



**General Teaching Council
for Northern Ireland**

Promoting Teacher Professionalism

Striking the Right Balance

Towards a Framework of School
Accountability for 21st Century Learning



**Response to the NI Assembly Committee for Education
*Inquiry into the Education and Training Inspectorate and the School Improvement Process.***

Striking the Right Balance

Towards a Framework of School Accountability
for 21st Century Learning

Endorsed by

NORTHERN IRELAND TEACHERS' COUNCIL

UNIVERSITIES' COUNCIL FOR
THE EDUCATION OF TEACHERS (NI)

CONTENTS (aligned to the Inquiry Terms of Reference)

Preface

Introduction

Overview

Summary and Recommendations

Elaboration of Evidence

Effectiveness of current approaches

- 1: ETI's current approach to school inspection / improvement
- 2: How ETI assesses value added in schools

Key issues for schools

- 3: Key issues impacting on schools experiencing difficulties

Gaps

- 4: Gaps in the ETI review process
- 5: Gaps in the support services provided by DE and ELBs

Alternative approaches and models of good practice

- 6: Inspection, value-added and improvement in other jurisdictions

Recommendations

- 7: To improve ETI's approach to the school improvement process
- 8: Alternative measures of achievement
- 9: Enhanced powers, improved governance and transparency

Conclusion

References

Appendix

Survey of Teacher Perceptions of End of Key Stage Assessment (June 2013)

Preface

This submission to the Northern Ireland Assembly Education Committee Inquiry into the Education and Training Inspectorate and School Improvement, has been developed by the General Teaching Council for Northern Ireland (**GTCNI**), in collaboration with the Northern Ireland Teachers' Council (**NITC**).

GTCNI is the professional and regulatory body for teachers, which is responsible for maintaining a register of qualified teachers; approving qualifications; promoting the highest standards of professional conduct, practice and professional development; future regulation (pending new legislation); and providing advice to the Department of Education and employing authorities 'on all matters relating to teaching'.

NITC is the teacher union side of the Teachers Negotiating Committee (TNC) and is responsible for negotiating on pay and procedures to regulate conditions of service, as well as advising on educational policy. It has representation from: the Association of Teachers and Lecturers (ATL); the Irish National Teachers' Organisation (INTO); the National Association of Head Teachers (NAHT); the National Association of Schoolmasters/ Union of Women Teachers (NASUWT); and the Ulster Teachers' Union.

The submission is also endorsed by The Universities Council for the Education of Teachers Northern Ireland - **UCET (NI)** - which has representation from St. Mary's University College, the Open University, Queen's University, Stranmillis University College and the University of Ulster. UCET (NI) acts in collaboration with the wider UK UCET network as a forum for the discussion of matters relating to the education of teachers and professional educators, with a view to contributing to the formulation of policy in these fields.

Introduction

On behalf of the profession GTCNI, NITC and UCET (NI) warmly welcome this important inquiry and commend the Education Committee for initiating it. From the outset we wish to state categorically that as a teaching profession we fully accept that we should be accountable for the effective education of our young people and that robust monitoring and evaluation (both internally and externally) is needed to ensure school accountability and continuous improvement so that young people, parents, politicians and the public can have confidence in our schools and in our teachers. This submission is therefore not about whether there should be an evaluation service but, rather, it is about the approach to providing that service, the driving forces underpinning its approach, the basis for the construction and validity of the targets that it responds to, the nature of the statistical evidence that it uses, the manner in which it reports, the impact that it has on schools, particularly those in challenging circumstances, and whether there are other way of achieving similar (or better) outcomes.

The purpose of this submission is to draw attention to the now considerable amount of research evidence available about different approaches to school evaluation, both internal and external, and the use of a wider range of comparative measures and value-added adjustments that may provide a truer picture of performance and may better serve school improvement.

Our hope is that the inquiry process and outcomes will have a wider constructive impact not just on future approaches to school evaluation and quality improvement in Northern Ireland but also on the entire ethos and culture of our education system; the focus of curriculum, assessment and examinations; the measures derived from these by which schools are held accountable; how these are reported to government and parents; and how these are monitored and commented on by the Northern Ireland Audit Office and within the media.

Our aspiration is to achieve an evaluation service that is strongly linked to adequate and on-going school support and a framework for career long teacher professional development as well as to inform and influence the coherence of: Department of Education policies in relation to school improvement; future school support structures; CCEA's processes and mechanisms for assessment and examinations; and future Programme for Government Targets.

This response is structured in accordance with the following Terms of Reference which aim to:

1. Review the effectiveness of ETI's current approach in respect of school inspection / improvement
2. Consider particularly how ETI assesses the value added in those schools which have lower levels of examination attainment;
3. Identify the key issues impacting on schools experiencing difficulties;
4. Identify any gaps in terms of the ETI review process;
5. Identify any gaps in the support services provided by the Department or the Education and Library Boards;
6. Identify and analyse alternative approaches and models of good practice in other jurisdictions in terms of school inspection and alternative approaches to the assessment of value added and improvement;
7. Consider what priorities and actions need to be taken to improve ETI's approach to the school improvement process, including the need for enhanced powers; alternative measures of achievement; improved governance; and transparency.

Overview

There are a number of important caveats to be acknowledged at the outset.

ToR 1 - In order to properly and fairly **review the effectiveness of ETI's current approach in respect of school inspection / improvement:** a proper independent research analysis needs to be undertaken into the conduct of ETI inspections, the appropriateness of the quality indicators that are used; how (and whether or not these are consistently) applied; the nature of the report back to schools; whether or not the basis of judgements arrived at and reported are transparent and fair; and the impact of ETI inspection on long-term school im-

provement. **This response can therefore only refer to ‘perceptions’ about the current approach, which lack a robust evidential base.** GTCNI intends to undertake an on-line survey to explore the evidence base of these perceptions. It is also accepted that ETI’s inspection processes are continuously evolving in response to circumstances and feedback.

ToR 2: In relation to how ETI assesses the value added in those schools which have lower levels of examination attainment : It is recognised that assessing value-added is a challenging issue not only for ETI but for all schools and education systems around the world (as well as for the health service, police force; governments etc). The issues raised are therefore not issues solely for ETI (or confined to schools which have lower levels of examination attainment). Rather these are issues for all schools and the whole system. It needs to be recognised from the outset also that the Department of Education, in hand with the Assembly Education Committee, set the Programme for Government Targets by which the system is measured, apply FSME as the main accountability-value-added indicator and created Annex C of the ESaGS policy. ETI merely responds to these directives. Also, while value added may be something that all schools should be trying to measure (and only a minority do so ‘effectively’) this is likely to be because schools have had little training to help them do so.

ToR 4-6: In relation to identifying gaps in the ETI review process and in the support services provided by the Department or the Education and Library Boards it is recognised that the Northern Ireland education system has been undergoing a period of unprecedented change at a time of major financial constraint and that planned change has been slowed by democratic scrutiny. Thus gaps in the ETI review process may be exacerbated by gaps elsewhere which are not of their making.

ToR 6-7: In relation to Identifying alternative approaches to inspection, value added and school improvement in other jurisdictions, the range of international evidence cited is an indicator of the extent to which other countries are engaging with issues similar to those identified by the Education Committee inquiry; that this inquiry is a healthy reflection of what we need to be doing constantly in relation to major education policies; and that the recommendations offered are meant to be positive and enabling in evolving towards a system that engages all partners in a clear shared moral purpose of doing the best to support our schools and our young people. Bearing in mind these important caveats, and wishing to contribute constructively to this inquiry and the recommendations that may emerge from it:

- Section 1 reviews **perceptions of the ETI’s current ‘risk-based’ approach to inspection**, including the potentially in-built socio-economic bias of this approach, the excessive data requirements reported in Union case-study evidence, concerns about the weighting given to numerical outcomes, evidence of minimalist written feedback and suggestions of an increasingly deficit approach, reinforced by the current proposals for changes to the Formal Intervention Process. The paper highlights the potential unintended effects of ‘short-termist’ approaches to school improvement that run contrary to robust evidence from international research, which stresses the length of time and support needed to bring about genuine and sustainable change in the ethos and culture of struggling schools.
- Section 2 considers **how ETI currently assesses value added** (noting that the challenges raised are not confined to ETI but to the whole education system) including: the unreliability of many of the measures used, such as free school meals; the potentially distorted picture of performance presented by a reliance on 5A* to C at GCSE; the

standard and random errors that are not reported; the lack of attention to confidence intervals; the complete lack of confidence in the numerical ('level') outcomes from statutory assessment evidenced by GTCNI's recent survey (June 2013) and by the 'Expert Panel on Assessment' (DfE 2011).

- Section 3 identifies **the key issues impacting on schools in challenging circumstances** (noting that these issues are not confined to these types of schools only) including: insufficient use of base-line measures; lack of cognisance afforded to research related to family and community factors; the peer effect and the impact of separating young people from the positive influences of their better off peer group at a vulnerable age; leading to pupil *'compliance without engagement'* (Harland et al., 2002) and student underperformance and drop out (Purvis et al., 2011).
- Section 4 identifies **Gaps in the ETI review process** including: lack of analysis of effect sizes and correction for student intake; over-estimation of the school effect which is considered to range between 5% and 18%; and conflation of the term 'effective' (a statistical term borrowed from economics) with the perception of 'good' (which is a value judgement) (MacBeath, 2012: 44).
- Section 5 identifies **Gaps in DE and ELB support** including: delays in strategy setting, for example, the decade-long delay in the Review of Teacher Education; the current gaps between policy direction and support capacity, for example the assumption of capacity within the support services to provide the level of tailored response likely to be needed as a consequence of proposed changes to the Formal Intervention Process; the overall run down in provision for teacher professional development; the gap in the policy drive towards 21st Century learning *'to ensure that 21st century skills that are considered important, become valued in the education system'* (OECD, 2011: 19); and the pressing need to develop a coherent professional development framework for teachers and to consult on the shape of a future advisory and support structure.
- Section 6 identifies and analyses **alternative approaches and models of good practice** including: Finland, which does not have an Inspection Service; Scotland, which has developed a constructive model closely aligned with support; New Zealand, which uses census information to stratify schools; Hampshire, where value-added estimates for primary schools were utilised by the authority and head teachers as an unpublished 'screening device' and a 'school improvement' tool; and good practice models from a range of other settings including Hong Kong, Germany, Spain Slovakia and The Australian Capital Territory.
- Sections 7-10 considers priorities for action to **improve the approach to the school improvement process**, including **recommendations** on construction approaches to school evaluation; more sophisticated base-lining and value-added calculations; the use of alternative measures of achievement; and the need for greater coherence in educational policy and sustained career-long professional development and support.
- Section 11: In **conclusion** the submission calls for a more constructive model of accountability, underpinned by proper base-lining and value-added measures which builds teachers' confidence and commitment. The overall recommendation is that future policies should seek to **strike the right balance** - *'between holding schools to account*

and allowing innovation and supporting school improvement' (Perry, C., 2012, P1, NIARIS).

Summary of Evidence and Recommendations

1 The current approach to inspection/school improvement may serve to:

- 'incentivise schools to prioritise compliance... over innovation' (Perry, C., 2012);
- prioritise performance data over other factors and 'pre-judge' outcomes;
- produce a range of undesirable practices with unintended consequences;
- confirm an 'in-built' social bias which in turn fails to recognise value-added;
- feed a form of 'blame culture', holding schools to account for failure to overcome the absence of family and community cultural capital (MacBeath, 2012);
- exacerbate fear and lead to a downward spiral towards school closure.

2 The current approach to value-added is fundamentally flawed because:

- It fails to take full enough account of factors which influence variations in pupil attainment; to analyse school effect sizes and correct for student intake (Sammons, 2007);
- Statistical differences tend to conceal more than they reveal (Mc Beath, *ibid*);
- Performance indicators lose usefulness when used as objects of policy (William, 2001);
- Reducing attainment to a single figure or grade, while attractive to politicians and the public '... masks complex nuances in ability and performance'. (Gipps, 1994);
- Trying to achieve multiple objectives with a single policy instrument is not feasible (Hanushek & Raymond, 2004).

3 Key issues for schools and gaps in support include:

- The lack of solid evidence that investing in increasingly sophisticated measurement devices drives change (OECD –Scotland report -2007);
- The constant focus on measurement may serve to place intense pressure on young people (MacBeath, 2012) resulting in 'compliance without engagement' (Harland et al, 2002); and 'disengagement' by many (Purvis et al., 2011);
- Selection exacerbates differentials by removing positive peer effects (OECD, 2011);
- The run-down of services associated with ESA has resulted in a deficit model of support; there is no coherent strategy for teacher professional development or evidence of change-management planning for a future school support strategy.

4 Alternative approaches/models that should be considered include:

- Finland, which does not have a school inspection regime at all;
- Scotland and Ireland, which emphasise a two-way collaborative approach; and
- New Zealand, which uses census and other information to stratify schools by socio-economic intake.

RECOMMENDATIONS: to improve the approach to school improvement

1. **Undertake a cost benefit analysis** of the relationship between inspection and school improvement (Whitby, K. 2010 in Perry, C., 2012, P21)
2. **Develop a supportive quality assurance model** (Finland/Scotland) which uses positive language (for example, *Very Confident, Confident, Not Confident* as in Scotland) aligned to support systems that involve more seconded teachers and principals;
3. **Stream-line future school evaluation processes** to provide clearer guidance on data requirements; permit verbal (and written) challenge; reduce reporting timescales; and improve the qualitative detail of unpublished reporting to schools.

