



# Creating and Sustaining Secondary Schools' Success: Sandfields, Cwmtawe, and the Neath-Port Talbot Local Authority's High Reliability Schools Reform

---

Sam Stringfield | David Reynolds | Eugene Schaffer

**Conference Paper**

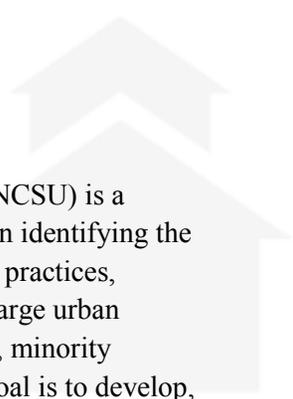
June 2012

Prepared for Achieving Success at Scale: Research on  
Effective High Schools in Nashville, Tennessee



THE UNIVERSITY  
of NORTH CAROLINA  
at CHAPEL HILL





The National Center on Scaling Up Effective Schools (NCSU) is a national research and development center that focuses on identifying the combination of essential components and the programs, practices, processes and policies that make some high schools in large urban districts particularly effective with low income students, minority students, and English language learners. The Center's goal is to develop, implement, and test new processes that other districts will be able to use to scale up effective practices within the context of their own goals and unique circumstances. Led by Vanderbilt University's Peabody College, our partners include The University of North Carolina at Chapel Hill, Florida State University, the University of Wisconsin-Madison, Georgia State University, and the Education Development Center.

This paper was presented at NCSU's first national conference, *Achieving Success at Scale: Research on Effective High Schools*. The conference was held on June 10-12, 2012 in Nashville, TN. The authors are:

Sam Stringfield \*  
*University of Louisville*

David Reynolds  
*University of Southampton, U.K.*

Eugene Schaffer  
*University of Maryland – Baltimore County*

---

\* Sam Stringfield is currently the Director of the School of Education at the University of Cincinnati.

This research was conducted with funding from the Institute of Education Sciences (R305C10023). The opinions expressed in this article are those of the authors and do not necessarily represent the views of the sponsor or the National Center on Scaling Up Effective Schools.

# Creating and Sustaining Secondary Schools' Success: Sandfields, Cwmtawe, and the Neath-Port Talbot Local Authority's High Reliability Schools Reform

## Abstract

In this paper we discuss a fifteen-year study of the High Reliability Schools (HRS) secondary educational reform effort in one Welsh Local Authority (LA) and two of its schools. After briefly reviewing research on educational improvement efforts that fed into this project, we discuss the evolution of the HRS project and its long-term effects.

Three-year pre-HRS student achievement data indicate that sixteen-year old students in the schools and their LA were performing well below the Welsh average. Over the subsequent four years, the two schools' students made gains at over twice the national average rate, and the larger LA's gains were nearly double the national average gain. Achievement data from 5- and 11-years post-intervention indicate that the schools and the LA have continued to rise to well above the national averages. Qualitative observations and interviews with the schools' and LA's leadership and student groups, combined with analyses of recent Her Majesty's Inspectors' reports indicate that schools are continuing using HRS principles and are further refining/evolving the principles. Implications for future school reforms and research on reform efforts are presented.

## Introduction

The effort to create highly reliable secondary school improvement had four research roots: school effectiveness, school improvement, research on loose vs. tight organizational coupling, and research on High Reliability Organizations. Each will be briefly described as prelude to discussion of the High Reliability Schools project. Our effort will not be to provide full literature reviews in any of those areas, but rather to highlight the connection to the evolution of the HRS project.

### *School Effects Research*

The school effects research base initially evolved through a series of case studies of unusually effective elementary schools (Weber, 1971; Brookover, 1979; and in Britain, Rutter et al., 1979). Of particular importance for the HRS project was the Louisiana School Effectiveness Study (LSES, Teddlie & Stringfield, 1993). LSES had involved an attempt to conduct a double blind in which teams of observers were sent to demographically matched pairs of outlier schools. The design called for analysis of low-inference quantitative and various qualitative data sets to determine consistent student, classroom, and school-level differences between the positive and negative outliers. However, one aspect of the “blind” design failed almost immediately. All observers on all teams correctly intuited which schools were positive and which were negative outliers. An analysis was conducted to determine how otherwise uninformed observers were able to predict, with 100% accuracy, each school’s status (Stringfield & Teddlie, 1991). It was during the months of qualitative and quantitative data reviewing associated with this task that Stringfield concluded that the issue wasn’t so much clearly exemplary schooling throughout the positive outlier schools as the striking unevenness in 12 aspects of instruction and management in the negative outlier schools. While the positive outliers weren’t always sparkling, they bore a striking resemblance to Pfeiffer’s (1989) description of studies of High Reliability Organizations. Perhaps producing strong results in student achievement was more a function of avoiding harmful dysfunction as of doing one or more things brilliantly<sup>1,2</sup>.

### *Research on School Improvement Efforts*

Economists are fond of referring to theirs as “the dismal science.” Were educational researchers a more competitive lot, we might contest that label. Relatively rigorous studies of school improvement efforts have not, in general, produced grounds for huge optimism or provided a great deal of clearly actionable direction. Volumes with titles like *So Much Reform, So Little Change* (Payne, 2008) and *The Implementation Gap* (Supovitz & Weinbaum, 2008) provide a flavor of much of what is known and not known about how to improve schools.

Three of the findings that had run from the *Eight Year Study* (Aiken, 1942) through the *Rand Change Agent Study* (Berman & McLaughlin, 1978; McLaughlin, 1990) to the *Special Strategies Studies* (Stringfield et al., 1994, 1997) were relevant to HRS, and they follow. (1) Many, perhaps most, school

---

<sup>1</sup> In subsequent evolution of the High Reliability Organization literature, this has come to be known as avoiding “the normalization of deviance” from acceptable practice (Vaughn, 1996). The normalization of deviance is typically a precursor to a seriously cascading failure..

<sup>2</sup> For updated reviews of school effects research, see Teddlie & Reynolds, 2000; Reynolds et al., in press.

that begin various reform efforts never fully implement them. (2) The level of implementation was often associated with the extent to which schools were able to achieve desired moderate-term outcomes<sup>3</sup>. (3) In general, obtaining implementation and desired outcomes has been documented much more often by reformers working with elementary than with secondary schools<sup>4</sup>. This finding presaged a conclusion drawn by Earl, Torrance and Sutherland (2006), in their decade-long study of focused secondary school reform efforts in Manitoba, Canada, “[T]here 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.” (p. 126, italics in original).

### *The Issue of “Loose Coupling”*

Much of the research on school effectiveness and improvement implies an increased set of linkages among various actors. The study of organizational change is the study of trying to get whole groups to move in a single, typically interconnected direction. By contrast, Karl Weick (1976) famously observed that American education is a “loosely coupled” set of systems. Weick was not being either complimentary or critical; rather, he was describing what was. The very existence of substantial research bases on teacher-, school-, and system-effects (Brophy & Good, 1986; Teddlie & Reynolds, 2000; Stringfield & Yakimowski, 2005; Zavadsky, 2009) implies substantial variance among units, and, in turn, relatively loose coupling between units. In a report on a nine-nation comparative school effects study (Reynolds & Stringfield, 1994; Reynolds et al., 2002) found that the within-country and within school variance in teaching were substantially greater in English-speaking countries than in two higher achieving Asian countries, where coupling was clearly tighter.

There are some advantages to loose coupling in systems. Individual units (teachers within schools, schools within school districts, districts within states) have greater freedom to experiment with new ideas and are relatively well buffered from dysfunctions in other units within the system (Glassman, 1973), such as ineffective teachers (Brophy, 1988), schools (Stringfield & Teddlie, 1988) or systems (Stringfield & Yakimowski, 2005). The buffers can work within one level, or between two or more levels, as when a highly effective teacher is buffered in an otherwise highly ineffective school or district (Stringfield & Teddlie, 1991). Relevantly to public education, Weick hypothesized that, “a loosely coupled system should be relatively inexpensive to run because it takes time and money to coordinate people.” (p. 8).

