally and uncritically adopted a reform *in toto*. Rather, in study after study, teachers and administrators at sites obtaining the greatest multi-year effects actively engage ideas and practices, and eventually are full partners in the creation of an improved school. Just as there never were 'teacher-proof texts', there are no 'school-proof reforms'.

Fifth, the ability of a school to engage with a design team in co-constructing school reform and thereby achieving reasonably strong implementation of most components within a specific context implies a school that either possessed a substantially healthy or 'effective' school culture prior to engaging in a specific reform, or developed such a local environment as a result of the reform. Either possibility suggests the relevance of the school effects research base. Following initially promising work by Edmonds (1979) and Rutter, *et al.* (1979), a steady stream of school effects research has developed. *The International Handbook of School Effectiveness Research* (Teddle & Reynolds, 2000) reviewed more than 1,400 articles and books in the field. The majority of school reform projects now assume the importance of such basic school effects

dimensions as 'instructional leadership', 'positive school culture/climate', and 'efficient use of school and classroom time'.

Sixth, research on the sustainability of various educational reforms and their outcomes, let alone continuing improvement on diverse measures post-reform, is extremely rare. Those studies that exist can be divided into two types: historical/'looking back'; and proactive. Reviewing the previous decades' reform efforts, Tyack and Cuban (1996) and Cuban and Usdan (2002) concluded that most reforms do not last, and those that do have strongly tended to involve structural additions to the system that do not threaten previously existing structures (e.g., the addition of near-universal kindergarten). Hargreaves and Fink (2006) examined eight secondary schools that had attempted reforms at least two decades prior to the authors' data gathering. Summarising their analyses of the various schools' whole-school innovation efforts, the authors conclude that, 'The overall evidence is not uplifting. The vast majority of reform efforts and change initiatives – even the most promising ones – are unsustainable.' (Hargreaves & Fink, p 252)

Proactive studies have the great advantage of having had direct experience with the reforms from the beginning, so that they need not rely on participants' reflections from five to 25 years past. They have the disadvantage of almost always being of shorter duration. Among the proactive studies that have attempted to examine both short- and moderate-term effects, two are particularly relevant here. Schaffer, Nesselrodt, and Stringfield (1997) examined three-year data from 25 schools, each of which was implementing one of ten diverse 'promising programmes'. The authors found that various combinations of ten factors lead to discontinuation of promising reforms. These included issues with finances, leadership, declining commitment among participants, public and parent perceptions, staffing changes, lack of professional development, curriculum misalignment with larger organisational goals, political changes, racial politics, facility issues, and practical problems with communication and/or scheduling. Datnow (2005) reported on

“ It would be wise to choose a reform that can help the school improve on state and district measures of accountability. ”

a study that followed 13 schools involved in various 'promising programmes'. She reported that several of the schools never achieved moderate levels of implementation and that only five continued implementation at moderate or higher levels after four years. She noted that the sustainability of reform appeared to be a multi-layered construct, with issues at the level of the state and the Local Authority (LA) directly impacting individual schools' decisions whether to continue any given reform. She concluded as follows:

1. Schools not committed to sustaining chosen reforms should probably not begin them.
2. Would-be reformers should not underestimate the substantial long-term efforts required to achieve sustainability.
3. To be sustainable, a reform should place few long-term demands for resources on the larger system.
4. Reforms that last help educators meet new (often changing) LA and state demands. In the current, high-stakes testing environment. 'It would be wise to choose a reform that can help the school improve on state and district measures of accountability.' (Datnow, p. 148)

5. A leader and faculty that believe a reform is working should keep higher administrators well advised as to the reform's value.

Datnow was clearly aware that such admonitions are easier to state than they are to live. Implementation and sustainability are both challenging and understudied. Without strong implementation there is little to sustain, and if a reform isn't sustained, the effort required to achieve implementation becomes hard to justify.

The current article reports on the four- and nine-year implementation and effects of an effort to produce sustainable reform. The reform is known as the 'High Reliability Schools' (HRS) project. Previous articles have described the theoretical underpinnings of the project (Stringfield, 1995) and initial results of the larger HRS project (Reynolds, Stringfield, & Schaffer, 2006); hence, this article only briefly summarises those areas, first by describing the principals of highly reliable organisations (HROs) in original contexts beyond education, and then the HRS project as implemented in the Welsh schools. The authors then focus on issues of implementation and sustained effects of the HRS project.

High reliability organisations

“The public would heavily discount several thousand consecutive days of efficiently monitoring and controlling the very crowded skies over London or Chicago if two jumbo jets were to collide over either city.”

For two decades, scholars in fields as diverse as political science, organisational behaviour and engineering have studied non-educational organisations that are required to work under the very unusual demand of functioning correctly 'the first time, every time' (LaPorte & Consolini, 1991; Roberts, 1993; High Reliability Organisations, 2007). Such complex social organisations as air traffic control towers continuously run the risk of disastrous and obviously unacceptable failure. The public would heavily discount several thousand consecutive days of efficiently monitoring and controlling the very crowded skies over London or Chicago if two jumbo jets were to collide over either city. Through fog, snow, total computer system failures, and nearby tornadoes, in spite of thousands of flights per day in busy skies, such a collision has never happened above any city – a remarkable level of performance reliability. Other examples of extremely high reliability can be found in electric power grids, nuclear power stations, and various medical environments.

Regardless of the sector of a society in which they work, HROs share numerous characteristics.

1. Social context: HROs evolve when both the larger society and the professionals involved in the working of the organisation come to believe that failure of the organisation to achieve its key goals would be disastrous. Thus, individual airlines are allowed to add and subtract specific routes, but both commercial aircraft maintenance and air traffic control are very closely monitored. Changing routes has little consequence for the larger society; however, one flight's total failure is unacceptable. Similarly, one badly cascading error in the 40-year life of an otherwise superbly performing nuclear power station is simply not acceptable, either for the surrounding community or the professionals working within.
2. Organisational reliability requires a clear and finite set of goals, shared at all organisational levels. No organisation performs huge numbers of tasks with very high reliability. Reliability requires priority setting and focus.