RECOMMENDATIONS: to improve the assessment of value-added

4. **Use NISRA census information and geographic information system (GIS)** to identify school characteristics and to stratify schools by socio-economic intake to help allocate resources effectively, target social need and calculate value-added.
5. **Assess productive language (oracy) on entry to school** as a key indicator of future educational potential and as a base-line measure of school value-added.

RECOMMENDATIONS: to improve system monitoring

6. **Use light sampling** to provide robust and independent monitoring data over time, disentangling teacher assessment from accountability (*Tymms & Merrill*);
7. **Use International data** (PIRLS, TIMSS and PISA) to provide additional quantitative and qualitative information as a broader comparative measure.

RECOMMENDATIONS for alternative measures of achievement

8. **Commission international research and development** to assist CCEA in developing innovative 21st Century assessments and examinations.
9. **Separate teacher assessment from accountability** to safeguard assessment for learning.
10. **Develop wider indicators** to 'enable progress in all important learning goals to be reported' (*ARG, 2008*) and to broaden measurement of 'value-added'.
11. **Use standardised testing data sensitively within schools only for** diagnostic, formative and value-added purposes to prevent teaching to the test.
12. **Use pupil attitudinal and 'well-being' surveys sensitively** to gain insight into the correlation between 'motivation', 'liking' and achievement (*Sturman, 2012*).
13. **Develop 'unseen' thinking skills assessments** 'to ensure that important 21st Century skills become valued in the education system' (*OECD, 2011: 19*).
14. **Develop new qualifications for N. Ireland** which reflect the needs of young people, the economy and employment in the 21st Century (*CBI, 2012*).
15. **Introduce a measure to reduce the number of pupils leaving school with no qualifications** by an agreed percentage.
16. **Review Programme for Government Targets and NI Audit Office Monitoring** to reflect these recommendations, based on an understanding of supportive accountability.

RECOMMENDATIONS for additional powers, governance and transparency

17. **Ensure accurate and transparent media reporting of educational outcomes.**
18. **Require that the evidence-base for ETI judgements is open and transparent.**
19. **Ensure that all future educational policy is based on sound research.**
20. **Invest in teacher professional development and improve political and public respect for teaching as a profession:** Re-route spending on statutory assessment and evaluation systems towards teacher professional development. Develop greater political and public appreciation of the complexity of education, issues of socio-economic deprivation and equity, and the quality of the public service which teachers provide.

1: Perceptions of ETI's approach to school inspection and improvement

1.1 **Perceptions:** At the outset it is important to state that ETI is funded directly by government, and while independent in its management and actions, is located within the Department of Education. The general perception therefore is that ETI acts in line with policy determinations from the Department of Education which are formulated in response to Programme for Government Targets endorsed by the Education Committee. Secondly, it is important to state that, in the absence of detailed research into schools' experience of school inspection, this response can therefore only refer to 'perceptions' about the current approach, which it is accepted lacks a robust evidential base. Thirdly, it is also accepted that ETI's inspection processes are continuously evolving in response to circumstances and feedback and that recent pilot approaches seek to take greater account of schools' own self evaluation evidence. It is therefore recommended at the outset that, in order to properly and fairly review ETI's approach to school inspection and improvement a proper independent research analysis should be commissioned into:

- the conduct of ETI inspections,
- the appropriateness of the quality indicators that are used;
- how (and whether or not these are consistently) applied;
- the nature of the report back to schools;
- whether or not the basis of judgements arrived at and reported are transparent and fair; and
- the impact of ETI inspection on long-term school improvement.

In the absence of that research, GTCNI intends to undertake an on-line survey to explore the evidence base of these perceptions.

1.2 **The shift towards a 'risk-based' approach to inspection:** A number of literature reviews (Penzer & Allen 2011) and comparative research studies (Ozga *et al.*, 2009-13; Ehren *et al* 2011-13; WBEE/EBT *etc*) explore different modes of inspection in different countries. These comparative research studies reveal that there is no single and unchanging form of inspection. Rather '*Inspection... remains unsettled and changeable, caught up in the processes of 'hyperactive' policy making and management*' (Clarke & Ozga 2011) and influenced by specific political, cultural and institutional conditions in each country.

While no official research study has, as yet, been undertaken into the changing nature of inspection in Northern Ireland, research undertaken for the Education Committee suggests that there has been a shift towards a more '*risk-based' approach*' with performance indicators becoming '*the major determinant of when schools should be inspected*' (Perry, 2012). It is known however that a key deciding factor in prioritising schools for inspection or identifying risk is provided by District Inspector local knowledge, as opposed to performance data on its own. Whatever the source, the shift towards risk-based inspection is confirmed by recently in proposals for changes to the Formal Intervention Process (DE, June 2013).

Education Committee research highlighted '*concerns around the pressures for organisations undergoing inspection and ...that evaluation can incentivise schools to prioritise*

compliance with requirements over innovation' (Perry, 2012). This observation is supported by several other research studies which have highlighted the increasingly '*performative*' character of the inspection process in many countries, with school staff using metaphors such as '*jumping through hoops*' and '*papering over the cracks*' (Plowright, 2007); or '*nominal compliance*' with the '*performance*' of accountability with good teaching on a '*stage managed*' basis (Case et al., 2000 in Clarke & Ozga 2011:18).

1.3 A potential inbuilt socio-economic bias: A number of critical concerns have been identified about the increasing use of school performance indicators as the major determinant of when schools should be inspected and their influence on inspection judgements.

- *The first is that performance needs to be contextualised and adjusted for the differential selection of students by schools in Northern Ireland and school examination results need to be adjusted for the intake achievements of students when they start at a school – so called 'value-added' ratings.*
- *The second issue is that the uncertainty surrounding any given ranking is very large, and in many important cases so large that no statistically meaningful comparisons can be made, nor can useful user choices be sustained (Foley & Goldstein 2012).*

It has demonstrated in the United States, for example, that '*many low-attainment schools are actually high-performing. The reverse is also true, though problems of poor performance are generally well hidden in high-attainment schools*' (Harris, 2010:3).

Analysis of inspection outcomes over the last few years suggest that schools from the least advantaged social band are four times more likely to receive an "inadequate" or "unsatisfactory" grade than those from the most advantaged intake, which are twice as likely to get an "outstanding" or "very good" inspection outcome (*Irish News, 26 February 2013*). This is substantiated by analysis in the United States which highlights that:

Attainment-based school performance measures like proficiency are systematically biased against schools serving low-attainment students. That is, by failing to account for factors affecting achievement that are outside the school's control, we systematically under-estimate the performance of low-attainment schools (Harris, 2010: 6).

It is argued that, if inspection took appropriate account of intakes characteristics, then schools in each social band should be able to achieve the same broad range of inspection grades. The following research observations will be elaborated more fully throughout this submission:

- *The first rule of accountability is that people can only be held responsible for the things over which they have control (Harris, 2010).*
- *The cause of 'differentials in performance lie largely outside schools and the classroom' (Purvis et al, 2011).*
- *The school effect is commonly agreed among researchers to be between 5 and 18 % (Chevalier, Dolton and Levacic, 2005; MacBeath, 2012: 44).*

1.4 Potentially excessive data requirements: Teacher Union case study data suggests that it is now the 'norm' in standard inspections for schools to return data in the range of

2 gigabytes (around 700 pages). C2kni guidance to schools on formatting pre-inspection reports runs to 52 pages. While the evidence which informs inspection judgements includes classroom observation, interactions with pupils, parents and staff, the perception is that pre-inspection data may serve to 'pre-judge' the actual inspection process with judgements likely to "follow the stats" (Mansell, W, 2007).

1.5 Nature of reporting: Although it is acknowledged that the oral report back which schools receive can be very detailed and helpful, written reports are described by many as lacking in detail and 'bland'. The perception – whether real or not, is that inspection reports in the past offered a richer, more rounded, picture of the school inspected.

The current practice in Scotland is to provide a short report of the DE type for publication and to provide a more detailed confidential report to schools. Schools have indicated that the lack of detail inhibits them from being able to challenge judgements that may be based on relative measures that could be 'subject to considerable margins of error' (ARG, 2008).

1.6 Consistency: Concern has been expressed about inconsistencies in the judgements made by different inspection teams, with insufficient transparent evidence provided to verify the basis of the judgements made. Representatives of the Irish-medium sector have registered particular concern about being inspected by personnel who do not speak Irish and may therefore be unable to recognise the language development of children or capture the detail and quality of the interactions and relationships between teachers and children in an Irish-medium classroom and the value-added by bi-lingual education.

1.7 An increasingly deficit approach: The perverse organisational effects of inspection have been much discussed in the research literature. Many studies point to the dislocation and distraction associated with being inspected. Some studies suggest that the impact of Inspection on school performance may be neutral or even negative, with some studies reporting lowered examination performance in the 12 months following an inspection (e.g., Shaw et al, 2003; Rosenthal, 2004). Counter-balancing this view, early evidence from a current study across a number of EU countries suggests that there is a positive effect from inspection. The degree of improvement, however, is significantly related to the promotion of self-evaluation and is moderated by whether the feedback is positive or negative (Ehren et al., 2013).

Viability audits associated with school rationalisation have exacerbated fears that a poor inspection grade can lead to negative media reporting, provoking parental 'stampedes' away from schools placed in "intervention", beginning a downward spiral to potential school closure. Again, whether evidence-based or not, the general view is that the inspection process in Northern Ireland is no longer perceived by the profession as the positive and constructive experience it once was, but is increasingly characterised as more akin to a judgemental, OFSTED-inspired, model.

Current proposals for changes to the Formal Intervention Process confirm these fears. The justification offered for the proposals is that:

- a number of schools in FIP (Formal Intervention Process) are not improving sufficiently quickly, despite action plans being developed and support being provided;

- a number of schools evaluated as 'satisfactory' have not been demonstrating any discernable signs of improvement over a number of years and would benefit from the support provided through the formal intervention process;
- a perception exists that schools in formal intervention evaluated as 'satisfactory' in a follow up inspection automatically exit formal intervention;
- there have been developments in other areas of education policy such as area planning which need to be reflected in the revised process.

(DE, 20 June 2013).

The proposed changes to the process intend that:

- A school in formal intervention which improves to a 'satisfactory' evaluation at the follow-up inspection, having had two years of tailored support, will have a further follow-up inspection within 12 months at which point it must have improved to at least a 'good' evaluation or further action may be considered;
- The timing of the follow-up inspection for a school with a 'satisfactory' evaluation will be shortened to between 12-18 months;
- It will be made more explicit in the FIP process that a school will not automatically exit FIP on an ETI evaluation of 'satisfactory'
- For any school entering formal intervention and identified as being unsustainable the Managing Authority will be required to bring forward to the Department a plan for the restructuring of education provision in the area (DE 2013).

If implemented, the impact of these proposals will be to assign a time-limit to a 'satisfactory' judgement with the threat that, if measurable improvement is not visible within a specified period, the school will technically be considered 'unsatisfactory', even if it has managed to sustain its initial improvements. The proposals threaten the ultimate sanction that if progress is not made the school may be amalgamated or perhaps even closed down. It has been shown elsewhere that:

The practice of increased frequency of inspection for 'unsatisfactory' or even simply 'weaker than average' schools may be an effective one in some circumstances but it may have a negative side effect in tending to reinforce a notion of 'inspection as punishment' (Vass and Simmonds, 2001).

It is possible that this may increase the tendency of schools to focus on 'passing' their next inspection rather than on learning from the previous inspection and using it as a catalyst for improvement (Penzer & Allen, 2011: 10).

It is suggested that punitive measures of this kind may 'help to push good teachers out of schools serving low-performing students, as these teachers become frustrated by a system that punishes them no matter how well they perform' (Harris, 2010: 3). In these circumstances it is unlikely that energetic and effective leaders will be willing to take on challenging schools.

Counter-balancing this view is the acknowledged regard for district inspectors, who are generally viewed as acting in a supportive role, promoting an understanding that inspection is not an event but a continuing process leading to improvement. It is this type of role which emerges in the research literature as one which schools value and which

promotes and enables genuine improvement. Indeed, the view has been expressed that ETI should adopt much more of a support role, informing (and perhaps leading – as in Scotland) other support services. The view expressed by some schools is that ETI, which observes school practice on a regular basis, is in a better position to advise on the nature of improvement than CASS colleagues who do not have the benefit of observational experience.

1.8 Unintended effects of 'short-term' accountability pressures: While the intention is that the publication of inspection reports should have positive effects it has been shown that overly strong accountability systems can produce a range of unintended and undesirable practices and perverse 'side effects' that leads to excessive focus on improving performance in narrow areas, to the neglect of other important areas of schooling and to the detriment of pedagogy and learning (OECD 2012). As external pressure on teachers to meet performance targets and maximise league table rankings increases, a growth has been detected in techniques linked to 'gaming' the system, spoon-feeding pupils, teaching to the test, 'nursing' the coursework and manipulating the grade boundaries. (Wilson, Croxson and Atkinson, 2006; Wiggins and Tymms, 2002; Visscher, 2001) These studies argue that in some cases institutions become so focused on the measures and standards employed by league tables that they begin to deliberately manipulate their data or behaviour to produce the desired results, regardless of potentially adverse effects (Foley and Goldstein, 2012: 29).