The disadvantages of loose coupling are equally obvious. By far the largest becomes involved when efforts are made to affect coordinated, directional change to a system. If a school, school system, or reform group wishes to bring a change to all the classrooms in that system of classrooms, tighter coupling would greatly enhance the speed and depth of implementation. The absence of such tight coupling could be hypothesized to lead to studies finding large variance in levels of reform-effort implementation (ex. Carter, 1984; Datnow et al., 2003; Payne, 2008; Stringfield et al., 2004, 2007; Supowitz & Weisman,

---

<sup>3</sup> For updates of this research base, see Nunnery (1998); Borman, Hewes, Overman, & Brown (2002); Datnow & Stringfield (2000), Supowitz & Weinbaum (2008); and Hopkins et al., (in press)..

<sup>4</sup> Muncey and McQuillan (1997) were particularly coherent on the challenges facing efforts at secondary school reform. In a subsequent meta-analysis of over 20 whole school reforms, Borman et al. (2002) found no secondary school reform designs that had produced consistently positive results.

2008) which, collectively, found substantial variance between classrooms and schools attempting to adopt various reform programs.

Weick observed that no organization could be tightly coupled on all dimensions. Rather, at best, organizations consciously choose a few dimensions on which to be more tightly coupled. The choice of a few dimensions on which to be relatively tightly coupled became a significant theme in HRS.

### *Sustainability*

It is hard to justify the effort required to create change if it will not be continued beyond the initial effort. Cuban (1992) observed that even school reform efforts that have been reasonably successful often do not last. From the *Eight Year Study* through *Special Strategies*, the majority of schools either never implemented their various reforms with any depth or discontinued their efforts within a few years. Hargreaves and Fink (2006) reflected on the rarity of sustained school reform efforts and stated that, “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.” (p. 2). Their long-term follow-up of eight initially successful secondary school reform efforts in New York State and Ontario, Canada concluded, “The overall evidence is not uplifting.” (p. 253). However, the authors made suggestions related to sequencing priorities, developing multiple indicators of collective accountability, systems for data management, providing mentors, developing and sharing leadership succession plans, formal and ongoing/informal professional development, and a regular auditing and pruning of past policies before creating new ones<sup>5</sup>.

Taken together, the research on teacher- and school-effects, and on school and program improvement efforts was interpreted by Stringfield (1995) as indicating that schools in America, and most Western societies, were operating in generally loosely coupled ways that made full implementation of almost any school improvement effort unlikely. He hypothesized the value of educators learning from the then-emerging research base on High Reliability Organizations.

### *Highly Reliable Organizations (HROs)*

Traditionally, theories of business and management have assumed a “trial and error” path to improvement. However, in the 1980s it became clear that there were organizations that were being required to operate under the previously believed to be impossible constraint of operating without major failing, “the first time, every time” (LaPorte & Consolini, 1991). Pfeiffer (1989) provided an early description of organizations such as air traffic control towers, electric power grids, aircraft carriers’ takeoff and landing operations within an emerging HRO framework. At the time, a small group of scholars at the University of California at Berkeley working in the diverse disciplines of business,

---

<sup>5</sup> Datnow (2005) has noted several characteristics of reforms that lasted at least four years. Those reforms: committed to and planned for the long-term, placed fewer long-term demands on the system and its resources, helped local educators meet current and emerging local, district, and state demands, built sufficient capacity to address high-stakes accountability pressures, and kept policy systems aware of their potential positive—and negative—impacts on reform efforts so as to maximize stability in the environment.

political science, and engineering had teamed together to study these phenomena. (Roberts, 1991; Roberts & Bea, 2001; Rochlin, LaPorte, & Roberts, 1987).

The early studies of HRO were descriptive case studies, either of HROs themselves or of similar units that seemingly should have worked reliably and did not. An example of the later was Roberts (n.d.) “Bishop Rock dead ahead: The grounding of the U.S.S. Enterprise.” This was a striking analysis of the very human errors that lead to the grounding of a very large, sophisticated, U.S. nuclear aircraft carrier, the only such grounding in modern naval history. This contrasted with the extremely error-resistant performance of the U.S.S. Carl Vincent (Pfeiffer, 1989). The technologies were virtually identical; the differences were in human performance and evolved human systems. The differences lay in how people were organized and treated.

In the years following Pfeiffer’s article in *Smithsonian*, a diverse group of persons working in areas ranging from hospital emergency rooms to wildfire fire fighters and special operations units of the U.S. military have adopted the core principles of HRO, often with striking success.

Impressed with the similarities between stable positive outlier LSES schools and HROs, Stringfield (1995) went on to describe characteristics of HROs as relevant to a teacher- and school-effectiveness researcher as follows<sup>6</sup>:

1. Organizational reliability evolves under a particular circumstance. HROs evolve when both the larger society and the professionals involved in the working of the organization come to believe that failure of the organization to achieve its key goals would be disastrous. By the mid-1990’s there was an emerging awareness of “a sharp acceleration in the growth of earnings inequality, particularly among men” (Levy & Murnane, 1992, p. 1333)<sup>7</sup>. While the difference in incomes in 1949 for young adult U.S. males who were high school dropouts vs. college graduates was approximately 30%, by the mid-1990’s it was over 100%. Educational failure had become so expensive that neither individuals nor society could afford it. (Today the difference is over 150%.) Similar changes were and are happening in Britain, and throughout the developed world.
2. HROs require a clear and finite set of goals, shared at all organizational levels.
3. An ongoing alertness to surprises or lapses. Small failures in key systems are monitored closely, because they can cascade into major problems. Weick and Sutcliffe (2007) refer to the solution as creating a “mindful infrastructure.” In order to sustain multi-level awareness, HROs build powerful databases. These databases can be described as possessing “Four R’s”: relevance to core goals; rich triangulation on key dimensions; real-time availability to all organizational levels; and rich triangulation on key dimensions.

---

<sup>6</sup> Various scholars working in the HRO field have described characteristics and processes involved in HROs somewhat differently (ex., Robert, 1993; Weick & Sutcliffe, 2008). Similarly, Bellamy (2011), describes possible uses of HRO in education with a greater focus on processes than in our listing.

<sup>7</sup> For more recent discussions of the education and growing economic inequality, see Friedman and Mandelbaum (2011) and Reich (2011).

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).
5. Highly reliable organizations actively sustain initiatives that encourage all concerned to identify flaws in standard operating procedures, and honor 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, extensive recruiting of new staff at all levels.
7. Constant, targeted training and retraining.
8. Rigorous, multi-way performance evaluation.

Further:

9. Key equipment is kept in high 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 behavior that emphasizes collegial decision-making, regardless of the formal position of the decision-maker.
11. Clear, regularly demonstrated valuing of the organization by its supervising and surrounding organizations. All levels work to maintain active, respectful communication geared to the key goals of the HRO. And importantly,
12. Short-term efficiency takes a back seat to very high reliability.

Two additional points related to the HRO characteristics. The first was that while these characteristics must necessarily be described separately, *their effects are presumed to be 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 honoring 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 organization continues a history of such poor accounting and economic prediction that it must periodically make drastic cuts in personnel, equipment, etc.

A second note concerned the description of the characteristics. It would be easy to regard each of the 12 HRO characteristics as existing in a stable state. In fact, all were assumed to be *dynamic and regularly evolving*. As technologies advance, systems have opportunities to create much richer databases. Last year's teacher recruiting effort, however successful, becomes the baseline for measuring this year's effort, and so on. In human organizations, reliability is a socially constructed, evolving—or devolving—phenomenon.