3. An ongoing alertness to surprises or lapses. Small failures in key systems are monitored closely, because they can cascade into major problems. In order to sustain multi-level awareness, HROs build powerful databases. These databases can be described as possessing 'Four Rs':
 - (a) Relevance to core goals
 - (b) Rich triangulation on key dimensions
 - (c) Real-time availability to all organisational levels
 - (d) Regularly used and cross-checked by multiple, concerned groups, both inside and outside the organisation.
4. The extension of formal, logical decision making analysis as far as extant knowledge allows. Regularly repeated tasks, which are effective, become Standard Operating Procedures (SOPs). This is in part to make 'best practice' universal, but also to allow a rich web of peer observation and communication.
5. Highly reliable organisations actively sustain initiatives that encourage all concerned to identify flaws in standard operating procedures, and honour the flaw finders.

Because high reliability is a social construction and requires high levels of individual professional decision making, HROs perpetually engage in the following three activities:
6. Active, aggressive recruiting of new staff at all levels.
7. Constant, targeted professional development/training and retraining.
8. Rigorous performance evaluation. In HROs, monitoring is necessarily mutual, without counterproductive loss of overall professional autonomy and confidence. This achievement is possible because organisational goals are clearly and widely shared, and because experience has taught all concerned that reliable success evolves through frank, protected, multi-way performance feedback.

Further,

9. Key equipment is kept in good working order.
10. Because time is the perpetual enemy of reliability, HROs are hierarchically structured. However, during times of peak activity, whether anticipated or not, HROs display a second layer of behaviour that emphasises collegial decision-making, regardless of the formal position of the decision-maker.
11. Clear, regularly demonstrated valuing of the organisation by its supervising organisation. All levels work to maintain active, respectful communication geared to the key goals of the HRO.
12. Short-term efficiency takes a back seat to very high reliability.

Two additional points relate to the above HRO characteristics. The first is that while these characteristics must necessarily be described separately, their effect is multiplicative, not additive. The total absence of any one can nullify great efforts to obtain others. Standard Operating Procedures can become mindlessly rigid in the absence of ongoing honouring of flaw-finders and process/program improvers. Aggressive recruiting in the absence of supportive, long-term professional development is futile. The first 11 characteristics, however laboriously put in place, cannot be sustained if an organisation continues a history of such poor accounting and economic prediction that it must periodically make drastic cuts in personnel, equipment, etc.

A second note concerns the description of the characteristics. It would be easy to regard each of the 12 as a stable state. In fact, all are dynamic and regularly evolving. As technologies advance, systems have the opportunity to create much richer databases. Last year's teacher recruiting effort, however successful, merely becomes the baseline for measuring this year's effort, and so on. In human organisations, reliability is a socially constructed, evolving phenomenon.

Stringfield (1995, 1998a) discussed each of these general principles in terms of educational reform efforts. Obviously many schools already exhibit several of the HRO characteristics. Diverse reform efforts are also particularly well/ill designed to achieve one or more of the characteristics. However, no previous effort focused specifically on assisting local schools in improving the reliability of their service delivery.

A previous paper reported on the four-year effects of an effort to implement HRO principles in three British LAs (Reynolds, Stringfield, & Schaffer, 2006)¹. The current article reports on a five-year follow-up within one Welsh district². Because detailed descriptions of the larger effort were provided by Reynolds *et al.*, this article will only summarise information on initial implementation and effects, and will focus more on long-term follow-up/institutionalisation. The next sections will present a brief description of the Welsh educational context, the nature of the outcome measure (GCSE examinations) and the reform as it was conducted in the Welsh HRS schools.

¹Note that the actual period of intervention was three years, but the intervention period spanned the time of four school years, so analyses of end-of-year GCSE testing is presented as a four year span between 'pre-' and 'post-intervention.'

²Analyses of data from other districts will be presented in subsequent papers.

The Welsh context

“Historically, the great majority of Welsh students were expected to leave school at age 16, and only a minority of students who were particularly interested in attending a college or university transitioned to a ‘sixth form’...”

Welsh schools are funded through and relate to Local Authorities (LAs). The LA allocates funds for refurbishing facilities, sets policies, assists in hiring teachers, provides professional development opportunities, and so forth. In Wales, as in England, power has been devolved to the teachers and Governing Boards to manage schools on a day-to-day basis. Most Welsh students attend state supported ‘infant school’ from age four until seven. At seven, students transition to primary schools, where they continue to age 11. Eleven year olds transition to secondary (comprehensive) schools, where they study through the spring of their age 16 year. The practice of ‘holding back’ or ‘retaining’ students is unheard of in British education. Groups of students who begin schooling together at age four typically complete their secondary education together.

Historically, the great majority of Welsh students were expected to leave school at age 16, and only a minority of students who were particularly interested in attending a college or university transitioned to a ‘sixth form’, where they studied for two additional years in preparation for college. However, that percentage rose steadily throughout the 20th century, and today more than 80% of Welsh (and other British) 16 year olds continue to sixth form.

As in England, the framework of the Welsh curricula is set nationally. The standardisation is not at the same level as, for example, the French education system, but it would be comparable to that of a relatively highly structured US LA. The curricula are chosen to prepare students for, among other things, successful performance on the national GCSE tests of students in their final year of secondary school.

The GCSEs

Great Britain has a long history of mandated, external-to-school assessment in students’ final year of secondary education (age 15–16) and, for those who stay on, at the end of sixth form (approximately age 18). The standard examinations for the end of secondary are the General Certificate of Secondary Education tests, or simply the GCSEs³. Virtually every British student sits for the GCSEs. The traditional measure of strong academic performance for a student is obtaining ‘five or more A*s to C’ grades on the various examinations (literature, mathematics, various sciences, etc.). Although scores have risen over the last decade, just over half of Welsh students (52.2%) obtained five or more A*–C grades in the 2005 administration of the GCSEs. The tests are viewed as relatively high stakes for both students and their schools. For students, a certain number of passing grades are required for such career options as becoming a policeman or postman, and a (higher) number is required for admission to various colleges.