- *When school performance is measured poorly it creates a variety of perverse incentives to do things that are clearly inconsistent with a school's mission (MacBeath, 2012: 22).*
- *Such a focus on 'doing well' could lead to distortion as a school puts its best foot, as distinct from its everyday foot, forward and may in extreme cases lead to deception (hiding known areas of weakness from inspectors). It gets in the way of inspection as a collaborative activity between professionals and encourages inspection as a competition between school and inspectors (Penzer & Allen, 2011: 10).*
- *'The higher the stakes are for school leaders and teachers, the more these unintended /undesired effects are likely to occur' (Hooge et al, 2012: 10)*

Smith (1995) sets out a number of means by which 'gaming' takes place:

- *concentrating on those students with whom most 'profit' can be gained to improve a school's Student Progression Information (SPI) while ignoring the needs of students at either end of the ability spectrum (This form of 'gaming' focus was part of the initial brief in the recent OFM/DFM initiative to employ c.270 new teachers in struggling schools to focus on Level 4 pupils at Key Stage 2 and Grade C boundary pupils at GCSE);*
- *selective student admissions and removing 'difficult' students (with students not being admitted into some grammar school 6th forms who have not scored a requisite number of grades at GCSE);*
- *concentrating on examination performance to the exclusion of other qualifications and teaching for the test;*

- *'creative reporting' of data; and /or depression of baseline/intake test scores to improve the value-added scores.*

(Foley and Goldstein, 2012: 30)

There is evidence to suggest that the results of such practices may in some cases actually prove detrimental to overall educational standards. A variety of teachers and head teachers interviewed by Wilson, Croxson and Atkinson (2006) reported that they did tend to focus extra resources on 'borderline' pupils (those who are likely to achieve C or D grades). This was acknowledged to have consequences for others. One interviewee admitted *'the bright kids still prosper... I don't think they miss out at all. But I think the lower ability ones potentially do'* (164). Others reported that they deliberately shifted these borderline pupils to vocational qualifications (ibid: 30)

A report in the Times Educational Supplement in mid August 2013 confirmed that GCSE grade deflation can in large part be explained by significant increases in early and multiple entries. Across all subjects there was a 39 per cent rise in entries from students who were aged 15 or younger. In mathematics, the proportion of entries from under-16s increased by 49 per cent so that the total number of entries amounted to nearly twice the number of 16-year-old students. The fall in performance is partly attributed to younger candidates' attaining lower results and reveals:

'... The full extent of the tactics used by schools caught between tougher government targets and exam watchdog Ofqual's clampdown on grade inflation. As Ofqual has intensified its "comparable outcomes" clampdown on grade inflation, school leaders are concerned about the impact of the watchdog's approach on their ability to meet government GCSE targets. "Schools are constantly trying to improve outcomes for pupils, whereas Ofqual and the exam boards are geared to making sure that there is no room for improvement. "The accountability system is built around a measure that [teachers] don't trust any more." (TES magazine on 23 August 2013)

It is important in the interests of balance to acknowledge Fisher and Downes (2008) research which concluding that while the propensity to manipulate metrics can be quite high, 'the deception is usually of a low level of ethical seriousness.' Nevertheless, MacBeath observes that: *'The higher the stakes for schools the more children are placed under intense and perhaps excessive pressure from policy driven demands'* (2012, 22).

Wiggins and Tymms (2002) compared the performance-measurement culture in England with a more supportive culture in Scotland. They found that the stress of performance targets is increasingly associated with a more 'short-termist' approach among English teaching staff and, in some cases, the development of a blame culture. They concluded that *'high-stakes, single-proxy indicators...can have significant dysfunctional effects'*.

The British Academy has called for:

More research [into] the effects of performance data on institutional performance. ...This evidence should pay particular attention to 'knock-on' effects whereby resources may be reduced for some important activities in order to improve performance (Foley & Goldstein 2012: 11)

Visscher (2001) has highlighted the institutional damage done by '*naming and shaming*' persistently under-performing schools. He argues that presenting simple comparative measures will always lead to some schools performing at a relatively lower standard, but that the focus should remain on whether each school reaches the standards considered appropriate by virtue of their intake.

In addition to these educationally undesirable pressures, the current 'Formal Intervention' proposals run contrary to a wealth of research findings which point to the length of time and support needed to bring about a genuine and sustainable change in the ethos and culture of struggling schools. For example, '*it may take approximately 30 hours of focused in-school, job-embedded learning before coherent improvements in teaching and learning become obvious*' (Reeves, 2006). Engendering such fundamental change often requires changes to leadership and collegial practice and also major change in the relationships with, and aspirations of families and communities.

Ben Levin in '*How to Change 5000 Schools*' emphasises that '*improving schools is hard work*' and needs to be done '*in ways that support positive morale among educators, students and parents*' (Levin, B., 2008:2). Substantial research evidence would suggest that the current proposals for changes to the Formal Intervention Process here, which are not accompanied by related plans to reinvigorate the advisory service, under-estimate the nature of the changes and support needed to engender sustainable improvement and are likely to exacerbate perverse behaviours.

2: How ETI assesses value-added

2.1 Pre-determined policy measures: It is recognised that assessing value-added is a challenging issue not only for ETI but for all schools and education systems around the world (as well as for the health service, police force; governments etc). The issues raised are therefore not issues solely for ETI (or confined to schools which have lower levels of examination attainment). Rather these are issues for all schools and the whole system. It also needs to be clearly recognised that it is the Department of Education (DE), in hand with the Assembly Education Committee (AEC) which sets the Programme for Government Targets by which the system is measured. (It is not known on what basis these targets are derived). It is DE and the AEC who apply Free School Meal Entitlement as the main accountability-value-added indicator. It is DE which has specified the Formal Intervention Process. ETI merely responds to these directives. Also, while value added may be something that all schools should be trying to measure (and only a minority do so 'effectively') this is likely to be because schools have had little training or help to do so.

2.2 Important trade-offs: The assessment of a school's performance monitoring must navigate some very important trade-offs between:

- *the accessibility and intelligibility of the information and measures used and the accuracy of that information;*
- *the availability of information and its validity as a performance measure;*
- *qualitative and quantitative measures;*
- *technical questions of adjustment and reliability (Bird et al., 2005, Goldstein and Spiegelhalter, 1996, Leckie and Goldstein, 2009) that limit the inferences that can be legitimately drawn, whether for the purpose of institutional accountability or for user choice.*

This is not to say that these kinds of issues always occur in practice, and indeed there are a range of other dangers associated with an entirely unregulated system. However, it is vital for policymakers to remain aware of such potential problems within performance monitoring frameworks.

(Foley & Goldstein, 2012: 20).

Since inspection feedback is insufficiently detailed in relation to the basis on which qualitative judgments have been made, it is not always clear that these important trade-offs are being taken into account so that the limits of the inferences that are being drawn are both apparent and transparent.

2.2 Perceived over-emphasis on numerical value-added: ETI undertake criterion referenced inspections using 5 indicators, supported by the detailed range of indicators set out in *Together Towards Improvement*, which vary slightly by sector. The perception of schools is that a stronger emphasis is placed on numerical evidence of having met performance targets (i.e. *the % of pupils achieving designated levels of attainment at specific key stages or 5A* to C at GCSE*). While ETI states that the full range of indicators are applied in a balanced way to arrive at criterion referenced judgements, the perception that numerical evidence has a stronger influence than other criteria is borne out by the emphasis in inspection reports.

In the absence of more finely tuned base-line measures of a school's intake profile, reliance on numerical data primarily is an insufficiently robust basis on which to assess school quality or value-added, which is why observation of practice by the inspectorate is an extremely important dimension in judging school quality. What needs to be made more transparent is the extent to which the interpretation of data influences the inspection outcome since:

statistical data remains problematic and potentially unreliable. The agents who collect it may try to manage the representation of performance; the indicators chosen may not be adequate to the reality they are intended to convey; and performance management systems are persistently vulnerable to problems of 'gaming' as evaluated organisations and actors try to produce success. As a result, the apparent 'hardness' of statistical fact is itself an artefact (Poovey in Clarke & Ozga, 2011: 5).

Over reliance on numerical data has been challenged by a number of research studies, as the following quotations illustrate,

- *Use of assessment evidence for accountability is based on the idea that measuring itself leads to improvement....Over the last 20 years there is no solid evidence from research or practice that investing in increasingly sophisticated measurement devices drives change (OECD –Scotland report -2007, p15);*
- *Performance indicators lose their usefulness when used as objects of policy.... When used as the sole index of quality, the manipulability of these indicators destroys the relationship between the indicator and the indicated' ('Goodhart's Law' - former chief economist at the Bank of England quoted in Wiliam, 2001: 2);*
- *'...Put bluntly, the clearer you are about what you want, the more likely you are to get it, but the less likely it is to mean anything' (Wiliam, 2001: 2);*

A wide range of research has questioned the value of Levels of Attainment in particular, as a robust numeric, highlighting that:

- *Reducing attainment to a single figure or grade while attractive to politicians and the public 'as a form of shorthand' in which to report performance masks complex nuances in ability and performance (Gipps, 1994: 27);*
- *No single measure can fulfil both the formative and summative functions (Harris, 2010);*
- *Assessments should be treated as approximations, subject to unavoidable errors (Gardner, 2008);*
- *'Trying to achieve multiple objectives with a single policy instrument is not feasible' (Hanushek & Raymond, 2004).*

2.3 Unreliability of measures of deprivation, attainment and progress

Compounding the problem of an over-reliance on data is the research evidence which suggests that the various components which comprise the data set are individually unreliable. There is not space to do justice to what is a contentious and well covered issue but some issues that require further research and reflection include:

2.3.1 Free school meals: The use of free school meal (FSM) data is widely prevalent in official estimates of educational disadvantage as well as in educational research reports in the UK. However, while there has been some concern expressed about the measure, there has, to our knowledge, been no systematic test of its appropriateness. Research at Bristol University has tested the use of FSM for appropriateness as a measure, taking into account the dynamics of poverty and the error that can be associated with its application in judging school performance. The research found that FSM is a coarse and unreliable indicator to judge school performance and leads to biased estimates of the effect of poverty on pupils' academic progress. Using county-wide data to assess the magnitude of error that can be introduced in estimates of the prevalence of economic disadvantage the associated error was found to be large (10%) and was also found to lead to an underestimation of the proportion of children who consistently remain below the income thresholds implied by the FSM-eligibility criteria by 50%. The research concludes that:

FSM eligibility is not just a coarse indicator of socio-economic of disadvantaged considerably... Moreover, the progress of children from very poor backgrounds early in life could also be overestimated in schools with low FSM take up rates. Finally, and most importantly these findings raise questions about the way progress in schools is 'officially' measured and raises doubts about the trust that is invested in FSM as a reliable indicator of deprivation. It also raises questions about the estimates of school effects based on models where FSM entitlement is used as a measure of disadvantage. This work questions the architecture of accountability which drives the state theory of learning in England (Lauder et al., 2006). Our findings suggest that many schools will confront far greater levels of disadvantage than what is currently measured by FSMs.... It is important not to see the problem of quantifying the poverty related educational disadvantage as just confined to measures such as FSMs (Miles & Evans, 1979). Rather, it can be argued that disadvantaged populations will always be difficult to 'capture' through single catch-all measurements from routinely collected administrative data such as FSMs (Kounali et al, 2012).

The findings raise important policy questions about the quality of indicators used in judging school performance. Recommendation 21 of the *Independent Review of the Common Funding Scheme* advises that 'ongoing investigation into an alternative, or adjunct measures to Free School Meals should continue' (Salisbury, 2013: ix).

2.3.2 5 A*s to C at GCSE and A level: It has long been acknowledged that the performance of some schools is flattered by the focus on the proportion securing 5 A*-C GCSEs. Recent detailed research from the University of Ulster (Borooah &

Knox 2013) highlights that there are schools which should be doing much better than they are given their intake.

*Attainment in national tests such as Level 4 or 5 in English and Mathematics at primary schools or GCSEs in post-primary schools are a crude indicator of the **value which schools add** to the pupils in their classroom and hence the quality of education on offer. For example, good results in GCSEs in grammar schools attracting academically able pupils might hide the fact that teachers added little to their performance. Compare this with less good results in a secondary school attracting a large percentage of pupils from disadvantaged areas where their teachers have added significantly to the performance of pupils.*

(Borroah & Knox 2013: 1)

It has also become apparent that the focus on the C threshold encourages schools to invest considerable resources at the C/D borderline which can drive perverse behaviours. It has been argued that a fairer way to judge school performance would be to measure the *attainment of all the pupils* rather than the sub-set who achieve the highest grades. In principle, this might be a step forward but, as ever with issues of assessment, it is complicated to calculate fairly. Crucially, this is still not a measure of the level of progress made by an individual or a cohort from entry to exit, which is a much more genuinely inclusive measure. That too is complicated to measure, as the education system is gently tilted back towards norm- rather than criterion-referenced assessment methods, so that not all pupils may be able to make three levels of progress. The key point is that:

'Accountability' for performance in education is complex. Developing measures which genuinely allow schools to demonstrate what they have achieved with young people is complex. Translating it into a readily understood format which can be communicated clearly is perhaps even more complex. At root, society needs clarity about what it wants to hold schools to account for: the progress made by individual pupils, in which case we should worry less about thresholds, or their ability to move all pupils to an agreed threshold, and threshold performance, or their ability to push the most able to elite levels of performance. We need to reflect on how to map the performance of all. Until we clarify that, we will struggle with inadequate measures in which we vest too much confidence" (Husbands, C. 2012)

2.3.3 Standard and random errors: The results of most tests are reported using either standard scores, percentiles or grades which purport to measure and describe how a student performed on a test compared to a representative sample of students of the same age from the general population. This comparison sample or group is called a norm group. Educational tests cannot by their nature measure abilities and traits perfectly so, no matter how carefully a test is developed, it will always contain some form of error or unreliability. This error may exist for various reasons that are not always readily identifiable.