David Reynolds subsequently presented the abstract ideas of schools operating as HROs to several groups of British school district leaders and Headteachers (equivalent to U.S. principals). Three groups of secondary schools quickly volunteered to become participants in an experimental effort to move these

abstract principles into practical school improvement programs. The first group of schools suffered from being exposed to a set of ideas that were not yet well-matched to the realities of schools and from continuous competition—rather than cooperation—among the schools. The second Local Authority (LA) experienced a deep recession almost as soon as they attempted to initiate the HRS effort. The cuts they were forced to make violated the 12<sup>th</sup> principle of HROs. While the schools and LA eventually produced large achievement gains, their implementation story is complex and not well suited to a brief paper. The third group was immediately more successful in implementing the HRO principles, both as individual schools and as a LA. The remainder of this article focuses on two participating schools in the Neath-Port Talbot LA that received the full, three-year HRS intervention without serious interruption<sup>8</sup>.

## Research Methods

Any effort covering 15 years will almost necessarily involve a complex set of data gathering efforts, and HRS is no exception. From the inception of the HRS project in the fall of 1996, the authors conceived the effort as both a secondary school reform effort in which we were significant participants and as a mixed-methods, longitudinal research project. Our active involvement in delivering content and providing formal feedback ended in the spring of 2000. For the subsequent 12 years our roles have been restricted to being gatherers, archivers and analysts of quantitative and qualitative follow-up data. In the paragraphs that follow, we will describe the project's sample, timeframe, processes and outcome measures.

### *Sample*

Neath-Port Talbot (NPT) is a county borough with a total population of approximately 137,000. The borough's population peaked in the 1930s, when coal mining and iron production were the main sources of jobs. The borough's two major towns are Neath and Port Talbot, with smaller towns and rural areas surrounding them. NPT is located along the Swansea Bay, on the southern boundary of Wales, an hour's drive west of the capital of Cardiff. Traditionally, the area's economy was based on coal mining and steel mills; however the last mines shut three decades ago and the steel mills subsequently contracted by over two-thirds. On a series of measures of poverty, NPT has been stable for over two decades, scoring 19<sup>th</sup> of 22 Welsh jurisdictions on measures of social deprivation.

The NPT LA supports 11 secondary schools. As can be seen in column 2 of Table 1, over the three years preceding the HRS project, NPTs age 16 students were achieving on the national examinations at levels well below the national average. Given the local economic context, this was not regarded as surprising and was causing larger concern only for the very highest poverty/lowest achieving school (Sandfields, described below). The LA was relatively new at the beginning of the HRS project, having been one of three districts devolved from a previous, large authority that the national government had not viewed as providing adequate services to all schools and students.

Four secondary schools had originally volunteered to participate in the HRS programme in late 1996.

---

<sup>8</sup> The intellectual underpinnings and full history of the High Reliability Schools project have been described in greater detail elsewhere (Stringfield, 1995; Reynolds, Stringfield, & Schaffer, 2006; Stringfield, Reynolds & Schaffer, 2008, in press; Schaffer, Reynolds & Stringfield, 2012).

However, their initial reports of the value of the project were sufficiently compelling that all 11 NPT secondary schools joined within a few months. All continued their participation for the next three-plus years.

Sandfields in 1996. The community of Sandfields is comprised of two electoral wards. In 2005, each was listed as being among the 50 most socially deprived wards in Wales. Sandfields is surrounded on two sides by the Swansea bay, on a third side by an industrial park, and on its long north-eastern boundary by the M-4, a limited access motorway<sup>9</sup>. The Sandfields community was built in the 1950's as a "housing estate" (in the U.S., public housing) to provide housing for workers in the nearby steel mill. The mill has subsequently reduced its workforce by over three-quarters, and unemployment is a major challenge. Nearly 40% of Sandfields' students qualify for free schools meals, nearly triple the national average of approximately 15%.

Sandfields Comprehensive School was built in 1958, and serves approximately 650 secondary school students (age 11-16) each year. The school has been described as being the first "purpose-built" comprehensive (secondary) school in Wales. At the beginning of the HRS project, the school itself presented a generally un-kept appearance.

In the three years before the project, Sandfields' 16 year olds had among the lowest percentages of students obtaining 5+ A\*-C GCSEs examination scores (described below) in Great Britain, with an annual average of below 14%. In 1996, just before the start of the HRS project, Sandfields had an HMI inspection that was strongly critical of most aspects of the school. The report included documentation of the school's students' very low scores on the GCSEs. The report took the very unusual step of recommending that if the school did not show clear improvement, it should be closed. When the opportunity to participate in the High Reliability Schools project came, Sandfields' was one of the first four Welsh schools to sign on—perhaps in part reflecting their faculty's and administration's anxiety to be perceived as making good faith improvement efforts.

Cwmtawe Comprehensive School in 1996. Cwmtawe was created in 1969 to replace three schools serving the upper Swansea Valley. The school moved into a purpose-built campus in 1996. The school has served between 1000 and 1300 students, ages 11-16, each year. The new school had been constructed on a pleasant piece of land, near a community recreation center. Approximately 19% of Cwmtawe's students each year qualify for free meals, again above the national average, but not at Sandfields' level of deprivation.

In the three years pre-HRS (1994-1996), an average of 31% of Cwmtawe's 16 year olds achieved 5+ A\*-C GCSEs. This was slightly below the district average of 34.5%, and well below the Welsh national average of 40.7%. Given the community's higher-than-average level of disadvantage, these numbers were not considered by the community to be deeply troubling.

### *Timeframe*

The Welsh HRS effort began in the fall/winter of 1996/7. Four schools initially signed on to the project, and by the early spring of 1997 they had been joined by all of the 11 secondary schools in the Neath-Port

---

<sup>9</sup> In the U.S., this would be an Interstate highway.

Talbot (NPT) school district. The initial plan was for the three authors to provide detailed professional development activities related to the 12 HRO characteristics over a period of three years. The total number of workshops and follow-up activities was to be approximately a dozen, though the specifics had not been worked out in detail at the project's beginning. This lack of initial clarity reflected, in part, the schools' desire to "get the show on the road" right away. At the same time, it reflected the conviction among the authors that the reform would either be co-constructed or would not be sustained (see Berman & McLaughlin (1978) on "mutual adaptation" and Datnow & Stringfield (2000) on active co-construction.) We assumed from the beginning that some of the HRO principles would require more explanation/professional development than we could prospectively plan, and others would require less. This presumption proved accurate. The formal set of trainings by the authors was, none the less, completed by the early spring of 2000.

By early 1997 an unanticipated component of the HRS project evolved. Several of the school Headteachers began holding monthly meetings to work through "the little things that matter" in implementing the HRO principles. These meetings were eventually endorsed by their LA, and often included all 11 secondary Headteachers. While the "formal" trainings for Headteachers and teachers ended in early 2000, these important, informal peer-problem-solving sessions continued for several additional years. To some extent, the bonds forged in the formal trainings and informal gatherings have continued to this day, and the schools' leadership teams are often unusually open in seeking shared solutions to schools' challenges.

#### *Data Gathering*

**Field Notes.** Beginning with our initial planning and workshops with educators, we endeavored to keep detailed notes on the processes and decisions being made by the three authors and by the schools. These notes have been repeatedly reviewed and enriched with recollections from both the researchers and groups of teachers and school Headteachers.

**Workshop Materials.** Copies of workshop materials were saved and modified, based on feedback from participating educators. These became part of the complex data bases of the research component of the project.

**Detailed Case Studies.** In 2000, at the end of the formal involvement of the authors in the HRS intervention, a series of detailed case studies were conducted on each of the participating schools. Data gathering included interviews with Headteachers, other key administrators, and teachers. Archival data were gathered at most sites. Data gathering and report writing was conducted by a separate researcher, paid through a small grant. The reports followed the HRS characteristics described above, and ranged from 10-20 pages per school, plus appendices.