Given that all English and Welsh secondary schools have essentially open admissions (i.e. students are not bound to attend the school in their specific geographic area), a rising or falling standing on the percentages of students passing five or more GCSEs can affect the number of students choosing to attend a school. In turn, teachers’ and administrators’ positions can be gained or lost. At the low end, the LA can close secondary schools that have a persistent pattern of very low scores. Well short of that extreme, it is not uncommon for a headteacher to lose his/her job if school-level GCSE scores fall for several consecutive years. The result is a testing programme that has relatively high stakes for students and educators alike.

³The GCSEs replaced the former Ordinary, or ‘0-level’ examinations. The age 18 exams are still the Advanced, or ‘A-level’ exams.

The evolution of the HRS programme in Wales

“ *One goal had to be a substantial five-year rise in the percentage of students obtaining five or more A* to Cs on the GCSEs.* ”

The High Reliability Schools (HRS) programme began in one English LA in 1995–1996 and rapidly spread to two additional LAs. Despite considerable interest from additional schools and LAs in England and Wales, the developers concluded that it would be better to focus efforts on those few locations rather than spreading efforts ever more thinly. The first district served as a pilot to subsequent efforts. At the pilot district, both the schools and the developers were co-constructing the HRS program month-to-month. The second involved a somewhat more developed implementation, and the third was the most nearly developed implementation from the start. Three-year data from all three sites have been presented in Reynolds, Stringfield, and Schaffer (2006). While there is some very promising long-term data from other schools and LAs⁴, this article reports on longitudinal follow-up from the LA that received the most mature HRS implementation, the Welsh district.

The district is an area of greater than average social deprivation, where closures and cutbacks in the basic industries of coal, steel, and engineering – historically the staples of the Welsh economy – have impacted upon activity levels and incomes. However, most parts of the district have retained elements of the traditional, cohesive community structure that has been the historic pattern for working-class areas of Wales.

The community served by this district has a higher than average rate of poverty. On a national index of social deprivation, the LA has held a steady rank of 19th of 22 Welsh districts (e.g. relatively highly deprived). Secondary school free school meals rate has remained steady in the 19%–23% range. Still the community generally possesses fewer of the extreme social problems of its inner-city counterparts, and schools in this district regard themselves as fortunate to have relatively stable foundations on which to build.

Eleven of the 12 Welsh schools in this study comprise all of the secondary schools in one LA. The 12th was from a high poverty region of a nearby LA. Of the 11 from one LA, ten were conventional state supported schools that were actively engaged in ongoing LA activities. The 11th was a ‘faith-based’ or ‘parochial’ school that was more engaged with – and accountable to – the wider community of Catholic schools and church leadership across Wales than to the LA.

The initial decision (by four relatively low-achieving schools) to join the HRS project was made in the late spring of 1996 with implementation steps begun in the autumn. The remaining eight schools joined and became fully participating members by the winter of 1996–1997. The formal programme, involving the external research team and organised engagement with schools from two English LAs continued for three years, meaning that the formal ‘HRS Project’ in Wales ran for three years. Informally, without support from the HRS design/implementation team, aspects of the project have continued within the district.

The Welsh agreement among the schools and the researchers was as follows:

1. All of the schools were to focus on 2–4 very ambitious goals. One goal had to be a substantial five-year rise in the percentage of students obtaining five or more A* to Cs on the GCSEs. A second had to be improved attendance. Each school was free to choose up to two additional HRS goals.
2. The headteachers would lead the efforts.
3. The heads and faculties implemented the HRS programme schoolwide from the start of the project.
4. All schools and departments within schools would agree to share successes and failures, and thus create learning communities across schools and LAs.

⁴One of the English LAs was recognized by the British Department of Education and Skills in 2005 as having the greatest 3-year GCSE gains in England.

“The researchers would present school-level series of workshops on the theoretical underpinnings of ‘High Reliability Organisations’, and the research bases on school effects, school change, and teacher effectiveness.”

Each school and department would commit to studying ‘best practice’, both from the international research bases and within and without the HRS schools in England and Wales.

The researchers would present school-level series of workshops on the theoretical underpinnings of ‘High Reliability Organisations’, and the research bases on school effects, school change, and teacher effectiveness. Armed with this knowledge, teachers would engage in both within- and between-school classroom observations and ‘no-fault’ feedback to peers. Importantly, all agreed that there would be no one piece of research or observational learning that would be required of any school or teacher. HRS was to rely on the well-informed and supported professional judgement of practicing educators in the diverse schools.

5. The researchers and the administrators of each school would support the faculties in becoming uniquely ‘data-rich’. Students would be given short tests as they entered the schools, and age/grade-level teams of teachers would meet and discuss how best to address each student’s needs and how to maximise each student’s chances of academic success.

The resulting student-level data sets were to be (a) rich in individual students’ academic histories, (b) available to all teachers and administrators, and (c) regularly shared/discussed by all grade-level teams within schools.

6. Almost all of the schools purchased a university-based system of storing and reporting initial intake and eventual GCSE scores. The system made it relatively easy for school personnel to compute a ‘value added’ measure.
7. All faculties and administrations committed to regularly review their organisation and processes to create widely understood, time-saving Standard Operating Procedures, and to identify and intervene in school-wide fashions with their pupils who appeared to be at risk of failure.
8. A focus on teacher effects/peer observations began immediately. This included both professional development

time to learn core aspects of the teacher effectiveness research field (ex., Brophy & Good, 1986), and for observation in classes within and among schools.