Random errors might seem innocuous because they are equally likely to arise with all teachers. But random errors are problematic because they call into question the conclusions we wish to draw from performance. Thus both systematic and random errors need to be taken into account when making decisions about performance measures (Harris, 2010:7).

2.3.4 Confidence intervals: In order to account for this error, confidence intervals can be calculated within which the student's true score is likely to fall over a certain percentage of the time. For example, if a student earned a standard score of 90 with a confidence interval of +5, it is more accurate to say that there is a 95% chance that a student's true performance on this test falls somewhere between 85 and 95.

Best practice in assessment and examining would make confidence intervals transparent; for example, New Zealand reports assessment scores to parents showing confidence intervals graphically. Similarly, confidence intervals exist in arriving at subjective judgements in inspection between the views of individual inspectors and between different teams of inspectors in different schools at different times. Therefore there needs to be greater transparency in accepting that 'assessments should be treated as approximations, subject to unavoidable errors' (Gardner, 2008).

When making comparisons between institutions it is assumed that we are interested not merely in how they happened to perform at the time when the data were collected, but how they compare in terms of their underlying 'effectiveness'. Thus, for example, to base a comparison using just one randomly sampled student from each school would be very unreliable and hardly acceptable. The question is then to determine how many students contributing to a school's score would be adequate. By providing a range or interval for each school we can indicate the relative accuracy for different schools, with larger intervals associated with less accuracy. Judgements can then be made about whether differences can be ascribed to chance variation due to small numbers of students, or may reflect real differences. (Goldstein and Spiegelhalter (1996) provide a detailed discussion).

(Foley and Goldstein 2012: 23)

Visscher (2001) points out that even if student achievement scores have been adjusted for relevant student background characteristics 'precise school performance remains uncertain as a result of large confidence intervals' (202). Large confidence intervals are just one of the results of the relatively small sample size constituted by the average school's yearly cohort.

In research on this problem in the United States, Kane and Staiger (2002) found that the median elementary school has only 69 students per grade (in the UK, the average primary school year group is just 40). They point out that 'the 95% confidence interval for the average fourth-grade reading or math score in a school with 69 students per grade level would extend from roughly the 25th to the 75th percentile among schools of that size (95). (ibid: 26)

2.3.5 Statutory assessment and levels of attainment: In a recent independent survey conducted by GTCNI (June 2013) which received 500 responses representing almost 50% of schools involved in end of key stage assessment, only a very small percentage of respondents considered numerical Levels were useful to:

- To Boards of Governors to understand value-added only 15%
- To parents to understand their child's progress only 10%

- To receiving schools to understand what a pupil knows only 18%
- To ELBs to understand the support a school may need only 18%
- To ETI to understand the value added by schools only 18%
- To DE and Politicians to understand system performance only 14%

The 'Expert Panel on Assessment' (DfE, 2011) advised that: *'The ways in which 'levels' are currently used to judge pupil progress, and their consequences actually inhibits performance, distorts and undermines learning and exacerbates social differentiation, rather than promoting a more inclusive approach'* (DfE, 2011).

Embargoed until 27 September 2013

3: Key issues impacting on schools experiencing difficulties

3.1 The theory of inspection leading to improvement: According to an EU project currently evaluating the impact of school inspections, the theory informing school inspections is that:

School inspection criteria and procedures and the feedback given during inspection visits are expected to enable schools and their stakeholders to align their views/beliefs and expectations of good education and good schools to the standards in the inspection framework, particularly with respect to those standards the school failed to meet during the latest inspection visit. Schools are expected to act on these views and expectations and use the inspection feedback when conducting self-evaluations and when taking improvement actions.

Stakeholders should use the inspection standards, or rather the inspection assessment of the school's functioning against these standards (as publicly reported), to take actions that will motivate the school to adapt their expectations and to improve. Self-evaluations by schools are expected to build their capacity to improve that will lead to more effective teaching and learning conditions. Likewise, improvement actions will (when successfully implemented) lead to more effective schools and teaching conditions. These conditions are expected to result in high student achievement. (Ehren et al 2012).

While the theory is that inspection will lead to improvement (and this is ETI's mission statement) extensive research suggests that external school evaluation has differing impact on schools and that certain conditions are associated with schools accepting and acting on feedback from external school evaluation.

3.2 Tensions between inspection and improvement: Evidence from across 17 countries reviewed by CfBT suggests that the conflation of 'inspection' and 'improvement' are in tension with each other. On the one hand accountability looks outward from the school (towards government and other stakeholders) and aims to be an objective process. Conversely, school improvement is focused inward and is achieved subjectively, by the particular people who work in and attend the school, with their own particular strengths, weaknesses, motivations etc. The 2010 CfBT report suggested that there is little evidence of a properly grounded, evidence-based effort to resolve the conundrum. In the real world, something more is needed to translate inspection outcomes into school improvement.

"Professionals need to be fully engaged in the change process and to feel a high degree of ownership about the outcomes. [This] requires an infrastructure for changing professional practice that ensures the profession owns and drives the change. "(Harris, 2010)

The first technical report from the current EU-project 'Impact of School Inspections on Teaching and Learning' suggests that stakeholder pressure and setting expectations do directly influence and affect improvement actions for school effectiveness which in turn is influenced by improvement in teacher cooperation transformational leadership and capacity building. The degree of improvement, however, is significantly related to the

promotion of self-evaluation and is moderated by whether the feedback is positive or negative (Ehren et al., 2013).

3.3 Styles of inspection/evaluation to promote improvement: The current EU study (Ehren et al. 2013) confirms that the way an inspection is performed and the way staff perceive it have a direct impact on the nature of their response to its outcome. Teachers' emotional reactions to inspection and its aftermath are critical to determining whether any improvements transpire. While the ultimate responsibility for staff morale rests with the school and in particular, its head, the issue of maintaining staff morale and self-esteem needs to be designed into any evaluation process as an important requirement and pre-condition to help persuade teachers to embrace the changes necessary for improvement. Researchers have identified four steps that are needed to achieve improvement:

1. *School governors, owners, management and teaching staff need to be persuaded and convinced that the conclusions of their inspection are valid, accurate and balanced, and that they encapsulate the most important issues for the school to address.*
2. *The school needs to obtain, or be given, the resources it requires in order to make whatever changes are desirable. By resources we do not mean just money, but also access to the skills and advice it needs and – if required – to training for its staff or, indeed, new staff.*
3. *Staff at all levels in the school must be motivated to alter their ways of working, and to have the self-confidence to take the risks which change and development programmes inevitably involve.*
4. *Finally, there need to be effective systems of encouragement and reward for the school as an institution and for its staff as individuals when they embark on, and successfully conclude, effective beneficial changes.[Only then might] there need to be sanctions to hand if they do not.*

(Penzer & Allen, 2011:11).

3.4 The cardinal rule of accountability is to hold people accountable for what they can control (Harris 2010). The consequences of school-level performance indicators are determined by the interaction between four broad groups of factors:

1. The nature of the information published and the validity of the measures on which the judgement is made;
2. The way in which the information is fed back to intended users, for example, whether it is accompanied by an explanation of what the data means, or whether complicated indicators are used without clear discussion;
3. the nature of the local school market and whether, for example, an alternative school exists for parents if their local school does not appear to perform well;
4. the extent to which government seek to take action to correct poorly performing schools (Visscher (2001).

The interaction of these four groups of factors can generate three categories of problems:

- *technical or analytical issues around the construction and aggregation of performance indicators;*

- *usability issues* related to the clarity, utility and comprehensibility of the data presented to service users; and
- *political or societal issues*, linked to the broader implications of the use of performance indicators on public service provision. We look at each of these sets of issues in turn.

(Foley & Goldstein 2012: 23).

3.5 Insufficient base-line measures: It is a well known that the cause of ‘differentials in educational performance lie largely outside schools and the classroom’ (Purvis *et al.*, 2011, 7) and that affecting change in schools can prove futile against the culture of the surrounding community, its attitudes, values, traditions and beliefs (Vollmer, 2010 in MacBeath, 2012: P42). Over three decades of research into school effectiveness and improvement in a range of countries (Sammons, P. 2007) highlights that the factors which most influence variations in pupil attainment are:

- Individual characteristics (age, birth weight, gender);
- Family socio-economic characteristics (particularly family structure, parental background: qualification levels, health, socio-economic status, employment/unemployment, and income level);
- Family cultural capital, (particularly the powerful impact of the child’s home learning environment, especially in the early years, as a predictor of attainment);
- Community and societal characteristics (neighbourhood context, cultural expectations, social structural divisions, especially in relation to social class); and last of all, educational experiences, where teachers and schools can add value.

Of these, the two most influential factors are socio-economic status and the quality of parenting. There is complete agreement across the research field that, *the interplay of school with family, neighbourhood and community needs to be taken into account in any judgement about teaching quality and effect* (MacBeath, 2012: 45).

3.6 Insufficient account of family and community factors: Since very little account is taken of the factors that influence variations in pupil attainment, the ‘blame’ for failure to overcome family and community cultural capital tend to be placed at ‘*the door of schools and on the shoulders of teachers*’ (ibid P21).

‘[Children] arrive at the classroom door with vastly different early childhood experiences and levels of readiness for school. For example, at the very beginning of kindergarten, high-income children have average test scores that are 60 percent higher than low-income children. Schools cannot have caused these “starting-gate inequalities,” because most students haven’t set foot in a classroom before.... Yet the inequalities are so large and persistent that even effective schools cannot completely overcome them. Non-school factors continue to influence children as they progress through school. These factors are outside the control of schools and failing to account for them, as attainment measures do, amounts to violating the Cardinal Rule of Accountability’ (Harris, 2010:3).

‘If children are not succeeding, it is obviously the fault of teachers, their low expectations or incompetence, the malign influence of unions on teachers, or failures of leadership to raise standards... There may be a nodding acknowledgement to social and economic factors but successive governments have regarded any reference to these as excuses and insisted that background factors can be overcome

by good teachers and inspirational leaders. This ignores the growing body of evidence about the crucial influences [for example, during pregnancy of the effects of smoking, drugs and foetal alcohol syndrome, poor stimulus and bonding in the first nine months after conception and poor child care in the early years] that are beyond the repair of even the most enlightened teacher' (ibid, 21-2).

'The task facing teachers and other professionals who work with children from disadvantaged backgrounds is, for these reasons, much more challenging now than it was a generation ago'. (Alexander and Hargreaves, 2007, 3).

3.7 Insufficient account of the peer effect. The major student achievement problem in the Northern Ireland schooling system is not the overall performance of pupils, but the levels of equity within that performance. Selection and increased school choice policies are correlated with an increase in the differentiation of pupils according to social background. The consistent message arising from a decade of international comparisons is that selective systems create wider differentials of achievement by separating young people from disadvantaged backgrounds from the positive aspirations of their 'better-off' peers at a vulnerable age.

The power of the 'compositional' or peer effect has been shown to be one of the strongest determining factors of achievement and attitude... The weaker the social and intellectual capital in the family, the stronger the influences of peers, which tends to find its level at the lowest common denominator.... Dominant forces in childhood and adolescence can be ascribed to 'significant others' who shape values and character often more insidiously and powerfully than parents and teachers which play out in school and classroom life on the one hand and in street and neighbourhood culture on the other hand.
(MacBeath, 2012, 47)

Thus in Northern Ireland, a 20% underachievement problem at primary level doubles to a more serious 40% problem at post-primary level.

3.8 Pupil 'compliance without engagement': The Northern Ireland Cohort Study (1996-2002) of 3,000 pupils over 7 years revealed that very many pupils viewed school as only relevant for jumping hurdles to pass exams, but of little relevance to real life, leading to a culture, even among high-performing grammar school pupils, of 'compliance without engagement' (Harland et al, 2002). As a result of their feedback and significant consultation with teachers and wider society, the Revised Northern Ireland Curriculum was introduced in 2007. Unfortunately the assessment and examination system has not been sufficiently aligned with the revised curriculum, inhibiting real changes in teaching and learning. The following quotation from Ravitch (2010) sums up the impact of the accountability agenda upon political and public perceptions of the responsibility of schools and teachers in the United States:

It would be good if our nation's education leaders recognized that teachers are not solely responsible for student test scores. Other influences matter, including the students' effort, the family's encouragement, the effects of popular culture, and the influence of poverty... Since we can't fire poverty,

we can't fire students, and we can't fire families, all that is left is to fire are teachers. (Ravitch 2010)

Embargoed until 27 September 2013

4: Gaps in the ETI review process

4.1 Pre-determined policy measures: ETI uses performance measures that are defined within Programme for Government targets and therefore in doing so they are adhering to pre-determined DE and Education Committee policy requirements. As suggested earlier in this submission, many of the limitations of these measures have not been fully explored and a great deal more analysis needs to be undertaken of the nature and reliability of the measures themselves and of associated effect sizes to ensure that the conclusions drawn from the use of flawed measures is robust. Additionally, it is recognised that the Northern Ireland education system has been undergoing a period of unprecedented change at a time of major financial constraint and that planned change has been slowed by democratic scrutiny. Thus gaps in the ETI review process may be exacerbated by gaps elsewhere which are not of ETI's making.

4.2 Analysis of performance measures and the way in which they are used: There is currently a great deal of scepticism amongst teaching professionals about the expanding role of performance monitoring (*Wiggins and Tymms, 2002*). Teachers working in areas of high social and economic disadvantage in particular often feel that, even with more contextualised data, that performance monitoring fails to provide an accurate reflection of institutional quality. The problem they say resides not with the performance measures themselves, but with the way that these measures are often used.