One-half-to-one day follow up site visits have been conducted at many of the sites. The follow up data gathering at the two schools described in this report happened during 2005, 2010, and in January of 2012. Each follow up visit was conducted by between one and all three of the authors, and included interviews with Headteachers, teachers, and, at some schools, classroom observations.

**HMI Reports.** All British schools receive a detailed inspection once every five years from Her Majesty's Inspectors (HMIs). The inspection team typically includes a half-dozen or more members, drawn from

both the national government and outside educators. The goals of the inspections are detailed at the official web page of the inspectors (<http://www.ofsted.gov.uk/schools/for-parents-and-carers/understanding-school-inspection-report>), and include providing a picture of each school's current performance together with making suggestions for further improvements, presenting an unbiased picture of students' academic and social progress, gathering and presenting data on how parents and students perceive the schools, providing an assessment of the physical condition of school buildings, noting the processes through which school personnel deal with problems and complaints, and the extent to which schools comply with local and national rules and regulations. The inspection results in a detailed, public report. Report results are taken seriously by the schools' teachers and leaders, by their LA, and their board. For the five years after an inspection, all schools' web pages include copies of, among other things, their most recent inspection reports. As with GCSE reports, a negative inspection report can have severe consequences in terms of numbers of parents choosing to send their children to a school, and the longevity of a Headteacher. Hence, the inspection process is taken quite seriously. In Wales these inspections are organized by a body entitled ESTYN and in England by one called OFSTED.

### 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" (students' age 17-18). The standard examinations for the end of secondary are the General Certificate of Secondary Education tests, or simply the GCSEs<sup>1</sup>. Virtually every British student sits for the GCSEs. The traditional measure of strong academic performance for a student is obtaining "5 or more A\* to C" grades on the various examinations (literature, math, various sciences, etc.). Under half of Welsh students obtained 5 or more A\*-C grades in the mid 1990's administrations of the GCSEs, and those percentages have risen to over half in the first decade of the 2000's. The tests are viewed as relatively high stakes assessments for both students and their schools. These examinations are specific to the course content taught during the years, making them somewhat similar to Advance Placement courses in the U.S. For students, a certain number of passing grades are necessary 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 5+ 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 program that is relatively high stakes for students and educators alike.

One condition for schools' participation in the HRS programme was an agreement that the percentages of students achieving 5+ A\*-C scores would be the primary outcome measure for each school and the overall programme.

### **The High Reliability Schools (HRS) Project in Neath-Port Talbot**

Having participated in several teacher- and school effects studies (Reynolds, 1987; Schaffer, Stringfield,

& Wolfe, 1992; Stringfield et al., 1994, 1997; Teddlie & Stringfield, 1993), the authors were aware of over a half-century's studies indicating that school reform efforts—and particularly very ambitious high school reform efforts—had typically foundered for lack of consistent implementation. Stringfield (1995) had hypothesized that in educational reform efforts, as throughout the social sciences, reliability sets the upper boundary of school reform validity. Reynolds, Schaffer, and Stringfield were enthusiastic about the possibility of using research on HROs to work with schools and local educational organizations to co-construct long-term secondary school reform.

The over-arching agreement between the Welsh schools and the research team was that we would do our best, within limits of time and resources, to bring research results in various “effectiveness” areas to the schools and district for discussion and implementation. Operating as full partners in the development and implementation process, the schools would co-construct the actual reform components at every stage of the reform. Our agreement was built on the oft-stated assumption that the researchers knew and would share substantial amounts of information on teacher-, school-, and program-effects and how those had been implemented at various sites. At the same time, we were explicit that they knew their schools at a much greater level of detail than we ever would. The important next sentence was always that, “The world’s leading authorities on your schools, their strengths, weaknesses, and how best to improve them are you, your schools’ and LA’s professional educators.” In stating this we were simply acknowledging that however often we provided workshops at schools, or the LA staff visited, the people who were in the schools six to ten hours a day were the teachers and administrators of the schools. They would always know more of the day-to-day realities of the students, staff, and curriculum than would three college professors or, for that matter, employees of the LA. The assumption from the beginning was one of goal-focused partnership. Because honest, accurate feedback was necessary both between the presenters and schools and among the schools and the LA, we all worked hard to create a shared trust that any criticisms would be delivered in a constructive way and, whenever possible, would be paired with compliments on things that were going well.

The NPT High Reliability School agreement among the schools and the researchers, as related to the 12 HRO principles was as follows:

1. Organizational reliability evolves under a particular circumstance. HROs evolve when both the larger society and the professionals involved in the working of the organization come to believe that failure of the organization to achieve its key goals would be disastrous. All participants took as a given that the economic and social impacts of education had changed dramatically over the previous generation, and that the rate of change was only likely to accelerate. In Britain, as in the U.S., the differential in incomes between people who had succeeded in school and those who had not were expanding rapidly, and have continued to do so. One implication of these rapid changes is that success in school is much more important today, both for individuals and for the health of the larger society, than in previous generations. For example, in the U.S., the average differential incomes for young adult males who had dropped out of high school vs. those who had graduated from college was 30% in 1949, and over 150% by 2000. Those and similar British data were presented to faculties at all schools.

2. HROs require a clear and finite set of goals, shared at all organizational levels. All of the schools agreed to focus on 2-4 very ambitious goals. One goal had to be a substantial 5-year rise in the percentage of students obtaining 5 or more A\* to C's on the GCSEs. A second had to be improved attendance. Each school was free to choose up to two additional HRS goals. The choices among additional goals often produced lively discussions within faculties.

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 R's": relevance to core goals; rich triangulation on key dimensions; real-time availability to all organizational levels; and rich triangulation on key dimensions. 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.

One almost immediate benefit of acquiring the additional data from the university-developed system was the introduction of testing of all incoming students. 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 2 years behind in reading, an effort was launched to coordinate the secondary school's literacy programs with those of the feeder primary schools.

Beyond that formal, externally-developed system, individual schools began building longitudinal student data bases. Those have continued evolving for 15 years and today are often very sophisticated.

As an example, Cwmtawe keeps elaborate data sets on all students and cohorts at their school. Every student's progress is reviewed by a multi-person team a minimum of three times per year. The focus is on making sure that any problems are identified and addressed as early as possible, and that no child to get accidentally left behind.

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). We were pleasantly surprised by the enthusiasm with which teachers at many schools embraced the development of standard solutions to widely-experienced problems. Teachers and Headteachers often saw these as time/energy-saving steps.

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 by a regular focus upon what came to be called "the little things that matter." Regular HRS meetings increasingly centred upon 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. When an effort had succeeded, it typically was tried at other schools. When one failed, other schools typically provided both encouragements to keep trying and tips on why the effort had succeeded at the other schools.

5. Highly reliable organizations actively sustain initiatives that encourage all concerned to identify flaws in standard operating procedures, and honor the flaw finders. Schools SOPs continue evolving, testament to their willingness to consider and act upon suggestions for improvements.

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, extensive recruiting of new staff at all levels. Schools reported an enhanced focus on recruiting new teachers, and especially new Headteachers. The typical practice in Great Britain has been that when a headship comes open, an advertisement is placed in the Times Education Supplement and the school's board and faculty interview and choose a person who seems well matched to the school. As in the U.S., this often produces a bouncing from one Headteacher's priorities and biases to another in a disorderly fashion. In NPT, schools began choosing new heads from the deputy heads in the LA who had shown a particular passion for moving the HRS programme forward. The result was a substantial increase in continuity of goals and SOSs.
7. Constant, targeted training and retraining. Additional time for professional development was built in. Headteachers and faculties attended regionally based residential sessions (two-day meetings at a conference centre), and also added national residential sessions, all aimed at enhancing knowledge transfers across schools. The Professional Development focus tilted strongly 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 schools.