9. A strong departmental effectiveness component was emphasised (High Reliability Schools, 2006)

Several additional components quickly followed:

10. When the assessment of incoming 11 year old students at some of the schools indicated that many of the students were entering secondary school more than two years behind in reading, an immediate effort was launched to coordinate the secondary school’s literacy programmes with those of the feeder primary schools.
11. Among the Welsh schools, a part-time ‘HRS Driver’ was immediately appointed by the LA to formally coordinate activities among the district’s schools. The effect was to have HRS continuously ‘on the radar screen’ at each school and in most departments of all schools.
12. The focus in the project on the ‘broad brush’ principles of HRS and the detailed organisational features of the HRS model as outlined in the components material was supplemented in Wales by a regular focus upon what came to be called ‘the little things that matter’. HRS meetings increasingly centred upon regular sessions in which each school explained to the whole group of Welsh schools the practical things that they had done at the ‘micro’ level to embed the concepts and the components in the form of practical organisational features at the point of delivery of education to pupils in classrooms and schools.
13. Additional time for professional development was built into the Welsh implementation. Heads and faculties attended regionally-based residential sessions (two-day meetings at a conference centre) for all headteachers and HRS representatives, and also added national residential sessions; all aimed at enhancing knowledge transfers across schools and LAs. The Professional Development focus in Wales tilted strongly

“ One secondary school went so far at to use some of its own resources to provide a literacy coordinator to the four primary schools whose students it served. ”

towards turning schools into 'knowledge generators' rather than passive knowledge recipients. Particularly, the HRS project focused upon the introduction of peer observation systems to permit the charting, generation, and transmission of good practice in classrooms, training some school personnel to use observation systems which were then cascaded around the entire school.

14. The team's focus upon improving schools' capacity to be reflective about their organisational functioning and outputs was enhanced, using additional training. Examples included the provision of sessions on the statistical analysis of data and the provision of a sophisticated, relational database that teachers could access to more efficiently analyse stored grades, background information and test scores of pupils.
15. The programme began to take a close interest in the effectiveness of the primary feeder schools that were generating intakes of pupils that, in the case of most schools, were regarded as unintentionally

setting 'low ceilings' on what it was possible to achieve. Primary senior management teams were invited to the secondary schools' HRS training days. One secondary school went so far at to use some of its own resources to provide a literacy coordinator to the four primary schools whose students it served.

16. Finally, the HRS representatives and headteachers were given additional bodies of knowledge to those given to all staff. These additional materials were of two types. Some materials were focused on topics around being effective managers of change. Additionally, bodies of knowledge that were to be shared with teachers were first previewed with the headteachers and HRS coordinators. This allowed the leadership to be prepared to answer staff questions and ease the material into schools.

The HRS project began in the winter of 1996–1997 and continued through the spring of 2000. Since the GCSEs are administered annually, the period between pre- and post-measurement was four years.

Methods

For the current article, multiple data sets were used to address three key questions:

1. Did the HRS schools obtain greater gains than the Welsh national totality of schools in their GCSE scores over the three-plus years of reform implementation?
2. Did the HRS schools demonstrate greater GCSE gains over the subsequent five years (2001–2005) than did the average of all Welsh schools?
3. Were there some common characteristics across all HRS school, and other characteristics that differentiated among schools that were able to obtain greater v. lesser levels of FIRS implementation and GCSE gains?

While a wide range of data have been gathered in the HRS project, three types of data are reported in this article. The first are annual GCSE results for the 12 Welsh HRS schools. These data are aggregated annually by the government of Wales. The schools' GCSE results will be contrasted to the Welsh national average⁵ and gains over the years 1994–2005.

In order to obtain as stable a 'pre-' measure as practical, the GCSE results from the three years prior to implementation (1994–1996) were averaged. Analyses will focus on three primary time points: 'pre-HRS' (1994–1996), 'post-FIRS' (2000), and five-year follow-up point (2005).

TABLE 1: School and Reform level GCSE gains over the Four Years of the HRS Intervention

5 or More A*–C GCSE's (%)									
School/Grouping	1994	1995	1996	3-year Mean	1997	1998	1999	2000	4-year Gain
School 1	23	21	25	23	27	42	41	44	21
School 2	32	40	40	37	43	42	49	46	8.7
School 3	15	13	13	14	15	29	30	35	21.3
School 4	27	29	30	29	23	31	25	47	18.3
School 5	29	38	39	35	43	50	51	49	13.7
School 6	49	38	52	46	57	62	69	59	12.7
School 7	36	36	48	40	42	39	44	50	10
School 8	33	28	32	31	41	45	52	51	20
School 9	26	21	25	24	32	38	32	45	21
School 10	41	46	48	45	50	50	50	55	10
School 11*	48	42	35	42	50	44	58	52	10.3
District 2, School 1**	13	32	26	24	27	25	29	35	11.3
Welsh 12 Mean	32.1	32.9	36	33.7	39.5	42.9	46.2	48.1	14.5
Welsh National Mean	39	41	42	41	44	46	48	49	8.3

⁵The use of Panel data to make longitudinal comparisons in this study is justified by two facts. First the N of GCSE-taking students in these groups of schools is always large (500+ within any group of schools), and relative to the US, British Schools serve very stable populations of students. When queried, heads estimated that the total in- and out-migration of students from their schools from age 11 to GCSE taking at 16 was consistently under 3%.

“...the government conducted a multi-level statistical analysis, entering students' prior attainment (e.g. SATS tests at age 11), gender, and date of birth at the student level...”

Second, results of an LA-level GCSE 'value added' analysis, produced by the Welsh government will be presented. For this analysis the government conducted a multi-level statistical analysis, entering students' prior attainment (e.g. SATS tests at age 11), gender, and date of birth at the student level, and at the school level entering percentage of students receiving free school meals and the school mean level test scores (age 11 SATS) to predict district-level GCSE results (see Reynolds, 2006, for details).

Third, detailed case studies were developed on each HRS school on an ongoing basis as the project unfolded. Each case was updated each year of the initial project, through a series of on-site observations,

and interviews with school heads, department heads and teachers, plus the administration of implementation questionnaires. These efforts resulted in files typically more than 100 pages per school. Individual school cases have been augmented with follow-up site visits to the schools, conducted by the research team during 2006 and 2007. Field notes have been transcribed, results reached through a constant-comparative method, and conclusions triangulated among the research team.

Results

“ Addressing the question of whether the reforms effects were sustained, Table 2 presents data from the follow-up phase of the study. ”

Analyses relevant to question 1 can be seen in Table 1 on page 14. This table provides GCSE data on individual schools and the 12 schools HRS cohort of schools during three pre-intervention years (1994–1996), and data during the four subsequent school years. The individual HRS schools varied significantly in their pre-HRS achievement status. While the Welsh nationwide average percentage of students obtaining five or more A* to C scores on GCSEs during the three pre-implementation years was 41%, the 12 individual schools' averages ranged from a low of 14% students obtaining 5+ A*–C to a high of 46%. However, the weighted-by-n's of students average for the HRS schools, 33.7%⁶, was well below the national average, reflecting the district's relatively low socio-economic standing.