4.3 Lack of analysis of effect sizes and correction for student intake: School quality is the degree to which a school scores better than other schools, corrected for student intake characteristics. An effect size is no more than a relative measure subject to considerable margins of error. Researchers are cautious about quantifying the language of effects, pointing out that statistical differences are often marginal and tend to conceal more than they reveal. This, however, has not prevented the term 'effective' (a statistical term borrowed from economics) with the perception of 'good' (which is a value judgement) (*MacBeath, 2-12: 44*).

4.4 Over-estimation of the school effect: The comparative importance of various factors in influencing pupil performance has been researched for many years and within a number of research traditions. An important categorisation is between factors internal and external to the school. The larger the sample under investigation, the smaller the influence of school factors has been found to be. There is a high degree of agreement between researchers from different traditions that approximately 85% of the variation in pupil achievement is due to factors external to the school. As a counter to the fatalism which might derive from such findings, the school improvement movement in Britain sought to identify characteristics of effective schools, on the assumption that the improvement in teaching and learning techniques would raise overall achievement. However, a review of this work by one of its most eminent practitioners (*Mortimer, 1998*) also confirmed that such internal factors were much less influential than external ones. A review of related studies (*Chevalier, Dolton and Levacic, 2005; Cassen and Kingdon, 2007*) also concluded that the variance in pupil performance due to schools ranged between 5% and 18%. The major gap in the DE policy of 'Every School a Good School' and in the ETI school review process is, therefore, the lack of analysis of effect sizes, which may be much less significant than implied, and the lack of appropriate correction for student intake.

5: Gaps in DE and ELB support

5.1 Alignment with a constructive support infrastructure: Matthews and Sammons (2004, p. 164) identify the following main conditions for the implementation of recommendations from external school evaluation:

“understanding and acceptance of the findings by the provider; leadership that can generate and implement strategy for implementing inspection outcomes, including effective action planning; identification of any resources and support needed to effect improvement; and planned external follow-up to assess the progress made”
(OECD 2013a: 390)

In order for external school evaluation to be effective therefore there needs to be a supportive infra-structure coming in alongside or behind it.

5.2 DE Strategy setting: It is recognised that the Northern Ireland education system has been undergoing a period of unprecedented change at a time of major financial constraint and that planned change has been slowed by democratic scrutiny. Nevertheless, the Department appears to be excessively engaged in short-term operational issues, tightly monitoring the performance of its Non-Departmental Public Bodies. It needs to create space to tackle more of the key strategic issues and develop a long-term strategy for education in light of foreseeable resource constraints. The development and implementation of key educational policies is too slow, for example, the Review of Teacher Education has been ongoing for over a decade. There is an urgent need for the development of a coherent professional development framework for teachers and for widespread consultation on the shape of a future advisory and support structure.

5.3 Gaps between policy direction and support capacity: ELB support is now targeted almost exclusively on schools identified by the ETI and management authorities as failing to meet the required academic standards. This approach has emerged, not as part of any strategic shift in the thinking but, rather, as a consequence of the vacancy control policy related to ESA. Schools which have not been identified as failing academically are now struggling to effect meaningful change due to shortfalls in expertise within their own staff and a shortage of finance to purchase this expertise from outside, even if it was available. Many ELB officers report that their task, post-inspection, is as much about restoring confidence and motivation after inspection trauma, as improving teaching and learning. As referred to earlier, the current consultation for changes to the Formal Intervention Process make reference *‘schools in formal intervention ... having had two years of tailored support’*. The proposals go on to suggest that: *‘Any school not improving to at least a ‘good’ evaluation by the time of its follow-up inspection will be placed in formal intervention, provided with tailored support and given a further 12 months to improve to at least a ‘good’ evaluation or further action will be considered’* (DE, June 2013).

These proposals assume a capacity within the support services to prove the level of tailored support suggested. The reality of shrinkage in the CASS service and the experience of schools would suggest that policy development is at variance with planning. Indeed, evidence over the past 6 years or more would suggest that the one

consistent characteristic of Northern Ireland's approach to educational change management is that written policy directives are issued from the centre and then schools are expected to interpret and implement them without any tangible sustained support to do so.

5.4 Gaps in provision for teacher professional development: The limitations of the current narrow focus on struggling schools is already manifest. To reiterate what has been said previously, the proposals are at variance with copious research evidence which highlights the length of time, range of measures and nature and depth of support needed to bring about a genuine and sustainable change in the ethos and culture of struggling schools.

It is now accepted internationally that '*Change is based on building the expertise of the profession*' (Hayward et al, 2012) and that '*No education system can rise above the quality of its teachers*' (McKinsey Report, OECD, 2007). While there is a growing acceptance that the best professional development is school-centred and focused on the core skills of better teaching, learning and assessment, this will not happen overnight or without a proper strategy and support. The Independent Review of the Common Funding Scheme has recommended that:

The proposed regional school development service should assign a central role to supporting peer support at area and school level, providing greater opportunities for teachers to work together in sharing good practice, while also able to draw on external expert advice, where needed.

(Salisbury, 2013: viii)

Initiatives have already been established by small clusters of schools, drawing on research insights from the highly effective 'London Challenge' strategy. There is a rich opportunity to capture, support and cultivate their innovative work, and establish collaboration networks among these teachers and students to build capacity and models for practice.

5.5 Gaps in the policy drive towards 21st Century learning: Concerns are increasingly being expressed about preparing young people for what has become known as the 'knowledge era', reflecting the exponential growth, ease of access to, and speed of flow, of all kinds of knowledge via the world-wide web and social media. This knowledge revolution has had a profound impact on our access to knowledge and our potential to learn. *The Global partnership on New Pedagogies for Deep Learning* advances the proposition that our education systems need new policies, measures and evidence-based pedagogical models to enable learning relevant for the knowledge-based, globalized era.

The crisis — and there is no other word for it — in public schooling is a function of the interaction of an enormous push-pull dynamic. The push factor is that students find schooling increasingly boring as they proceed across the grades. Studies from many countries show that less than 40% of upper secondary students are intellectually engaged (Jenkins, 2013; Willms et al., 2009). And, not unrelated, signs of teacher frustration are growing. Teachers and students are psychologically if not literally being pushed out of school. Education under these terms needs to be radically re-thought (Fullan & Langworthy, 2013: 7)

A recent OECD report (2011) highlights how already high-performing countries have taken action *'to ensure that 21st century skills that are considered important become valued in the education system'* (OECD, 2011: 19). The outcomes of these changes in assessment policy are believed to be already bearing fruit a decade later (ibid.). A survey of seventeen countries (OECD, 2009) found that, while most countries refer to 21st century skills and competencies in their guidelines for compulsory education, few specific definitions of these skills and competencies exist at national or regional level and virtually no clear formative or summative assessment policies for these skills. The only evaluation regarding their teaching is often left to external inspectors as part of their whole school audits (Ananiadou & Claro, 2009: 4). Northern Ireland is an exception, having put in place definitions of these skills and competencies and valuable support materials since 2003, as well as support through for example, the Accelerating Children's Thinking Skills (ACTS) Project since 1996 (see McGuinness references). There are however, gaps in system-wide support, assessment and examination.

6. Alternative approaches in other jurisdictions

6.1 Alternative approaches: All countries want their education system to be as good as possible and school inspection, which inevitably comes at a price, should be able to demonstrate that it is worth the cost. It has the potential to deliver on two fronts, accountability and improvement. The balance between a focus on accountability and a focus on improvement varies from one country to another.

The range of international evidence cited below illustrated the extent to which other countries are engaging with issues similar to those identified by the education committee inquiry; that this inquiry is a healthy reflection of what we need to be doing constantly in relation to major education policies; and that the recommendations offered are meant to be positive and enabling in evolving towards a system that engages all partners in a clear shared moral purpose of doing the best for our young people.

6.2 Finland has been heralded as one of the world's most successful education systems ever since the Organisation for Economic Co-operation and Development (OECD) began publishing international league tables more than a decade ago.

Prior to 2000 Finland rarely appeared on anyone's list of the world's most advanced nations, let alone education systems. Many young people were leaving the system relatively early, and Finland's performance was never better than average on five different international mathematics or science assessments of the International Association for the Evaluation of Educational Achievement (IEA) between 1962 and 1999. However, over the past decade Finland has been a major international leader in education. It has consistently ranked in the top tier of countries in all PISA assessments since 2000, and its performance has been notable for its remarkable consistency across schools. No other country has so little variation in outcomes between schools, and the gap within schools between the top and bottom achieving students is extraordinarily modest. Finnish schools seem to serve all students well, regardless of family background or socio-economic status. (OECD, 2012: 94)

In the mid 1990's fiscal control of schools was moved to the districts, spending was entirely devolved to municipalities and state school inspections were eliminated. Schools are accountable for spending to municipal and regional offices, who are also responsible for scrutinising a school's examination performance, although results are not usually made public (Sahlberg, 2012: 27-30) Instead of inspection, teachers undergo a yearly evaluation with the school leadership. Pupils and parents are both offered questionnaires. The Education Evaluation Council works with Government to provide schools with support to evaluate their own performance. The aim of evaluation is seen as gathering and analysis of information to develop education generally, rather than to direct improvement in individual settings—supporting the focus on a fair and balanced system, rather than on changing individual school practices. A sample-based educational evaluation system is used to help monitor the overall performance of the educational system. Feedback is given to participating schools to inform changes to teaching (the same type of system is used in Scotland) (ibid).

One of the main reasons for Finland's performance is its focus on improving equity – not achievement or results. The country has invested fairly and more heavily in schools within disadvantaged communities and insisted the best way to provide equal educa-

tional opportunities for all is through public schools. Between 1970 and 1981 a comprehensive system was introduced, which ended the previous divisions between grammar and technical and vocational schooling. All pupils of 7 - 16 years of age are educated in local schools, without any kind of streaming. The number of students in a class is also much smaller than other countries, normally between 15 and 25. Schools alone are responsible for assessing student achievement and there are no examinations until the age of 16, after which students choose to attend either general or vocational schools. A high-performing school is seen as one where all students perform beyond what would be expected based on their socio-economic background.

Finland places a very high value on education, which is supported by a very strong focus on teacher recruitment, training and development (NESC, 2012: 56). Teaching is a much-admired profession, with only around 12% of applicants being accepted for training, and there is very little central prescription. All teachers and administrators must have high academic credentials and must update their knowledge and skills continuously. Finland invests 30 times more funds in the professional development of teachers and administrators than in evaluating the performance of students and schools, including testing. (This ratio is the opposite of many countries with testing-intensive education systems, where the majority of funding goes to evaluation and standardized testing). In 2012, for example, the state allocated more than \$30 million to the professional development of teachers and administrators. Finnish teachers and administrators each spend, on average, seven days annually in professional development activities; half of that is on their personal time.

Finland also places a strong element of child well-being and care. Schools maintain strong support systems for all learners – healthful nutrition, health services, psychological counselling and student guidance are normal practice. Finland's special education system is also cited as a major reason for the country's world-class ranking. A core principle is early identification of learning difficulties before a child even starts school. Regular free assessments of the physical, mental and social development of newborn and pre-school children is provided by a network of child health clinics which are located across the country. Multi-professional teams comprising a public health nurse, medical doctor, speech therapist and a psychologist, if necessary, do the evaluations. These checks are carried out according to national guidelines that specify the timetable for child well-being checks. All schools have 'welfare boards' concerned with the broader well-being of students. Particular attention is paid to children who need more help becoming successful, compared to other students while allowing the student to remain in class with his/her peers (ibid: 28).

6.3 Scotland's inspection service increasingly emphasises a two-way collaborative approach, aiming to work with staff in a “*constructive, positive and professional manner*” (ibid). Several changes have happened over the past 2-3 years, the most significant of which is the much closer alignment of the Scottish Inspection Service with the school support service within a new amalgamated structure, under the banner of 'Education Scotland'. Revised inspection arrangements place a stronger focus on: school self-evaluation; analysis of a wider range of outcomes; and a wider range of “continuing engagement” or “improvement visits” carried out by non-HMI development officers and/or senior education officers who work within Education Scotland. (Such visits can involve HMI from time to time). This engagement aims to offer support more directly or to capture and publish innovative or creative work noted on inspection. It also includes use of: *The PRAISE self evaluation framework* which is used after each inspection to evaluate HMI performance on inspection at individual and team level; *A New Scottish*

Benchmarking approach to assessing added value which takes into account a wider range of qualifications and learning programmes, including post-school participation; and *Scottish School Improvement Partnerships* programme led by Education Scotland working with local authorities and professional associations have been set up to tackle the link between socio-economic deprivation and low educational attainment. The absence of centrally designed and monitored end of key stage standard assessments in Scotland *'has meant that data gathering and use is much less intensive within the Scottish system than in England'* (Ozga et al., 2009: 20). Data has been found to play a much less significant role in influence 'the government of education in Scotland'. Although it was important *'it was being actively used more to support self evaluation and hence self government'* (ibid. 22). A survey of almost one thousand teachers in Scotland and England found that:

'Teachers in Scotland and England are more positive about Quality Assurance processes over which they have some degree of control, rather than those that are top down; Teachers in Scotland highlight the importance of self regulation and feel less regulated 'from above' than do their English colleagues (ibid.,.)

Interestingly, however, one of the less expected findings of an earlier survey of teachers in England and Scotland (Wiggins and Tymms 2002) was that Scottish primary schools (whose results are not publicised in league tables) felt under greater pressure to meet performance targets than teachers in England. In addition, schools deemed by performance monitoring to be 'good' were just as likely to find performance indicators problematic as 'poor' schools.