For three years, all schools focused their professional development activities on one reform. As an example of the HRS professional development, teachers received training on how to conduct low- and high-inference classroom observations on dimensions identified as improving teacher effectiveness (Good & Brophy, 1986, 2009), and to providing accurate yet supportive/"no fault" feedback after observations. Then demographically similar but high achieving schools were identified in specific areas and teachers were sent to observe and to provide feedback (ex., math teachers observing math teachers at a high achieving school). This produced rich discussions among teachers upon their return to their home schools or departments. As a result of this substantial emphasis on within- and between-school targeted intervisitations and observations, teachers became accustomed to observing one another's lessons, and levels of open discussion and trust between teachers within departments, within schools, and within the LA rose significantly.

A strong "departmental effectiveness" component evolved as the project developed. This facilitated among-teachers, within-school learning.

As a third example of this key issue, all of the schools and departments within schools agreed to share successes and failures, and thus created learning communities within and across schools. Each school and department committed to studying “best practice,” both from the international research bases and within and without the HRS schools. Every time a school or department learned something new and valuable from another school or department, either through the second school sharing a success or a failure to be avoided, collective trust rose.

In years two and three, 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. Second, 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. Headteachers also selected among possible staff development alternatives based on their perceived needs of faculty and previous efforts that had been successful. On occasion, schools requested additional training on topics that did not appear to have yet been successfully implemented in a school.

8. Rigorous, multi-way performance evaluation. Through the HRS project, heads and deputy heads also spent much more time in classrooms, teachers provided much more feedback to administrators, and schools have evolved richer sets of relationships with parents. Students, parents, teachers, and administrators all became more aware of the specifics of their responsibilities in improving the academic achievements of all students.

Further:

9. Key equipment is kept in high working order. Sandfields began their HRS involvement with a general clean-up-fix-up campaign. Other schools focused on ensuring that overhead projectors, computers and such were kept in a higher state of order. In the years of follow-up school visits, the authors have often been struck with how smoothly and efficiently various aspects of schools now operate.
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 behavior that emphasizes collegial decision-making, regardless of the formal position of the decision-maker. In the two English groups of schools, deputy heads had been tasked with managing the HRS work. This had the unintended effect of signaling to teachers that HRS wasn't critically important. Seeing this, the Welsh Headteachers immediately declared that they would lead their schools' efforts. Formal organizational hierarchy had a positive effect. Beyond that, authority was widely spread, depending on the tasks. For example, at Cwmtawe, there is no question who is the Headteacher, yet to visit the school is to observe a wide range of administrators, teachers, and students in leadership positions. As an example of the later, in 2011 the school launched a “School Nutrition Action Group,” led by students. The group got rid of all snacks that included salt and “fizzy drinks” (in the

U.S., sodas). The same group initiated “walking Wednesdays” in which all students and faculty get additional, moderate exercise. The initiatives were from students, and hence weren’t viewed as a top-down fun-removal, but rather as a part of a whole-school pride-in-health initiative.

11. Clear, regularly demonstrated valuing of the organization by its supervising and surrounding organizations. All levels work to maintain active, respectful communication geared to the key goals of the HRO. The Neath-Port Talbott LA supported the HRS initiative throughout, and continued supporting it for several years after the formal programme ended. A part-time “HRS Driver” was 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.

The district has gotten a good deal of positive attention as a result of their students’ progress, and that may have helped district leaders remain loyal to the effort, but regardless of motivation, their support has been significant over the years. This is not to say that they haven’t gotten pressure from the national assembly to initiate new initiatives on a virtually annual basis, but they have allowed schools to retain a focus.

12. Short-term efficiency takes a back seat to very high reliability. The LA attempted throughout the programme and in the follow-up years to sustain levels of funding that would permit HRS to continue. There have been years when funding was cut, but relative to other public services, funding cuts to achieve short-term efficiencies have been minimal—and much less than the cuts to several other social services— in NPT.

## Results

Two types of outcome data are presented in this section. For the LA and each of the schools, longitudinal GCSE data will be presented, and contrasted with Welsh national results. In addition, for both Sandfields and Cwmtawe, data from their most recent HMI inspection will be provided.

For the district, aggregated GCSE data will be the primary outcome variable.

### *Sandfields Secondary School*

From 1994 to 1996, Sandfields Secondary was both one of the highest poverty and one of the lowest achieving secondary schools in Great Britain. With over 40% of Welsh 16 year olds scoring well on national examinations, Sandfields was averaging under 14% (see Table 1). This percentage was substantially below the 20% rate that was regarded as the absolute minimum acceptable rate nationwide. In 1996, the once-every-five-years national inspection report by HMIs recommended considering closing Sandfields if large improvements were not made.

In the fall of 1996 Sandfields became one of the original four Welsh secondaries to join the HRS project.

Over the next four years, Sandfields more than doubled its students' success rates on the GCSE national tests to 35% in 2000. In 2000, Sandfields was recognized by the Welsh Secondary Schools Association (WSSA) as the most improved school in Wales<sup>10</sup>. Continuing to use HRS principles through an additional change in school Headteachers, Sandfields rose to 51% in 2005. In 2011 fully 54% of Sandfields' age 16 students obtained 5+ A\*-C scores on the various GCSE examination. As can be seen in Table 1, this rate of gain in student achievement is nearly double the national average (30% vs. 16.1%). Over the 15 years since the beginning of the HRS project, this rise has meant that literally hundreds of high-poverty students have graduated with at least a shot at going on to college and/or landing and succeeding in 21<sup>st</sup> century jobs.

Table 1

*Longitudinal Percentages of Students with 5+ A\*-C scores on the General Certificate of Secondary Education (GCSE) Age 16 Tests*

Time Frame/ Group	1994-1996 (pre-)	2000 (post-)	<i>Initial Gain</i> (pre- to 2000)	2005 (5-yr. follow-up)	2011 (11-yr. follow-up)	<i>Longitudinal Gain</i> (pre- to 2011)
NPT LA	34.5%	49.4%	+15.9%	57.6%	75%	40.5%
Sandfields	14%	35%	+21%	51%	54%	30%
Cwmtawe	31%	51%	+20%	72%	93%	62%
Wales	40.7%	49.0%	+8.3%	52.2%	56.8%	16.1%

Sandfields' 2011 HMI's report gave the school an over-all "excellent" (highest) rating for both current status and "prospects for improvement" (Estyn, 2011). Across the current HMI categories, Sandfields had the very unusual experience of obtaining ratings of "excellent" in every major category: student outcomes, student wellbeing, and provision of services to students, including teaching and support and guidance, leadership and management, partnerships, and resource management. The nine-member HMI team could hardly have been more positive about the school's operation or its results.

In a January 2012 interview, the Headteacher stated that the fundamental starting point for Sandfields' turnaround, and a continuing source of guidance, was the HRS experiences. He expressed particular pride in the very sophisticated, relational data base the school had evolved over the past 15 years and in the quality of his teachers. A wide range of student-specific data, ranging from daily attendance to frequencies of discipline issues, are available to every teacher in the school from their desktops.

<sup>10</sup> The school coming second was another NPT secondary.

### *Cwmtawe Secondary School*

During the three years pre-HRS, from 1994 to 1996, an average of 31% of Cwmtawe Comprehensive School's students had scored well on the national GCSE examinations. As can be seen in Table 1, this was somewhat below the Welsh national average of the time. In the four years from pre- to post-HRS's formal programme, Cwmtawe rose to having 51% of its students obtaining 5+ A\*-C GCSEs. This 20 percentile gain was more than double the nation's 8.3% rise and moderately greater than the over-all district's 15.9% improvement. Five years post-intervention (2005), fully 72% of Cwmtawe's students were obtaining 5+ A\*-C's. By this point, Cwmtawe, like Sandfields before it, had been recognized as one of the nation's most improved secondary schools. In 2011, a stunning 93% of Cwmtawe's students achieved at that level. The school's students' gains in levels of success have been so great, and the school leaders' belief in data-driven success is so high that simply tracking the numbers of students with 5+ A\*-C's no longer created high enough goals. The school now tracks the numbers of students that achieve 10+ A\*-C scores, an accomplishment virtually unimagined for any student 15 years earlier. For 2011, 52 students reached that very high plateau, including three students who achieved at least 11 A\*s and A's. Not only are Cwmtawe's 5+ A\*-C scores—for a higher than average free lunch school—highest in the LA, they are, by 14%, the highest of any school in their economic band in all of Wales.