By the end of the formal project, spring of 2000, the Welsh national average had risen to 49%, a laudable gain of 8.3% more students obtaining 5+ A*–Cs. Over the same four years, including the three full years of FIRS participation, the 12 HRS schools raised their average percentage of students obtaining 5+ A*–C grades on the GCSEs (weighted by the n of students per school) to 48.1% for a student-mean gain of 14.5 percentages. This rate of gain was fully 75% more rapid than the nation's laudable gains. Note that the HRS school with the least gain had gained more than the national average during those years (8.7% v. 8.3% nationally), and that five schools made gains that more than doubled the national average gain. At the end of the intervention phase, the HRS project had produced substantial gains in the percentages of students succeeding on the end-of-secondary-schooling tests.

Addressing the question of whether the reforms effects were sustained, Table 2 on page 17 presents data from the follow-up phase of the study. These data extend the analyses to five full years post-HRS intervention. Note that the research/

development team had spent no time in the HRS schools during those years. Between 2000 and 2005, the percentage of Welsh students nationally obtaining five+ A*–C's had risen to 52.2%, a rise of 3.2 additional percentage points. By contrast, the 12 HRS schools, that served, on average, relatively deprived catchments areas, had 55.0% of their students with 5+ A*–C's, a rise of 6.8 percentiles. Since the beginning of the HRS project, nine of the 12 HRS schools had risen more on the national tests than had the nation as a whole. The overall group mean had risen at a rate 90% steeper than the nation as a whole (21.3% v. 11.2%). This was in a region in which the overall economic conditions had not changed significantly. A reasonable conclusion to draw from Tables 1 and 2 would be that the HRS project had substantial implementation- and follow-up phase impact on students' performance on national examinations.

An additional method for addressing the effects of the project would be the production of a 'value added' analysis. In 2005 the Welsh national government produced such an analysis on an LA-by-LA basis. As noted previously, eleven of the 12 Welsh FIRS schools comprised all of the secondary schools in a single district. Hence, while the analysis would not include all 12 schools, it would not be an unreasonable source of additional data on the project. (The one school from another LA is marked with ** in Tables 1 and 2.) Table 3 on page 18 presents data from the Welsh national government's LA 'value added' assessment of secondary education. As noted above, their analyses examined 2005 GCSE data in light of three student-level variables (date of birth, date of birth, and pupil's attainment on age 11 tests before entering secondary school). At the school level, free school meal mean, mean achievement at intake, and distribution of intake scores were entered. The Welsh government concluded that based on 2005 data, among the 22 Welsh LAs, the 'value

⁶Note that the weighting allows the sum of the data to reflect means of students (as opposed to schools' scores). Note also that in the tables, addition and subtraction occasionally appear to be slightly off, this is due to rounding.

TABLE 2: Five-Year Post-Intervention GCSE Scores for HRS Schools and Welsh National Means

School/Grouping	94–96	2000	Implementation Gain	2001	2002	2003	2004	2005	Follow-up Gain 2005–2000	2005–Pre
LA 1, School 1	23	44	21	42	46	49	48	44	0	21
School 2	37	46	9	52	48	51	55	60	14	23
School 3	14	35	21	39	41	34	33	51	16	37
School 4	29	47	18	35	40	41	36	30	-17	1
School 5	35	49	14	57	51	50	51	57	8	22
School 6	46	59	13	66	65	59	68	60	1	14
School 7	40	50	10	50	47	51	55	63	13	23
School 8	31	51	20	59	59	64	68	72	21	41
School 9	24	45	21	40	41	43	43	48	3	24
School 10	45	55	10	49	58	62	59	63	8	18
School 11*	42	52	10	52	55	55	60	52	0	10
LA 2, School 1**	24	35	11	35	24	34	31	30	-5	6
Welsh 12 Mean	33.7	48.2	14.5	49.6	49.5	51.2	52.8	55	6.8	21.3
Welsh National Mean	41	49	8	49.8	50.5	51.1	51.4	52.2	3.2	11.2

added' of secondary education ranged from a low of -8.6 percent five+ A*-C GCSEs obtained (e.g. the average school in that LA was graduating students with five+ A*-C at a rate that was 8.6 percentiles lower than would have been predicted), to a high of +6.7 percentages. The + 6.7 was the HRS district, judged the most 'value added' in Wales. This would compute to a district-level effect size of +1.82, clearly a large positive effect.

Finally, one additional, more speculative analysis of the GCSE data is possible. One of the 12 Welsh secondaries was from a separate LA, and that a second was a faith-based school with additional reporting responsibilities to its religious affiliation. While the schools were moderately involved in the implementation phase, the two were almost necessarily less connected than the average during the implementation years,

and often much less connected during the five subsequent years. The two generally less connected schools produced an overall mean gain modestly above the Welsh mean gain during the implementation years (10.9 percentiles gain vs. 8.0 nationally). However, during the follow-up years, when the HRS project itself wasn't providing regular connections, the student-weighted mean of the two less-connected HRS schools actually dropped by 3.2 percentage points. The net nine-year effect was that while the Welsh gain was 11.2 percentage points, and the ten more institutionally connected HRS schools gained, on average, 23.8 percentage points, the two less connected schools had risen only 8.0 percentage points. These last analyses, although speculative, appear to offer support for the value of systemic connections during and after the implementation phase of school reform efforts.