There was agreement across teachers in both Scotland and England that external, standardised performance indicators were not particularly good at judging overall performance and that internal systems controlled by schools themselves would be more effective (Foley and Goldstein 2012: 29). The overwhelming impression from research in England is that *'the education system has become so demanding and so data heavy that its intelligent use is compromised'* (Ozga et al., 2009: 21). This finding endorses the Finnish approach to inspection and accountability

6.4 Singapore emphasised accountability in their inspection service in the 1980s and 1990s but found that while it contributed to the improvement of academic performance over the years, it led schools to focus too much on examination results, with little room or motivation for schools to take responsibility for bettering themselves. A new system was introduced in 2000, based on school self evaluation, with a system of rewards to encourage, motivate and reward for successful schools as an integral part of its school excellence model.

6.5 New Zealand makes use of a socio-economic 'decile system' which informs school base-lining, value added, resource allocation and other services: Census information is used to place schools into ten deciles. Student addresses are assigned to the smallest Census areas, called mesh-blocks, which contain about 50 households. The mesh-block is examined against five socio-economic factors drawn from census data, including: *parental educational qualifications; parental occupation; household occupancy; household income; and Income support.*

Schools are ranked in relation to every other school for each of the five factors. Each school receives a score according to the percentile that they fall into. The five scores for

each school are added together (without any weightings) to give a total. This total gives the overall standing of a school in relation to all other schools in the country, enabling the Ministry to place schools into ten groups, called deciles, each having the same number of schools. A school's decile rating informs resource allocation and other services.

Analogous contextual information— with the exception of household income – is available in Northern Ireland. There are potential linkages here to the recommendations contained in the Salisbury report (2013).

- 6.6 Hampshire (England):** In an experiment in one English local authority (Hampshire) in the late 1990s, value-added estimates were introduced for primary schools and utilised by the authority and head teachers as an unpublished 'screening device' and a 'school improvement' tool. The detailed yearly scores were fed back in confidence to schools as one item of information within an inspectorial system so that it could be used alongside other information (*Yang et al., 1999 in Foley & Goldstein, 2012: 28*).
- 6.7 In Germany** inspection reports are confidential to the inspectorate and the institution inspected and it is generally accepted that the prime responsibility for ensuring that a school provides a good standard of education rests with the school itself, and not with the inspectorate (*Penzer & Allen, 2011*).
- 6.8 In Hong Kong** a school can decide whether or not to make its report public but, having decided to do so, it cannot reverse the decision next time it is inspected (*ibid*). Hong Kong has recently developed an External School Review approach which has been designed to be 'improvement-oriented'. The Education and Manpower Bureau of the Government of Hong Kong has produced an 'On-line Interactive Resource on Enhancing School Improvement through School Self Evaluation and External School Review' (*ibid: 12*) (*See also 7.4. below*).
- 6.9 In Spain** inspection does not 'aspire to classify schools but [rather] to help them know themselves more deeply' (*SICI European Inspectorates' Profiles 2009: Spain*).
- 6.10 In Denmark** 'very infrequent' inspection is regarded as all that is needed to check and to keep a school accountable or focused on the provision of excellent education *SICI European Inspectorates' Profiles: Denmark, 2009*).
- 6.11 In Slovakia** inspectors provide in-service training for teachers.
- 6.12 The Australian Capital Territory** is in the process of introducing a well structured periodically validated self evaluation system

7: Recommendations on approaches to school improvement

7.1 Devise a supportive stream-lined evaluation process: A recent OECD Review of Evaluation and Assessment in New Zealand (OECD 2012) highlights the need to provide a coherent framework for evaluation and assessment approaches at student, teacher, school and system level, outlining how the different elements are interrelated and describing for each individual component:

- (1) the purpose and goals of the process;
- (2) evidence-based principles of effective practice;
- (3) available tools and reference standards for implementation; and
- (4) reporting requirements and/or intended use of results.

The process of developing such a framework document of evaluation and assessment levels would provide an opportunity to analyse the various linkages between different components and identify missing links and articulations in need of strengthening.

Whatever the future process, clear guidance needs to be provided on data requirements; constructive challenge should be allowed; reporting timescales should be reduced to a maximum of 8 weeks, as in Scotland, but avoiding the OFSTED 15 day schedule (which is inadequate for appropriate reflection).

7.2: Closely align evaluation and support services: Inspection results need to be presented in ways that recognise the real constraints on action that any school faces, followed by sustained access to good professional advice and support (and improvement tools) when considering, planning and implementing the changes it needs to make over time.

Hong Kong initiated its system of self-evaluation and external review a decade ago. It was accompanied from the start by a longitudinal external evaluation and consultancy. The development of school self-evaluation (SSE) and external school review (ESR) followed the well known pattern (Rogers, 1962) of innovators, early adopters, early majority, late majority and laggards. The key to the diffusion of innovation was to learn from the innovators and early adopters and from how the wave of change is enabled to move through the system. Drawing on the experience and expertise of the leading-edge schools, principals and school staff were engaged as ambassadors and as conference and workshop leaders, as members of external review teams and as foci for good practice case studies. The development of an on-line interactive resource gives schools access to review tools and to testimonies from students, parents, teachers and principals discussing challenges and achievements. A revised version in 2010 included a range of classroom lessons with accompanying observation and evaluation questions to illustrate how self-evaluation can be embedded in day-to-day practice. Source: MacBeath (2009 in OECD 2012: 104)

In Scotland, provision of such support is now fully built into the inspection system. Consideration might be given to replicating the Hong King and Scottish model.

7.3 Widen the composition of any future inspection/evaluation service:

The OECD highlights that a key factor in the effectiveness of evaluation 'on whether those who evaluate and those who use evaluation results at the different levels of the system have the appropriate competencies (OECD 2012:133). The perception of the

composition of the inspectorate is that it has insufficient complement of people with actual experience of leading schools and that the balance of background is more heavily weighted towards the grammar school sector. The recent recruitment in June 2013 of 200+ Principals and middle managers as Associate Assessors is welcomed.

To ensure that inspectors maintain credibility with schools there is a view that the number of permanent inspectors should be reduced to a smaller core team supported by serving teachers and principals seconded as Associate Assessors either for a specified number of years or on a part-time basis 2/3 days per week. In addition, it is suggested that inspectors should be seconded on a periodic basis to school management teams for significant periods to refresh their authentic experiential awareness of the challenges of the environments they evaluate. It is also felt that external school evaluation should focus less on inspectors being the arbiters of the quality of subject learning and teaching and more on the evaluation of school leadership teams as the internal arbiters of quality.

7.4 Strengthen the focus on school self-evaluation: Perceptive self evaluation is known to be the best and most secure foundation for school improvement. The requirement for each school itself to reflect on the quality of its work has great potential when it is done seriously and honestly, and it does not depend on inspection for its effectiveness. A recent OECD review (2012) recognises that: *'schools know their contexts best and allows professionals to adopt a diversity of evaluation and assessment practices, thereby creating conditions for innovation and system evolution'* (OECD 2012:133).

7.4 Strengthen the focus on school self-evaluation exemplification and tools: More guidance and case-study evidence could be offered about the documentation and evidence which schools should provide and more resources need to be allocated to strengthening and supporting robust school self-evaluation so that schools themselves are the main agents of change and improvement.

7.5 Strengthen the focus on school leadership development: At the same time it also recognises the complexity and breadth of school leaders' and teachers' responsibilities regarding evaluation and assessment, requiring a new set of skills which many may not have acquired in their initial training (Ibid).

In the context of self-management, individual schools can be relatively isolated and may have limited opportunities for learning from effective practice from across the region or the country. Continuing to build the capacity of teachers, school leaders and Boards of Trustees for effective evaluation and assessment remains a priority.

7.6 Strengthen the focus on Board of Governor training and development Boards of Governors and Trustees also play a key role in planning, reporting and self-review tasks but their preparedness and capacity to fulfil this role is highly variable. There may be a need to remunerate of Boards of Governors to attract high calibre recruits to this important role who are prepared to invest the considerable time needed to undertake this challenging role.

7.7 Research and disseminate best practice: Inspection should be influenced, at least in part, by its role as a system-wide research tool. Decisions about which schools to inspect should be determined partly by a view about which have features from which others can learn, so that insights into best practice are gathered and disseminated widely

and persuasively through in-service events and the publication of thematic insights into what has been found to work (*Prender and Allen 2011*).

The 'Sustaining Improvement Inspection' pilot work undertaken in June 2013 in primary schools, to be followed up in this autumn in special schools and in May 2014 in post primary) is to be welcomed. This work allows schools where provision previously has been evaluated as very good or outstanding to demonstrate how it has developed its capacity for further improvement. These schools are provided with the opportunity to take greater control over the inspection process by identifying priorities within the school's Development Plan where they school feel they have made advances since the baseline inspection. The potential to extend this emphasis on partnership should perhaps be an element of all inspections to allow all schools to show-case their strengths and to identify for themselves initially the areas for further development.

8: Recommendations to improve value-added calculations

8.1 Utilise socio-economic base-line data: explore the potential to use NISRA census information to calculate the socio-economic intake of schools to:

- stratify schools (into deciles) according to the socio-economic intake;
- map school/pupil catchment areas and journeys;
- allocate resources more effectively to target social need;
- calculate value-added on the basis of better base-line data (see also recommendation 2 about base-lining pupil's productive language on entry to school).

8.2 Utilise school catchment analytics: Develop a GIS system (geographic information system) to capture, store, manipulate, analyze, manage, and 'map' all types of statistical analysis and databases to produce detailed educational analytics; to compare actual with expected school catchments and to consider daily spatial moves for different groups by gender, FSM status, and social class and so on. The data could be collected through existing administrative procedures or using the 2011 Census to calculate school catchments and pupil journeys to school. Spatial information of this kind could make a useful addition to a multi-level framework that includes individual and household level information.

8.3 Utilise educational base-line data: Undertake oracy assessments (productive language on school entry) on entry to school as a key determinant of ability to learn. There are a whole range of baseline measures that might be used to assess spoken language on entry. One well-known example is The Renfrew Bus Story (RBS) - a short screening assessment of receptive and expressive oral language for young children age 3 years to 6 years 11 months. Using 'narrative re-tell', the RBS provides a quantitative and qualitative assessment of each child's oral language skills based on rich language data. It has been shown to be able to identify children with language impairments, as well as to be predictive of later language and academic skill (Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998). The assessment is quick to administer and enjoyable for children, using a technique that is familiar to most children – storytelling. Other comparable examples might be researched and trialled for suitability.

8.4 Utilise sampling for system monitoring: Politicians and DE only need to know how the system is performing generally – not at individual school or pupil level. A system relying on 'light sampling' of 10% of schools will provide stable and robust information for the purposes of accountability and policy formation. Recent advice in Scotland (Hayward et al., 2012) endorses this and suggests the potential for enhanced targeted sampling in areas where there are concerns, to provide robust and independent data.

8.5 Utilise international data critically and objectively for system monitoring:

The Department already has a wealth of quantitative and qualitative sampled data from international testing, together (PIRLS, TIMSS and PISA) with detailed qualitative information on the sampled population. This needs to be properly analysed and fed back to participating schools as part of the improvement process – as well as a broader comparative measure for the whole system. Care needs to be taken in data analysis and reporting to avoid simplistic rank ordering and the tendency to misinterpret

significance and to overlook the limitations of this data, not least the difficulties of cross-cultural comparison.

8.6 Develop models of value-added: The goal is to create a measure of performance that fits the Cardinal Rule of Accountability. Value-added does this in two ways: 1) taking into account where students start when they first walk into school and 2) comparing schools that are similar in terms of measurable school resources or, more specifically, using a prediction approach that gives a reasonable head start to schools that operate with fewer resources, making more reasonable comparisons possible.

Borooah & Knox (2013) have already developed a workable model of value-added and applied it in Northern Ireland which identified those schools which add most educational value to their students.

1. *Using official data gathered through the viability audits, the Education and Library Boards and the Department of Education, we examine those **factors which best explain education performance** in primary and post primary schools.*
2. *As a result of understanding the relationships between those variables which explain education performance we derive equations (primary and post primary) which allows us to predict, within a range of significance levels, what results schools should achieve, given their circumstances. We can then examine the difference between actual results achieved against those which we can predict. This allows us to say whether a school is '**over-performing**' or '**under-performing**'.*
3. *The corollary of point 2 above is that we can estimate the **value which teachers add** to their pupils' performance through good teaching, leadership, expertise and so on. We can also compare the way in which the Department of Education currently measures school performance with our own proposals.*
4. *Given our specific interest in shared education and the educational benefits associated with its provision, this approach will also allow us to compare the quality of education performance of those **schools engaged in cross-community collaboration** with those which operate as discrete units.*

The outcomes of this model to calculate the value-added by schools in Northern Ireland makes startling and salutary reading. As policymakers move forward toward productive experimentation with value-added, they should avoid becoming over-confident in the ability of these measures to accurately distinguish performance with any degree of nuance. Value-added measures have potential, but we cannot lose sight of their limitations or of their larger purpose: measuring performance in a way that drives genuine improvement in teaching and learning. (Harris, 2010: p10)

9: Recommendations for broader measures of achievement

9.1 Separate teacher assessment from accountability - Teacher assessment for learning only: The clear recommendation from assessment experts (*The Assessment Reform Group; Gipps; Tymms etc*) is that processes of teaching, learning and assessment should focus on improving learning only and should not be over-burdened with bureaucracy or exposed to potential manipulation for accountability purposes. Virtually all of the research into the use of teacher assessment (and levels of attainment in particular) advises against the use of numerical assessment outcomes for target setting and accountability purposes. Instead, it advises that school evaluation should be disentangled from accountability, and that monitoring standards over time should operate outside an accountability framework, otherwise the accountability pressures distort the processes of learning and the outcome data.