Although official national data are not gathered on the percentages of students obtaining 10+ A\*-C scores, Cwmtawe is almost certainly a national leader in that new category. At the other end of the scale, recognizing the full range of students' abilities, fully 98% of Cwmtawe's students obtain 5+ A\*-G scores, meaning that almost all of their students graduate with some level of legitimate credentials. Summarizing the schools GCSE data: first, the school has made remarkable strides in the initial goal of increasing the percentages of students achieving 5+ A\*-C scores. Second, having made those strides, the school set a higher target of increasing the numbers of students achieving 10+ A\*-Cs, and has virtually created a new, demanding statistical category, one in which this higher-than-average-poverty school's students are now excelling. Third, not content just to "raise the ceiling to new heights," Cwmtawe has consistently shown concern for the less able students in the community, and has raised the percentage of students achieving at least some level of academic success across at least five areas to 98%. Any one of these achievements would be highly laudable. Together in one community school, they are remarkable.

Cwmtawe's most recent HMI inspection was in 2006 (Estyn, 2006). The group's summary statement was that, "Cwmtawe Comprehensive School is a good school with many outstanding features" (p.2). That was the highest rating category available at that time. The HMIs noted that the school had a culture of continuous improvement combined with high expectations; that the students had "mature and responsible" attitudes (p. 2); that the school had very good standards of achievement, teaching and learning; an excellent system of monitoring student progress; and "outstanding leadership" both in the Headteacher and the school Management Team.

Members of the research team were able to conduct follow-up interviews of Cwmtawe's Headteacher, leadership team, and groups of students in both 2007 and January of 2012. What was most impressive about the interviews with school leaders was their depth of knowledge about the detailed functioning of their schools. Using HRO and shared teaming principles, the leadership team had created systems that work very well at Cwmtawe for identifying the skills and carefully tracking the academic progress of every child in this large school. Every child has high academic challenges and at least one school

employee monitoring their progress. Students who aren't making progress over their first few years at Cwmtawe are assigned an individual tutor. That tutor may be the Headteacher, who every year is assigned a small group of students with whom to work, a classroom teacher or a peer tutor, depending on the leadership's informed decision as to what will work best for each individual student.

This elaborate system of tracking each student's progress is made possible through a thoughtfully evolved data system that the school has been developed and refined over the HRS and post-HRS years. Each student is tested on entry to the school, and an elaborate system of follow-up data is gathered on each student's progress. Results are color-coded, resulting in printouts that are simultaneously elegant and easy to quickly absorb on both the individual and group levels. The data base allows for rapid examination of individual departments' progress (math, biology, languages, vocational preparation, etc.) in assisting all students, while also monitoring and suggesting targeted interventions for individual students. At the macro level, the data base has become an excellent system for monitoring the school's progress, but its core strength lies in its ability to help faculty, students, and parents see—and, clearly, achieve—individually demanding goals.

As part of the above-mentioned follow-up site visits, the team conducted interviews with groups of students in both 2007 and January 2012. The students knew exactly their personal academic goals, the benchmarks associated with those goals, and the work they needed to do to achieve those goals. As one example, a group of three young ladies were being team-tutored by the Headteacher. They were students who, based on their age 11 test scores would not have been predicted to do well on the GCSEs four and a half years later. However, the Headteacher had worked with the girls on everything from homework skills to general motivation, and all eventually cleared the 5+ A\*-C bar. Their pride and emerging self-confidence was palpable.

For the researchers, even more impressive than the students' academic accomplishments were their levels of mature academic confidence and engagement in a broad range of pro-social activities. Two examples follow: In order to raise funds to support various school activities, the senior students had, for several years, offered one week per year during which younger students could bid for the right of having an older student help with everything from carrying books between classes to providing academic assistance on coursework. The seniors were proud of the prices their collective services were commanding, and more proud of the materials and service their auctioning of their efforts brought in each year.

As a second example, a student group had been formed, and supported by both teachers and canteen management, as the "School Nutrition Action Group (SNAG). The SNAG had examined both school lunch menus and snacks that were available for purchase in the school, and made a series of recommendations for healthier food and snacks. The school implemented the SNAG's recommendations, with the effect that students took ownership and pride in the fact that they were all eating healthier foods.

Cwmtawe's rise has been so dramatic, and has been sustained for so many years, that the school receives a regular flow of visitors from schools across Wales. The visitors are trying to learn and copy Cwmtawe's "secret." In our January 2012 visit, the deputy head who manages the school's now-highly-sophisticated student tracking and goal-setting system, reported with some frustration that visiting educators often latch onto Cwmtawe's data warehouse as though that is "the answer." He firmly asserted that the data base was only one part of a much larger system, evolved over years, and that focusing on any one part, however interesting, would not produce Cwmtawe-like results. Rather, what Cwmtawe has

evolved over 15 years is an integrated systems approach to educating a community's schools, and does so with virtually unprecedented levels of reliability.

### *The Neath-Port Talbot District*

As can be seen in Table 1, pre-HRS, NPT's secondary school students were achieving at levels moderately below the Welsh national levels (34.5% vs. 40.7%). However, by the end of the formal HRS programme, the district's age 16 students were achieving at the national average (49.4% vs. 49.0%). The 11 secondary school district's gains had nearly doubled the nation's impressive 8.3% gains. At that point, the HRS intervention team ended its formal involvement in NPT.

However, with the support of the LA, the Headteachers continued meeting and solving individual and group problems, using the HRS model and structure. As can be seen in the remainder of Table 1, by 2005, the NPT students had, on average, moved substantially beyond the Welsh average (57.6% vs. 52.2%). By 2011, the student in this relatively high poverty district was scoring fully 18.2 percentile points above the Welsh national average. So substantial and so sustained have been NPT's gains, that in late 2011 the Welsh government took the unprecedented step of asking NPT's administration to effectively take over and operate another LA that was experiencing much less success.

## **Discussion**

Sandfields, Cwmtawe, and the Neath-Port Talbot district all began their participation in the High Reliability Schools project at a time when the schools and district had been in a sustained period of stable, below-national-average achievement results. Given that the communities served by the NPT district were—and remain—relatively economically challenged, the lack of specific push for educational improvement from the community had not been atypically. However, several of NPT's Headteachers had tired of seeing reform effort after effort come and go with no noticeable effects, and latched onto the HRS project. They were particularly attracted to the focus on a few measurable goals, the uses of “effectiveness” research and data richness, and to the idea of locally co-constructing the reform effort.

The participant-research-team members insisted on holding to no more than four goals per school, with one of the goals being very high scores on the national GCSEs. The team helped schools begin developing much more timely and teacher-useful student data bases, and trained teachers in classroom observation systems for use both in self-assessment and in intra- and inter-school classroom visitations. As full partners and “co-constructors” of the HRS project, Headteachers, deputies, and hundreds of teachers examined research, their own practices, and “best practices” from across their district and across Great Britain. The four-year intervention student achievement data indicated that the HRS project had been a success throughout the district. Stringfield, Reynolds & Schaffer (2008) reported that each of the 11 secondary schools in NPT had made four-year GCSE gains that exceeded the national average, and the district as a whole had substantially exceeded national levels of gain.