TABLE 3: Welsh Government District 'Value Added' Assessment

'Value Affect Size' (unweighted)		
HRS District	6.7	1.82
2	4.3	1.17
3	3.8	1.03
4	2.6	0.71
5	2.4	0.65
6	1.8	0.49
7	0.4	0.11
8	0.3	0.08
9	0.3	0.08
10	0.2	0.05
11	0.1	0.03
12	-0.1	-0.03
13	-0.5	-0.14
14	-0.8	-0.22
15	-1.2	-0.33
16	-1.4	-0.38
17	-1.8	-0.49
18	-2.4	-0.65
19	-4.1	-1.11
20	-6.6	-1.79
21	-6.6	-1.79
22	-8.9	-2.42

Reflective themes

“ Each interview began with an open-ended invitation for the educator(s) to reflect on the original HRS process, and lessons that the educator(s) believed abided. ”

During 2006 the authors returned to the Welsh schools and interviewed heads, deputies, teachers, LA personnel, and others who had been in the schools in one capacity or another throughout the intervention and subsequent years. A total of six schools were visited. These were strategically sampled to include schools that had made among the largest and smallest GCSE gains during the intervention and in the follow-up years. They included schools serving student populations that were among the largest and smallest in the district. The six included schools that had been served by one head from initial implementation to 2005, and schools that had changed heads at least once post-implementation. Six heads, nine deputy and assistant heads, and more than a dozen heads of department, heads of year and teachers were interviewed, as were the LA's Director of Education and three LA officials. A semi-structured interview protocol was followed. Each interview began with an open-ended invitation for the educator(s) to reflect on the original HRS process, and lessons that the educator(s) believed abided. Through a constant comparative process, the authors identified nine themes in the educators' perceptions of the HRS project. Seven themes were derived from those interviews. They are summarised below.

Theme 1:

The importance of a *finite number of goals that are evolved and shared schoolwide*. Every educator talked about the importance and – at that time – the novelty of all teachers and administrators working toward a modest number of clearly articulated, shared goals. Heads and teachers talked about the value of believing in reasons for everyone to talk and work together toward shared objectives.

Theme 2:

The centrality of, and evolving sophistication with, *data and data analysis* for practical improvements. Beginning with fairly crude data analysis and basic disaggregation, several of the schools have evolved sophisticated disaggregation mechanisms and analysis. The district has recently bought an integrated

data warehousing software which is not ideal, but which is allowing further, more efficient breakdowns of data in ways the heads are quickly exploring.

Theme 3:

Standard Operating Procedures: Standard Operating Procedures were found to benefit the schools and teachers for responding to consistent problems in school, including absences, behaviour management, data collection, and other daily or weekly activities. Teachers and heads alike discussed the value of moving to increasingly standardised procedures for handling the necessary drudgery of educational systems.

Theme 4:

Seeking Best Practice. This had several subcomponents that various interviewees noted. One was the formal 'requirement' (in reality, often viewed as a novel privilege) of observing in higher achieving departments and schools. This was assisted by having attended workshops in which multiple dimensions of 'effectiveness' research had been presented, so that teachers acquired skills in accurately measuring student engagement/time-on-task, teachers' questioning patterns/wait time, etc., and could look for examples of positive school culture, data use, etc. Among the teacher interviewees, the experiences of visiting classrooms in other, demographically similar but higher achieving schools was one of the most often mentioned components of the programme.

Theme 5:

Off-Site Residentials. Much more so than the researchers/developers, the heads and teachers recalled specific learnings from the two-day residential retreats that focused on cross-school examination of 'best practice'. In the residentials, each school would present examples of HRS-related activities that they perceived to be working well in their context. The groups would then discuss how these strategies might carry on to other schools, what impediments they might face, and how the strategy would improve the quality of teaching and learning.

Theme 6:

Skillfully managed leadership successions.

Eight of 12 Welsh HRS schools changed heads between 1998 and 2006. In seven of eight change cases, the former head was replaced by someone who had worked in one of the more aggressive HRS-engaging schools. In almost every case the new head immediately recommitted to – and occasionally reenergized – HRS-related reforms. Several of those schools are now among the most achievement-gaining schools in the district/project. Whereas in typical US school reform efforts, the initiating principal is replaced by someone unfamiliar with the effort, in the Welsh HRS case HRS leaders were replaced by HRS leaders. In each case, the message to the faculty and staff was, 'This [HRS] is the path that we are going to pursue. We are going to have a few shared goals. We are going to be closely advised by our student data, and we are going to seek out best practice wherever we can find it'. Often the programme prospered more under the second HRS-trained head than the first.

Theme 7:

Practitioners noted that the HRS process had a *cyclical effect*. The initial gains in achievement encouraged staff and allowed for a second level of effort. When that succeeded, a third level became possible. The fact that School 1 (in Table 1) jumped from years in the low teens to the high 20s over two years gave credibility to the new head's drive to use data and a search for 'best practices' to inform decisions. This combination of teacher

enthusiasm and leadership authority lasted for several years and provided a new head with the authority to push forward with HRS, eventually resulting in a remarkable gain to 51% in 2005. When the school was at 13% 5+ A*-C, there was nothing anyone could say to convince the majority of the faculty that 30% was an even theoretically achievable goal and hence not worth the effort. But once the school achieved 30%, 40% seemed possible and new initiatives seemed worthwhile. The path from 13% to 51% was iterative, and not always linear. In fact, in no case was the route from the beginning of HRS to 2005 linear. It was more accurately described as cyclical. Similarly, expectations for achievement outcomes rose in stages. The head of one of the most successfully rising schools looked back after nine years in the interview, and observed how he once thought it would be a miracle if his school ever achieved a 40% five+ A*-C rate. That school is now at 50% with plans for achieving 60%. Several less disadvantaged schools in the project, none of which had previously imagined 60% are now working toward what they believe to be an achievable goal of 80%. The quantitative goals are nothing more than measurable artifacts of larger, less readily quantifiable but equally laudable goals. In each case, the achievement of a readily observable goal has allowed the upward re-benchmarking of a range of desirable outcomes. When asked in 2006 about their initial goals, the great majority of HRS-school personnel smiled and stated confidence in their ability to achieve new, much higher goals. These are, in themselves, laudable achievements.

Discussion

“ Research on Highly Reliable Organisations suggested that in a range of settings, improving reliability might be a path to improved performance. ”

The fundamental hypothesis of the HRS project was that by working with schools to heighten the *reliability* of school functioning and reform, it would be possible to improve desired student outcomes. We hypothesised that it would be possible for a group of researchers, working together with local educators, to co-construct more nearly highly reliable schools. Four- and nine-year changes in school processes and increases in the GCSE scores of students in the HRS schools support that hypothesis.