9.2 Develop and use wider indicators: Experts in the field have called for the gathering of 'multiple indicators of standards by combining information of different kinds' to 'enable progress in all important learning goals to be facilitated and reported' (Assessment Reform Group, 2008: 5; Tymms and Merrill, 2007: 14; Gardner et al, 2008: 5) and, to inform decisions about expenditure, the allocation of time and resources and to provide potential 'value-added' insights.

The British Academy inquiry into school measurement has called for 'Ways to rely less on a small number of indicators [...], as well as those which cover more aspects of learning' (Foley & Goldstein, 2012: 11, *British Academy Policy Centre*). The Director of the CBI, John Cridland, in a recent speech to launch the CBI's 'First Steps' called for:

'A rigorous and demanding accountability regime that assesses schools' performance on a wider basis than the narrow measure of exams. We need to define 'a new performance standard based on the whole person'...and ' a shift to new style [inspection] reports which will assess both academic rigour and the broader behaviours and attitudes that young people need to get on in life'.

CBI First Steps Report, 2012

The following suggestions, which are not exhaustive, illustrate the potential for improving the range and quality of data that might be garnered to facilitate a more sophisticated analysis of the value-added by schools.

9.3 Limit the use of standardised testing in schools for diagnostic and formative purposes and insights into progress: Assessment experts, examiners and statisticians argue that any test is only a short snapshot of a pupil's potential performance at any given time, which is subject to unavoidable errors and therefore needs to be treated with caution and sensitivity. A well-designed standardized test can, however, offer a relatively reliable way of estimating how an individual pupil has performed on a specific day, based on the population as a whole. Careful analysis of detailed feedback from such tests, over time, can help to identify individual learning difficulties or areas of misunderstanding that help teachers to target individual pupil learning needs. However, the use of such tests for summative accountability purposes runs the inevitable risk of teachers being pressurised to teach to the test and therefore corrupting the diagnostic and formative properties of the results. Sensitive analysis of pupil 'percentile ranking' or 'stanine' characteristics over time by comparison to baseline characteristics could be used to provide insight into individual progress over time, with

the caveat that pupils do not all progress at the same rate and may be subject to 'learning spurts' in the same way as they are subject to 'growth spurts'.

9.4 Develop more appropriate statistical analysis models: The recent British Academy report on 'Measuring Success' has called for:

More appropriate statistical analysis models should be used to describe institutional differences that allow for differential performance for different groups of students. In particular, there should be a shift away from the comparison of individual institutions towards research that helps to identify modifiable factors that appear to be related to good performance.

Foley, B. & Goldstein, H., (2012)

9.5 Utilise attitudinal data sensitively. Attitudinal surveys are a potential proxy for actual measurement. There is a well-established correlation (for example, in PIRLS & TIMSS 2011) between being a 'motivated or somewhat motivated reader' and between those who 'liked learning Mathematics/Like Learning Science bands' and the highest achievement in the subject. The better readers, for example, were also the more confident readers. The pupils who reported being most confident in mathematics and science were also the pupils who had higher average achievement scores. If we could teach towards motivation and enjoyment then achievement (and life-long learning dispositions) would follow. There are also a number of measures of social, emotional and personal well-being which might be investigated (for example the ACER scale) and of creativity and dispositions to learn (Bristol University and Antidote) which also could be considered in any holistic assessment of a quality education.

9.6 Maintain a proportionate focus on the 'old' literacies: The relentless focus on literacy and numeracy, while important, ignores the evidence that 80% of the school population is doing relatively well (Tymms, 2004) and that pupils are in danger of being turned off by too much drill and lack of creativity in education. The proportions of pupils in Northern Ireland who do not like reading was higher than the international mean (Sturman et al., 2012).

9.7 Increase the focus on 'new' literacies: The European Commission has highlighted that 'the key challenge for education systems in many Member States is the assessment of these competences. Assessment is one of the most powerful influences on teaching and learning but it tends to put too much emphasis on subject knowledge, and less on skills and attitudes, and to neglect altogether the increasingly important cross-curricular competences such as learning to learn or entrepreneurship. (European Commission, 2012)

There is a need for a profound shift in conceptions of learning and knowledge 'rigour' that moves away from memorisation of traditional knowledge towards more creative conceptions of learning associated with research, information management, knowledge construction and creativity across traditional subject boundaries. In other words our main educational focus should be on the Northern Ireland Framework for Thinking Skills and Personal Capabilities (CCEA/DE 2006), which in turn require more complex forms of assessment that are not readily achieved through traditional examinations.

9.8 Research and develop innovative 21st Century assessment and examining:

The recently published *OECD Review of Evaluation and Assessment in Education: Synergies for Better Learning - An International perspective on evaluation and assessment* (April 2013) recommends that countries should ‘align assessment with educational goals, designing fit-for-purpose evaluations and assessments, and ensuring a clear understanding of educational goals by school agents’ (OECD, 2013).

The Global partnership on New Pedagogies for Deep Learning (2012) highlights that one of the fundamental barriers to the development of 21st Century skills is the inadequate dissemination of new pedagogical models that foster deep learning and the inadequate development of ways of measuring and assessing deep learning.

Proactive research should be commissioned, possibly from the OECD or from leading international assessment organizations (for example the Australian Council for Educational Research - ACER) to assist CCEA in identifying, trialling and evaluating innovative 21st Century assessment and examinations mechanisms to move the field forward. The opportunity should be taken in the review of GCSEs and A levels to develop new qualifications for Northern Ireland to be taken at the appropriate age (17-18) which reflect the 21st century needs of young people, the economy, employment and life-fulfilment.

9.9. Assess 21st Century thinking skills and capabilities: The European Commission has recently highlighted the key challenge for education systems in many member states, is the assessment of 21st century skills and competences. The OECD has recommended that, rather than testing the content of learning, assessment should focus on cognitive skills such as problem-solving, communicating and reasoning which would give teachers more scope to put in place innovative teaching/learning strategies. They suggest that more use need to be made of innovative assessment methods (OECD Looney, 2009: 1). ‘Unseen’ assessment mechanisms might be used at key Stages 2 and 3 and synoptic assessment of skills might be undertaken at GCSE/A Level (similar to Queensland) which focus on thinking skills that are central to the NI Revised Curriculum (including, information management, problem-solving, decision-making, and creativity). This would mean that assessment and examining would serve the curriculum (and the skills needs of the economy) and drive pedagogy in the right direction. If teachers were teaching to these types of 21stC tests they would at the same time be teaching towards the skills identified by the EC, the OECD and the CBI as vital to future learning. Note that an assessment of cross-curricular problem solving was in PISA 2003 and a computer based version is in PISA 2012¹. There is also a big international project on Assessing and Teaching 21st century skills, with a focus on cooperative problem solving².

9.10 Build assessment literacy: CCEA moderation processes should support the development of better assessment of literacy through supportive internal moderation and cross-sectoral agreement trails for professional development.

¹ See http://www.oecd-ilibrary.org/education/pisa-2012-assessment-and-analytical-framework/problem-solving-framework_9789264190511-6-en

² See <http://atc21s.org/>

10 Recommendations for improved governance & transparency

10.1 **Review the influence of ‘governance by targets’:** National inspection systems in different countries can sit at various points on a spectrum, for example, being within the Government department responsible for education as is the case in Northern Ireland, Ireland and Flanders, or be totally independent of Government, as for instance in Sweden. In all cases, whether fully or partially associated with government or independent, there is a perception that inspection systems are potentially an instrument for implementing policy or achieving Government targets. Governments’ desire to foster greater accountability within public services, as well as to allow wider user choice, has been central to the growth of performance indicators for schools. The key driver of inspection approaches is therefore government targets and expectations.

10.2 **Review Programme for Government Educational Targets and NI Audit Office Educational Performance Monitoring**

However, a number of studies have been critical of governments’ lack of awareness and responsiveness to the challenges posed by league tables. Kane and Staiger (2002) highlight the tendency to ‘draw unwarranted conclusions on the effectiveness or ineffectiveness of policies based upon such short-term fluctuations in performance’ (p. 102). This is reinforced by the findings of Leckie and Goldstein (2009), which show that past performance is poorly correlated with future performance.

A further fundamental problem that surrounds discussions of public sector performance monitoring is the lack of systematic evaluations of its effects. Hallgarten (2001) points out that:

‘It should come as no surprise that targets and performance indicators change an organisation’s priorities. That is precisely their purpose. The concern occurs when such indicators skew priorities to the extent that other, normally less measurable, goals are relegated or jettisoned’. (ibid: 18)

This absence of sound evidence has made targets and performance measures a highly contentious area. Smith (1995) lists a number of problems which performance monitoring may generate which are all identifiable in political and civil service circles and replicated in our schooling system. These include:

- *Tunnel vision: a focus on quantifiable phenomena at the expense of all others.*
- *Sub-optimisation: the pursuit of narrow objectives at the expense of the aims of the organisation or system as a whole.*
- *Myopia and measure fixation: a focus on measures of success rather than underlying objectives.*
- *Misrepresentation: deliberate manipulation of the data collected.*
- *Misinterpretation: accidental misreading of the data, or unawareness of its limitations.*
- *Gaming: deliberate manipulation of behaviour to maximise league table position.*
- *Ossification: organisational paralysis due to an excessively rigid system of performance management.* (ibid: 20)

The Assembly and its Education Committee needs to reconsider its whole approach to educational monitoring based on a proper understanding of the impact of targets and Goodhart's law, whether or not they promote or inhibit improvement. Similar consideration needs to be given to whether or not the Audit Office should be making judgements about educational performance based on limited and flawed statistical evidence.

Hood (2007) introduces *'the idea of 'intelligence systems', which gather background information on the quality of performance with the intention to improve knowledge about the factors affecting the performance of a system, without focusing on particular measures or incentives to affect [and distort] the behaviour of the actors in that system'* (ibid: 16).

Since many of the factors affecting the performance of schools lie outside schools and, therefore largely outside schools' control, this would be insightful for politicians and policy makers. The British Academy inquiry into accountability and measurement advises that:

Consideration should be given to alternative ways of using quantitative information to monitor educational performance generally. This can be achieved by in-depth study of a sample of schools and students within a national database. A useful model is the Assessment of Performance Unit that was set up in the 1970s in England and discontinued in the 1980s (Gipps and Goldstein, 1983). Consideration should be given to using performance information as a screening device... accompanied by an emphasis on evaluation and inspection systems that are designed to emphasise ways of assisting schools to cope with problems rather than 'exposing' them using public rankings [reporting] (ibid: 12).

Foley & Goldstein, 2012

10.3 Review the audience 'transparency' and process of reporting: One of the basic principles of evaluation is to meet the demand for transparency. Two issues are important here – firstly transparency of the evidence used to arrive at inspection judgements and secondly the audience and purpose for which the report is written and how that affects the nature of the report. Any serious criticisms of a school should have to meet a higher evidential standard - beyond reasonable doubt – as opposed to a balance of probabilities in order to make acceptance of criticism more palatable.

Secondly, the publication of inspection reports, usually seen as highly desirable for reasons of transparency and accountability, may increase the pressure on schools to act defensively when criticised. Those countries where reports are kept in confidence between the inspectorate and the school (such as Hesse, Saxony and Rhine Palatinate in Germany) may avoid the issue (Penzer & Allen, 2011). Scotland, for example, provides two reports – a short one for publication and a more detailed one for detailed discussion with schools. Schools should also be at liberty to question those inspection judgements they disagree with. If ETI's mission is principally to 'promote improvement' then this should inform the style of reporting the clarity of its argument, the persuasiveness of the evidence it marshals and the timeliness of its publication. Timing can be an important factor – how soon after an inspection the report is finalised so as to build on any momentum established by the inspection itself.

10.4 Review the contribution of inspection systems to school improvement and the role and status of ETI:

Good evidence as to the benefits of inspection judgements in contributing to school improvement is in short supply (Foley & Goldstein 2012) given that: *'there is relatively little proof of the relationship between inspection and school improvement'* (Whitby, K., 2010 in Perry, C., 2012, P21). A study of inspection systems across 17 countries (Prenzer and Allen, 2011) found little evidence of deliberately designed systems to turn inspection into improvement. The British Academy recommends that: further consideration should be given to the role of inspection and accreditation agencies ... *especially when they are perceived to be instruments of government* (ibid: 12). Any such review should take account of international research and be subject to extensive debate and consultation with stakeholders. The OECD (2013b) recommends giving a prominent role to independent evaluation agencies but also integrating evaluation and assessment frameworks and aligning these with educational goals and student learning objectives so as to secure link to the classroom and draw on teacher professionalism. One consideration might be to separate ETI from DE and link it to the CASS service, outside of ESA, as an independent evaluation and support service as in Scotland.

10.5 Implement an ethical code to govern the publication of school performance reporting:

Wherever Institutional judgements or rankings are produced they should be accompanied by clear evidence and accompanied by prominent 'health warnings'. . An ethical code should be formulated (Goldstein and Myers, 1996) based on the two broad principles: *that unjustified harm to those to whom the information applies should be prevented, and that there should be no absolute publication rights for performance data* (ibid). One of the basic principles of evaluation is to meet the demand for transparency.