Those initial gains were particularly impressive when considered in light of a range of studies indicating that great variance in levels of reform implementation should be regarded as almost a rule (Nunnery, 1998; Supovitz & Weinbaum, 2008). Arguably more impressive have been the schools' and the local authority's subsequent gains. Five and 11 years post-intervention, the schools and the district have not

only held onto their initial progress, they have continued making dramatic gains. Not only had the HRS's local leaders made gains, they had learned how to create continuous progress at unprecedented levels and over an unprecedented time frame.

### *Implications for Practice*

The first implication for practice is that, in contrast to much of the frustration and skepticism regarding turning around historically low achieving secondary schools, long-term improvement is clearly possible. This is a very optimism-justifying conclusion that HRS clearly demonstrated.

In the case of the High Reliability Schools effort, a second implication is that what was needed for virtually unprecedented levels of student success at Sandfields, Cwmtawe and throughout the Neath-Port Talbot LA, was not so much any one specific reform idea, but an overarching reform logic. All ideas for improvement were funneled through a system of goal checks, then implementation commitments and checks, feedback from within and among schools, and monitored carefully through gradually evolving data systems. Some of those data systems are now remarkably sophisticated and intuitive. They are available to teachers and administrators in real time, providing data on everything from attendance histories to projected/targeted achievement levels.

Returning to the core thesis, no one component was or is “the key.” Have the schools evolved systems for using current teacher- and school-effects research? Yes. Elegantly evolved data bases for used in decisions at all levels? Yes. Greatly improved systems of pastoral care for all students? Clearly. More involvement in sports? Healthier eating? Yes, yes, and more. However, no one or group of those has been the guiding or the driving force in these schools. Rather, creating a highly reliable system for thinking and acting so that each student, teacher, and leader can use other innovations to achieve continuous improvement has been “the key.”

Third, while the research team helped the schools make productive use of well established research bases on teacher- and school-effects, school improvement, classroom observation techniques, and the need for efficient, teacher-usable data bases for decision making, the HRS project demonstrated the critical importance of local educators as “the world’s leading authorities on their own schools.” “Co-construction” was much more than a soon-to-be-trendy catch phrase in HRS, it was a cornerstone.

We believe that among the conditions necessary for HRSs to have evolved at Sandfields, Cwmtawe, and across Neath-Port Talbot were the following:

- a. As a first condition, professional educators, and, eventually the public, must realize that in the 21<sup>st</sup> century, the costs of educational failure are catastrophic both for the individual students who do not achieve their full potential and for all the rest of us in our society. This is a dramatic shift from much-discussed “sorting” function of schools from 50 years ago.
- b. Firm, committed buy-in from both the LA and the schools’ leadership to a focused set of goals was critical. These schools eschewed familiar “Christmas tree” habits of adding new programs whenever they were offered. Rather, schools focused on achieving 2-4 core goals, one of which was dramatically improved measures of academic progress. In one sense, these first two were reforms, “before the beginning,” and they were non-negotiable.
- c. A perception was created early on and sustained, that failure to achieve core goals was not

- only not inevitable but was unacceptable.
- d. A minimal level of leadership stability, together with carefully orchestrated leadership transitions was necessary, and it sustained organizational reliability.
  - e. The HRO characteristics described earlier in this paper were followed in detail.
  - f. As a result, school leaders, including leaders among the teachers, not only made 4-year improvements in their schools' functioning, they learned how to learn to make continuous improvements that have lasted 11 years post-initial HRS training.

In the introduction to this paper we briefly noted that a set of schools in an English district had joined the HRS project and produced no 4-year gains at all. While they are not the subject of this paper, we believe that clear contrasts are available, and we provide them here. We believe that viewed from an HRS perspective, the conditions that predicted educational reform failure are also knowable. They include, but are not limited to the following:

1. A lack of initial buy-in to the idea that dramatic improvement in student outcomes is possible.
2. Attempting to attend to and balance many diverse goals. Holding 10 “core” goals has the same effect as holding none.
3. Attempting to implement reforms that are not clearly informed by reasonably rigorous research.
4. Lack of multi-year commitment to targeted, sustained, shared professional development.
5. A failure to understand—and be open to—the idea that any reform, including HRO-based reform, is a combination of external ideas and continuously evolving local contexts. Just as there are no two air traffic control towers that are alike, there are no two schools needing exactly the same reforms, the same data to inform future actions, the same Standard Operating Procedures, etc. Further, any one school's need for any one SOP may change over time. Dynamic organizations must be improved dynamically.
6. Leadership- and staff-instability, especially if not accompanied with careful, real-time induction into HRS principles, are prescriptions for failure.

Our two over-arching conclusions follow. The first is that the conditions now exist in which substantially higher educational reliability is possible. Highly reliable systems evolve when the costs of failure become too high in the eyes of the public and the professionals working in an organization. In 21<sup>st</sup> century education, the costs of failure—both for the individual and the society—have become too great for unreliable educational practices to continue. Hence, we believe that countries' fundamental choice is not whether to become more reliable, but whether to stumble forward, feeling our way and making many, many mistakes, or to understand and control more efficient processes of increasing educational reliability.

The second conclusion allows for a good deal of optimism. Our data indicate that, under specifiable conditions, High Reliability Organization principles can be productively applied in school and district contexts. Thoughtfully, consistently applying HRS principles has produced dramatic results in Neath-Port Talbot, and could do so again in other schools and districts.

At first brush, some may worry that HRS models have a “mechanistic” feel to them and that they appear to inhabit a “check-list” or “tick-box” mentality. Nothing could be further from the truth.

In many schools and indeed countries, schools and systems are “tight” on a few processes meant to affect achievement and “loose” on the systems to achieve these processes. HRS is the opposite—loose on the precise organizational processes needed, leaving those to be determined in detail by schools. Where HRS has created more tightly coupled system components is on specifying the concepts and systems that schools should use to generate their often different processes, and evolving mechanisms for monitoring the effects of improvement efforts. This has allowed for the standardization of procedures that have been shown to produce successes, and the gradual weeding out of less successful processes.

### *Implications for Future Research*

We believe that the successes in the High Reliability Schools project have clear implications for future research.

First, we believe that researchers, like practitioners, need to attend much more carefully to the initial, pre-intervention conditions that exist in schools. The leaders of the schools and district described in this paper were focused, and bought-in, in advance of starting the effort. They understood that there was to be no “silver bullet” or “miracle of the month.” Initial conditions matter, and are rarely reported in research reports.

Second, it may be the case that exactly which improvement-focused reform is implemented—from a range of potentially valid options—is not as important as whether the reform is implemented reliably. The HRS project never mandated that a school implement any one of the research-based improvement ideas that the team forwarded. What the team insisted on was that any intervention component that a school chose to implement, be implemented reliably. Some schools worked much harder on developing relational data bases for teacher and Headteacher use than others; some conducted many more cross-classroom, cross-departmental, and cross-school observations than others, etc. The things that were always critical were intentionality and a focus on reliability of implementation of reform components, once chosen by individual schools<sup>11</sup>.

Third, getting reform to scale has proven extremely challenging (Fullan, 2010; Marzano & Waters, 2009; Payne, 2008; Zavadosky, 2009). Yet the HRS project demonstrates that secondary school reform is possible, at least at an 11-secondary school scale. The gains of the two schools profiled in this paper were clearly reflected in the larger Neath-Port Talbott district, and in the larger Newcastle English District HRS effort described elsewhere (Stringfield, Reynolds, & Schaffer, 2010). The Earl, Torrance and Sutherland (2006) observation that “[T]here are no examples anywhere of successful *whole district high-school reform*.” (p. 126), is no longer accurate. More studies of successful efforts at scale are clearly needed. This study suggests that HRO can provide a useful framework for analyzing why reforms are/are not succeeding at scale.

---

<sup>11</sup> Matt Miles (1986, personal communication) once observed that the most likely route for a school to take to successful reform was to “pick a reform and go at it hard.” We believe that the act by a school and a Headteacher of picking a reform to go at hard is a significant part of the reform, happening “before the beginning.”