A range of authors has observed that there are few-to-no prior examples of multi-school, coherent examples of achievement-enhancing secondary school reform. We believe that the substantial rise in academic achievement within the HRS project is the clearest evidence to date that substantial change-bearing reform of secondary schools is possible at scale.

The effectiveness of this programme, by contrast, is likely to reflect on HRS programme content, and processes that, rarely in other designs, involved:

1. Schoolwide focus on a finite set of goals.
2. The relentless gathering and use of data of all types. Getting the data in the hands of all persons involved in the education of each student: the head, the department heads, the teachers and the students.
3. Giving schools existing 'good practice' from the academic literature, and also helping schools develop relatively high quality data systems.
4. From the beginning the HRS project built on a combination of technical, scholarly knowledge of HROs and various 'effectiveness' literatures together *in equal partnership with* local educators' knowledge of and skill in working within local educational contexts. The HRS principles and programme may have been a catalyst, but the skill and commitment of local educators produced the outcomes. The HRS project was co-constructed from beginning to end.
5. Working with and through multiple levels in the education environment. This included

focusing upon the classroom and the middle-management tier of schools in addition to a focus at the conventional school level.

6. Building capacity at school site level to continue educational development after the formal end of the project.

As such, we believe the HRS programme is representative of the 'new wave' of school improvement argued for by Hopkins and Reynolds (2001).

To summarise, the authors began with the hypothesis that part of the challenge to more consistent positive results lay in improving reform reliability. Research on Highly Reliable Organisations suggested that in a range of settings, improving reliability might be a path to improved performance. Several practical conclusions may be drawn from the Welsh HRS work:

First, measurable-change-bearing reform is possible at the secondary school level. The fact that an additional 6.7% more students than predicted of an entire LA's secondary students achieve highly in a national testing programme, producing an effect size of +1.82, constitutes an existence proof.

Second, in the HRS case, the gains were guided by prior research on school effects and teacher effects, thus lending support to both fields.

Third, the test of whether a core challenge was increasing the reliability of implementation vs. the addition of esoteric 'new ideas' produced results that favoured increasing reliability. HRO principles of a focus on a small set of goals, a relentless examination of data, and the development of Standard Operating Procedures, among other HRO characteristics, were associated with greater success.

Fourth, the gains were clearly the result of a team embracing co-construction. Rather than pushing relentlessly for 'fidelity' of implementation of an original set of ideas, the

HRS development team pushed for an on-the-ground co-construction of school reform. By 2006, each successful school's HRS included some general principals in common (focus on a few goals and on data, search for best practice, collaboration across departments and schools, etc.). Each also claimed unique, locally-developed properties that helped the school become more nearly reliable in the services it provided to students. Among other advantages, the co-construction process filled in details as to how to successfully implement abstract academic concepts and greatly increased local ownership of the process.

A related observation would be that it was the combination of external facilitators and local professional educators that achieved positive results. Co-construction has no meaning if there aren't two partners. In every 2006 follow-up interview, the local educators stated that it was the combination of challenging, externally-derived principles with local energy and application that resulted in success.

Each school's evolution of standard operating procedures (SOPs) became a case in point. Some of the SOPs became standardised across virtually all schools, but each school evolved widely-agreed-upon SOPs that were unique to that school. Regardless of the specifics of a school's SOPs, professionals at every school participating in the follow-up interviews mentioned the advantages of having evolved SOPs that worked in their context.

Sixth, and as in most previous research on school change, the role of the head was pivotal in schools' progress. Where a head was focused on HRS goals yet open minded as to the specifics of achieving the goals, gains were achieved.

Seventh, the HRS project produced data that appear to support a focus on systemic supports for reform. At the least, the data suggest that having formal mechanisms and informal norms supporting reaching out across schools to answer practical problems was associated with success.

Finally, while the definition of 'institutionalisation' remains imprecise in the scholarly literature, the HRS study makes clear that not only are

student achievement gains possible during a multi-year implementation of a reform, but that those gains can continue for at least five years post-intervention.

The HRS project also has implications for future research. First among them is the need to replicate the study's findings in additional settings. A single study, however promising, is one study. Replication would be a valuable addition to the field.

Second, our field needs studies exploring the upper boundaries of the possible. Within the HRS project, one school that prior to HRS had consecutive years of only 13% of its students obtain five+ GCSEs, saw its 2005 cohort rise to 51%. A school that had 28% rose to 72%. Both of these gains occurred in catchment areas that remained essentially unchanged over time. In 1997, neither school's head nor faculty would have said such gains were humanly possible. When the Motorola Corporation declared the goal of 'six sigma' product reliability, engineers and researchers around the world scoffed. Then Motorola achieved it, and has now set the bar higher. What is possible for secondary schools in high poverty and other contexts? Multiple proactive studies are needed.

Third, studies of promising whole school programmes in the UK, the US, and elsewhere have typically failed to produce measurably positive student outcomes. The HRS project raises the possibility that the problem was not the reform ideas but the reliability of implementation. Studies are needed of diverse promising programmes being implemented through HRS principles.

Finally, the HRS project suggests the need for more proactive, long-term, follow-up studies. It is worth remembering that the goals of education for students are not short term. Similarly, a school that produces dramatic gains but does not sustain them has contributed little to the over-all educational environment. In HRS, the nine-year gains were perhaps as compelling as the four-year gains. Has this been true of previous reforms but researchers discontinued data gathering too quickly to observe the effects? Only long term, proactive data gathering can address that issue.