Related to issues of scale is the issue of loose/tight coupling, foreshadowed by Weick (1979). Clearly, the HRS teachers and schools have been more tightly coupled along key dimensions. They share finite numbers of goals across classrooms, departments and schools, gather and use real-time data much more extensively and productively, discuss successes and failures across classrooms and schools more, and share professional development activities. However, Weick observed that (a) no system can be tightly coupled on all dimensions and that (b) loose coupling offers many important advantages. Which dimensions are key for which levels of coupling, and why? From HRS, we believe that goals and accessible, relevant data should be toward the top of the list of candidates for tighter coupling, along with increases in well-prepared, targeted classroom intervisitations. Others deserve study beyond the reach of our case studies of schools and districts.

Fourth, there are not nearly enough long-term studies of school reform effects. While the four-year effects of HRS were valuable, the continuing follow-ups have added immeasurably to the significance of this effort. As the Borman et al. (2002) meta-analysis demonstrated, there are too few follow-up studies in education. There is too little evidence of what abides, what does not, and why? Datnow (2005) and Hargreaves and Fink (2006) have suggested initial characteristics of sustainable changes. Surely there are others.

In this paper we have provided clear evidence of secondary school “turn around,” of taking reform to scale, and of the High Reliability Schools’ core components being sustained 15 years post-reform implementation and 11 years post-intervention. Replications of this success are needed in the UK and here in the US.

---

i The GCSEs replaced the former Ordinary, or “O-level” examinations. The age 18 exams are still the Advanced, or “A-level” exams.

---

## References

- Aiken, W. M. (1942). *The story of The Eight Year Study*. New York: Harper.
- Bellamy, G.T. (2011). *High Reliability leadership for educational change*. Noteworthy Perspectives. Denver, CO: McREL.
- Berman, P., & McLaughlin, M. (1978). *Federal programs supporting educational change. Vol. VIII, Implementing and sustaining innovations*. Santa Monica, CA: Rand.
- Borman, G.D., Hewes, G.M., Overman, L.T., & Brown, S. (2003). Comprehensive School Reform and achievement: A meta-analysis. *Review of Educational Research*, 73, 125-230.
- Carter, L. (1984). The sustaining effects study of compensatory and elementary education. *Educational Researcher*, 13 (4), 4-13.
- Cuban, L. (1992). What happens to reforms that last? The case of the junior high School. *American Educational Research Journal*, 29(2), 227-251.
- Datnow, A. (2005). The sustainability of comprehensive school reform models in changing district and state contexts. *Educational Administration Quarterly*, 41(1), 121-143.
- Datnow, A., Borman, G., Stringfield, S., Rachuba, L., & Castellano, M. (2003). Comprehensive school reform in culturally and linguistically diverse contexts: Implementation and outcomes from a four-year study. *Educational Evaluation and Policy Analysis*, 25(2), 143-170.
- 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. H. Chrispeels (eds.), *Improving schools and educational systems: International perspectives*. London: Routledge.
- Friedman, T.L., & Mandelbaum, M. (2011). *That used to be us: How America fell behind in the world it invented, and how we can come back*. New York: Farrar, Straus and Giroux.
- Glassman, R.B. (1973). Persistence and loose coupling in living systems. *Behavioral Science*, 18, 83-98.
- Hargreaves, A., & Fink, D. (2006). *Sustainable leadership*. San Francisco: Jossey-Bass.
- Hopkins, D., Stringfield, S. Harris, A., Stoll, L., & Mackay, T., (in process). School and system improvement: A state of the art review.
- 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.
- Levy, F., & Murnane, R.J. (1993). U.S. earnings levels and earnings inequality: A review of recent trends and proposed explanations. *Journal of Economic Literature*, 30, 1333-1381.
- McLaughlin, M. (1990). The Rand Change Agent Study revisited: Macro perspectives and micro realities. *Educational Researcher*, 19 (9), 11-16.

- 
- Muncey, D.E., & McQuillan, P.J. (1996). *Reform and resistance in schools and classrooms*. New Haven: Yale.
- Nunnery, J. A. (1998). Reform ideology and the locus of development problem in educational restructuring. *Education and Urban Society*, 30 (3), 277-295.
- Pfeiffer, J. (1989). The secret of life at the limits: Cogs become big wheels. *Smithsonian*, 20 (4), 38-48.
- Reich, R.B. (2011). *Aftershock: The next economy and America's future*. New York: Random House.
- Reynolds, D. (1987). *The comprehensive experiment*. London: Falmer.
- Reynolds, D., Sammons, P. DeFraine, B., Townsend, T., Van Damme, J., & Stringfield, S. (submitted). Educational effectiveness research: A state of the art review.
- Reynolds, D., & Stringfield, S. (1994, January). *The International School Effectiveness Research Programme: Case studies from low socioeconomic status, positive outlier schools in eight nations*. Symposium presented at the meeting of the International Congress for School Effectiveness and Improvement. Melbourne, Australia.
- 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. (n.d.). *Bishop Rock dead ahead: The grounding of the U.S.S. Enterprise*. Berkeley, CA: Haas School of Business.
- Roberts, K. (1993). *New challenges to understanding organizations*. New York, NY: Macmillan.
- Roberts, K., & Bea, R. (2001). Must accidents happen? Lessons from High Reliability Organizations. *The Academy of Management Executive*, 15(3), 70-79.
- Rochlin, G., LaPorte, T., & Roberts, K. (1987). The Self-Designing High-Reliability Organization: Aircraft Carrier Flight Operations at Sea. *The CEO Refresher*. Downloaded from [http://www.caso-db.uvek.admin.ch/Documents/MTh\\_SM\\_22.pdf](http://www.caso-db.uvek.admin.ch/Documents/MTh_SM_22.pdf)
- 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., Stringfield, S., & Wolfe, D. (1992). Two-year effects of a sustained beginning teacher induction program on classroom interactions. *Journal of Teacher Education*, 43(3), 203-214.
- 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., 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). *Urban and suburban/rural special strategies for educating disadvantaged children. Final report*. Washington, DC: U.S. Department of Education.

---

Stringfield, S., Reynolds, D., & Schaffer, E. (2008). Improving secondary students' academic achievement through a focus on reform reliability: 4- and 9-year findings from the High Reliability Schools project. *School Effectiveness and School Improvement, 19*(4), 409-428.

Stringfield, S., Reynolds, D., & Schaffer, E. (2010, October). *Toward high reliability, high quality public schools*. Invited paper presented at the Best in the World Education Consortium. Denver: McREL.

Stringfield, Reynolds, D. & Schaffer, E. (in press). Creating high reliability schools and school systems. *Phi Delta Kappan*.

Stringfield, S., & Teddlie, C. (1991). Observers as predictors of schools' multi-year outlier status. *Elementary School Journal, 91*(4), 357-376.

Stringfield, S., Winfield, L., Millsap, M., Puma, M., Gamse, B., & Randall, B. (1994). *Urban and suburban/rural special strategies for educating disadvantaged children. First year report*. Washington, DC: U.S. Department of Education.

Teddlie, C., & Reynolds, D. (eds.) (2000). *The international handbook of school effectiveness research*. London: Falmer.

Teddlie, C., & Stringfield, S. (1993). *Schools make a difference*. New York: Teachers College Press.

Vaughn, D. *The Challenger launch decision: Risky technology, culture, and deviance at NASA*. Chicago, IL: University of Chicago Press.

Weick, K. (1976). Educational Organizations as loosely coupled systems. *Administrative Science Quarterly, 21*(1), 1-19.

Weick, K., & Sutcliff, K. (2007). *Managing the unexpected (2<sup>nd</sup> ed.)*. San Francisco: Jossey-Bass.

Zavadsky, H. (2009). *Bringing school reform to scale*. Cambridge, MA: Harvard Education Press.