References

- Aikin, W. (1942). *The Story of the Eight Year Study*. New York: Harper.
- Berman, P. & McLaughlin, M. (1977). *Federal programs supporting educational change, Vol. VII, Factors affecting implementation and continuation*. Santa Monica, CA: Rand.
- Borman, G., Hewes, G., Overman, L., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73(2), 125–230.
- Borman, G.D., Slavin, R.E., Cheung, A., Chamberlain, A., Madden, N., & Chambers, B. (2005). The national randomized field trial of Success for All: Second-year outcomes. *American Educational Research Journal*, 42, 673–696.
- Brophy, J.E., & Good, T. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Handbook of research on teaching, third edition* (pp. 328–376). New York: Macmillan.
- Grandall, D.P., Loucks-Horsley, S., Bauchner, J.E., Schmidt, W.B., Eiseman, J.W., Cox, P.L. (1982). *People, policies and practices: Examining the chain of school improvement* (Vols. 1–10). Andover, MA: The NETWORK.
- Cuban, L. & Usdan, M. (eds.) (2002) *Powerful reforms with shallow roots*, New York: Teachers College.
- Darling-Hammond, L., Hightower, A., Husbands, J., LaFors, J., Young, V., & Christoher, C. (2006) Building instructional quality: 'Inside-out' and 'Outside-in' perspectives on San Diego's school reform. In A. Harris and J. Chrispeels, *Improving schools and educational systems*, London: Routledge.
- Datnow, A. (2005) Happy marriage or uneasy alliance? The relationship between comprehensive school reform and state accountability systems, *Journal of Education for Students Placed at Risk*, 10, 113–128.
- Datnow, A., Borman, G., & Stringfield, S. (2000) School reform through a highly specified curriculum: A study of the implementation and effects of the Core Knowledge sequence. *Elementary School Journal*, 101(2), 167–192.
- Datnow, A., Hubbard, L. & Mehan, H. (1998) *Educational reform implementation: A co-constructed process*. Santa Cruz, CA: University of California, Center for Research on Education, Diversity, and Excellence.
- Datnow, A. & Stringfield, S. (2000). Working together for reliable school reform. *Journal of Education for Students Placed at Risk*, 5(1), 183–204.
- Earl, L., Torrance, N., & Sutherland, S. (2006). Changing secondary schools is hard: Lessons from 10 years of school improvement in the Manitoba School Improvement Program. In A. Harris and J. Chrispeels, *Improving schools and educational systems*. London: Routledge.
- Edmonds, R. (1979) Effective schools for the urban poor. *Educational Leadership*, 37, pp 15–27.
- Fullan, M.G. (1991) *The new meaning of educational change*. New York: Teachers College.
- Hargreaves, A., & Fink, D., (2006) *Sustainable Leadership*. San Francisco: JosseyBass/Wiley.
- High Reliability Organizations. Retrieved March 17, 2007 from <http://highreliability.org>.
- LaPorte, T. & Consolini, P. (1991) Working in practice but not in theory: Theoretical challenges of high-reliability organizations. *Journal of Public Administration Research and Theory*, 1(1), 19–48.
- McLaughlin, M. (1990) The Rand Change Agent Study Revisited: Macro perspectives and micro realities. *Educational Researcher*, 19(9), 11–16.
- Reynolds, D. (2006) World class schools: Some methodological and substantive findings and implications of the International School Effectiveness Research Project (ISERP) *Educational Research & Evaluation*, 12, 535–560.
- Reynolds, D., Hopkins, D., Potter, D., & Chapman, C. (2001). *School improvement for schools facing challenging circumstances: A review of research and practice*. Exeter, UK: University of Exeter.

- Reynolds, D., Stringfield, S., & Schaffer, E. (2006) The High Reliability Schools Project: Some preliminary results and analyses (pp. 56–76). In J. Chrispeels & A. Harris (Eds.), *School improvement: International perspectives*. London: Routledge.
- Roberts, K. (1993). *New challenges to understanding organizations*, New York: Macmillan.
- Ross, S., Sanders, W., Wright, P., Stringfield, S., Wang, L., & Alberg, M. (2001) Two- and three-year achievement results from the Memphis Restructuring Initiative. *School Effectiveness and School Improvement*, 12(3), 323–346.
- Rutter, M., Maugham, B., Mortimore, P., Oston, J. & Smith, A. (1979) *Fifteen thousand hours: Secondary schools and their effect on children*. Cambridge, MA: Harvard University Press.
- Schaffer, E., Nesselrodt, P., & Stringfield, S. (1997) *Impediments to reform: An analysis of destabilizing issues in ten promising programs*. Arlington, VA: Educational Research Service.
- Sebring, P., Bryk, A., & Easton, J. (1995) *Charting reform: Chicago teachers take stock*. Chicago, IL: Consortium on Chicago School Research.
- Slavin, R., Madden, N., Dolan, L., Wasik, B., Ross, S., Smith, L. & Dianda, M. (1996). Success for All: A summary of research. *Journal of Education for Students Placed at Risk*, 1(1), 41–76.
- Slavin, R., Madden, N., Karweit, N., Livermon, B., & Dolan, L. (1990). Success for All: First-year outcomes of a comprehensive plan for reforming urban education. *American Education Research Journal*, 27(2), 255–278.
- Stallings, J. & Kaskowitz, D. (1974). *Follow through classroom observation evaluation 1972–1973*. (SRI Project URU–7370). Menlo Park, CA: Stanford Research Institute.
- Stringfield, S. (1995). Attempts to enhance students' learning: A search for valid programs and highly reliable implementation techniques. *School Effectiveness and School Improvement*. 6(1), 67–96.
- Stringfield, S. (1998a). An anatomy of ineffectiveness. In L. Stoll and K. Myers (Eds.), *No quick fixes: Perspectives on schools in difficulties* (pp. 209–211). London: Falmer.
- Stringfield, S. (1998b). Organizational learning and current reform efforts. In K. Leithwood and K. S. Louis (Eds.) *Schools as Learning Communities* (pp. 255–268). Lisse: Swets & Zeitlinger.
- Stringfield, S., Millsap, M.A., Herman, R., Yoder, N., Brigham, N., Nesselrodt, P., Schaffer, E., Karweit, N., Levin, M., & Stevens, R. (with Gamse, B., Puma, M., Rosenblum, S., Beaumont, J., Randall, B., & Smith, L.) (1997) *Special Strategies Studies Final Report*. Washington, DC: US Department of Education.
- Teddlie, C. & Reynolds, D. (2000). *Handbook of research on school effectiveness and improvement*. London: Falmer.
- What Works Clearinghouse. Downloaded April 7, 2007 from <http://www.whatworks.ed.gov/>.



CfBT Education Trust
60 Queens Road
Reading
Berkshire
RG1 4BS
0118 902 1000
www.cfbt.com