10.6 Ensure accurate and transparent media reporting of educational outcomes:

Despite school league tables being abolished by Ministers in Northern Ireland the media have taken initiatives to compile league tables. This has become a global activity, part of The Global Education Reform Movement (GERM) which is responsible for standardised testing, teacher accountability, school inspections and centrally imposed curricula. Critics believe that GERM is like a virus which has lowered standards, not raised them (Sahlberg 2012).

The media often fails to highlight that (1) often tables are based on results of a small group of pupils, which in itself makes the findings unreliable and (2) the missing critical factor is the background children bring to any particular school with them, negative and positive. Tables apparently showing a school high up the charts may just tell us a school takes in well-motivated and able pupils. Even the 'value-added' tables that are now produced which do take into account some of the pupils' backgrounds may not give us a reliable picture of school life, because they average over all pupils and may hide some pupils consistently doing well, others doing worse. A British Academy Inquiry advises that

The government should consider ways to prevent league tables being exploited by the media, such as ensuring that measures of uncertainty are provided around any institutional results. Associated with this there could be a campaign to better inform the public at large about the strengths and limitations of league tables, although any such attempt poses considerable challenges. Wherever league tables are published they should be accompanied with appropriate and prominent 'health warnings' highlighting their technical limitations. These should include assessments of the statistical uncertainty, often large, that may limit their usefulness. They should

also include statements about the quality of the measurements that go to make up the indicators, including the effects of aggregation. In a broader context, there is a need for a debate about whether simply making data available to citizens will encourage good use of them. In the absence of professional support and advice, data analysis can be very difficult for those with limited experience or expertise. Deliberate or unintentional misuse of statistical information should not be encouraged and there is a real danger that this could occur increasingly unless public awareness of the issues improves (Foley Goldstein, 2012)

Some countries make it an offence for newspapers to publish school outcome information. The Education Committee and DE should consider ways to prevent league tables being published or exploited by the media, by requiring that measures of uncertainty are provided in relation to all measures and institutional judgements, and challenging distortion of educational data. This may help to reduce deficit reporting and enhance understanding and respect for the important contribution which the teaching profession and schools make to the well-being and success of young people, society and the economy.

11: Conclusion

On 14th April 1970 the commander of the Apollo 13 space mission James Lovell used the phrase '*Houston we have a problem*' to calmly convey a message to mission control in Houston Texas that the space shuttle had suffered a major failure in technical design which led to a near fatal explosion that incapacitated the mission. The phrase has become synonymous with reporting any kind of critical design fault or problem. The shuttle designers immediately set about reviewing all of the steps in the design process to solve the critical problem they faced. Mission control's approach was that '*failure was not an option*'.

At the moment we seem to be facing a critical design problem in relation to the coherence of education policies. We can be assured about one thing – we are not alone in this regard. Indeed, the fact that we are asking so many questions at the moment about our education system is to be applauded. We have just had a major inquiry by the OECD into assessment which is due to report in the autumn. We are in the midst of a review of assessment and examinations. We are in the midst of a review of the school estate and are currently consulting on school funding. We have an on-going review of teacher education for some considerable time, almost as long as the review of administrative and support structures. Now we have this major inquiry into ETI and school improvement. These are all important system design issues and they are all interconnected, like the control panels on Apollo 13. A weak link in one area can destabilise the whole enterprise. That's why we need the system policy designers at mission control to stand back and join up the insights into one coherent policy that enables our schools and teachers to get on with the job that they want to do, that of improving teaching and learning.. If schools are expected to accept the challenge from inspection reports to continuously improve their policies, approaches and outcomes, then as the saying goes – what is sauce for the goose is sauce for the gander.

The evidence presented here aims to prompt discussion about the health of our education system right now, the stress being placed upon pupils, teachers and schools and the image of our education system that is being presented to the public and our politicians, and to get everyone in the system to objectively consider where we are right now; where we want to go; what we want to achieve in the future and what we need to do to get there.

Where are we now? The analysis of current and proposed accountability policies would suggest that we are now headed in the direction of hyper-accountability based on dubious measures that present a distorted view of achievement, which flatters schools with selective intakes and is patently unfair to non-selective schools in the most challenging circumstances.

The evidence presented demonstrates that schools are a reflection of the selective communities they serve, the aspirations and cultural capital of the families from which pupils are drawn, and the ethos and impact of the education policies which drive them. It illustrates the complexity of the challenge, the inadequacy and unreliability of the accountability mechanisms currently used, and the fragility of the assumptions on which they are based. It contends that there are no quick fixes, no simple solutions and no fast-track routes to sustainable success. What is instead required is a much more sophisticated approach to joined-up social and economic, health and education policies to uplift family and community circumstances and aspirations from the cradle to the grave. In the case of education, the influence also needs to be pre-natal as well as in the early years. The bottom line which politicians, civil servants and parents must understand is that schools and teachers are far from the sole cause, and certainly not the sole solution, to the challenges which face our

economy, our society and neighbourhoods. By all means hold teachers and schools to account, but recognise the communities they reflect and the things they can and cannot control, not least the impact of selection which separates many young people at a very vulnerable age from positive peer influences.

Where do we want to go? The evidence from other systems endorses a constructive and supportive model of accountability, which builds teachers' confidence and commitment (as opposed to a deficit model which engenders fear and which may encourage perverse practices and unintended outcomes to achieve compliance and avoid retribution). The clearest analogy is that of parenting a child. If you encourage and support, you create confidence and self-esteem. If you constantly criticise and sanction you create resentment and disempowerment. We need to applaud our strengths, as well as challenge our weaknesses.

To use a Scottish analogy, we can take the "High Road" or the "Low Road". The "Low Road" is characterised by systems of micro-accountability, league tables, excessive testing, bureaucratic assessment and data driven evaluation, in which teaching is treated as a low skill, low discretion craft. The "High Road" is characterised by a reflective, high skill, autonomous profession, where teachers are recognised and appreciated for their knowledge, expertise and judgement. We have sufficient evidence across the UK and worldwide to show which approach bears fruit. We need to develop a new accountability system with broader value-added measurements which can motivate and encourage schools in challenging environments and better identify need and enable resources to be channeled toward those needs.

What do we want to achieve? By virtue of our size and the talent of our teachers Northern Ireland has the potential to be, not just a good, but a great education system. We want to do that on the basis of an informed understanding of what works internationally. To progress from 'good to great' or indeed from 'great to excellent' (McKinsey, 2010) requires that policy makers support and nurture a high trust, high autonomy, high discretion profession and a broader vision of education that will develop young people with 21st century skills. We are a small place in a small geographic space, where there are no natural resources at our disposal except the ingenuity and creativity of our people. The quality, motivation and creativity of the young people that our education system produces are central to our economic survival in an increasingly competitive world. Our education system in Northern Ireland is internationally recognized as being ahead of the game, having put in place specific definitions of 21st century skills and competencies at regional level (Gallagher, Hipkins, McGuinness & Zohar, 2011). So the 'leap' now required is that these 'new literacies' find their way into the accountability system, alongside better use of socio-economic data and appropriate baselining to assess value-added, in a supportive accountability framework.

How do we get there? Reflecting on a long career in the Civil Service, Sir Gus O'Donnell recently reviewed some of the policy triumphs and failures of his period of service and summarised his reflections for how the public sector is run and how it needs to evolve in 10 commandments of good policy making. Four of these are pertinent to our current scenario.

- Thou shalt be clear about the outcomes that you want to achieve: Lack of strategic clarity, of knowing the problem you are trying to solve, is a cardinal sin.
- Thou shalt evaluate policy as objectively as possible: Be clear about how you to determine success and relate success measures to desired outcomes.
- Honour the evidence and use it to make decisions

- Thou shalt not kill the messenger. If you don't encourage internal debate you will learn about your mistakes from your enemies not your friends.

The range of international evidence cited illustrates the extent to which other countries are engaging with issues similar to those identified by this education committee inquiry. In offering the messages within this report, the intention is to encourage collaborative internal debate within the system towards developing joined-up supportive policies, structures and resources that enable us to put our energies into encouraging innovative teaching, learning and assessment to support 21st century skills. Supporting this we need to develop proper base-lining and value-added measures, accompanied by supportive accountability.

The bottom line is that all of us who are engaged in advising on, developing and implementing policy to support schools need to articulate a common moral purpose to inform our roles and remits and collaborative actions in support of schools and each other. GTCNI published a charter for education some years ago. A refreshed charter should perhaps emanate from the Education Committee and the Department of Education in consultation with all partners and be signed up to by all.

The evidence in this submission aims to offer constructive insights to enable our system to **strike the right balance** 'between holding schools to account and allowing innovation and supporting school improvement' (Perry, C., 2012, P1, NIARIS). The key to achieving the right balance is the development of a coherent and supportive framework of accountability that unleashes the creativity and energy of teachers, pupils and schools towards 21st century learning.

Appendix 1: GTCNI Survey of Teacher Perceptions of the usefulness and manageability of End of Key Stage Assessment Arrangements June 2013

Your Views on Statutory Assessment

Purpose

The purpose of this independent GTCNI survey was to gather teachers' views about the manageability, reliability and usefulness of the end of key stage statutory assessment arrangements introduced in 2012-13.

Responses

- **One response** was invited from each school.
- **500 responses** were received, representing potentially 50% of schools involved in statutory assessment at Key Stages 1, 2 and 3.
- **227 Primary School responses** were received from: 34 KS1 teachers, 51 KS2 teachers, 21 assessment co-ordinators, 45 teaching principals and 76 non-teaching principals.
- **250 Post-primary School responses** were received from 186 KS 3 teachers; 46 assessment co-ordinators; and 18 post-primary principals.
- **33 other teacher responses** did not identify a category.

Main Findings

(1) Usefulness of numerical outcomes (levels):

The great majority of respondents (*between 82% and 89%*) regarded numerical outcomes (*Levels*) **to be of limited or no use to ...**

	Useful %	Not Useful %	Total of Responses
• Individual pupils in informing their on-going learning	17%	83%	469
• Teachers in informing their teaching	23%	77%	469
• Board of Governors in judging value added to pupil learning	16%	84%	468
• Parents in understanding their child's progress	11%	89%	468
• Receiving schools informing them what a pupil knows	18%	82%	463
• ELBs to inform the support your school may need	18%	82%	466
• ETI in understanding school value-added to pupils	19%	81%	467
• Politicians in understanding system performance	14%	86%	467

(2) Usefulness of level statements of attainment:

The great majority of respondents (*between 82% and 89%*) regarded the statement of attainment **to be of limited or no use in informing ...**

	Useful %	Not Useful %	Total of Responses
• Teacher judgements about pupils' work	34%	66%	461
• Teacher feedback to pupils on next steps of learning	25%	75%	461
• Adjustments to teaching	27%	73%	461
• Annual Report to parents	26%	74%	456

(3) Internal moderation:

Almost 60% of respondents (*268 teachers*) found involvement in internal school moderation useful or very useful in informing their understanding of progression and standards of pupils' work.

4) External moderation:

While many respondents said they did not participate in external moderation, of those who did:

- Only 1 of the 459 respondents considered the process easy to manage.
- 91% considered the process burdensome or very burdensome.
- 42% of those who prepared portfolios said they prepared over 30 portfolios each.
- 51% of these included 4-6 pieces of individual pupil's work in each portfolio.
- 86% of 351 respondents considered the timeframes associated with end of key stage assessment to be inappropriate.
- 70% (*310 of 437 teachers who responded to this question*) considered external moderation of limited usefulness or not at all useful in informing their understanding of progression and standards of pupils' work.

- Almost half of all respondents (246 teachers) answered an open question about the usefulness of CCEA feedback from external moderation. The majority of comments were negative. The researcher who compiled this report independently selected the following as an illustration of the views expressed:

The feedback from moderation was hugely disappointing. It undermined the professionalism of the staff concerned... The result of moderation changed the levels of pupils ... that they were now seen to be achieving a level comparable with other pupils who are known to be clearly working at lower levels. (Key Stage 3 teacher)

It was useful to see the feedback as it confirmed awarded levels and procedures. As it was the first year of this assessment this was useful in an unfamiliar area. (Key Stage 3 teacher)

The feedback was one line to say that the levels were not verified and no other attempt to give us feedback was made... found it highly unhelpful and disheartening. (Key Stage 2 teacher)

Feedback was extremely poor. It was vague and didn't give you any specific advice on how to 'fix' pupil portfolios. Overall, to have spent so much time preparing these portfolios for moderation, the feedback that was given was disgraceful. (Key Stage 2 teacher)

Our feedback was not helpful and I had to follow it up with an email. I then received a phone call which clarified things somewhat but we do not feel that we will be any more aware of how to show enough evidence next year. (Teaching Principal, Primary)

(5) Reliability of Level outcomes:

The majority of respondents considered the numerical end of key stage outcomes (*levels*) to be of limited reliability or not at all reliable for:

	Very Reliable	Reliable	Limited	Not at all
Their Class (430)	7%	27%	48%	18%
Their School (433)	7%	28%	47%	18%
Other Schools (433)	>1%	11%	56%	32%
NI system as a whole (437)	>1%	11%	49%	39%

This distrust of the reliability of level judgements indicates that an external moderation system, no matter how rigorous, is unlikely to be able to guarantee the reliability of end of key stage assessments, simply because teachers do not value other schools' judgements.

(6) Use of Levels for accountability purposes:

71% of respondents (313 teachers) felt that using teacher assessment for accountability purposes could distort the process. Only 8% (36 teachers) believed that this would not be the case. The comments below illustrate this:

There is no doubting the distortion within the reporting of levels. Any of the main stakeholders, ETI/ELBs/Principals will all 'off the record' admit the results are completely at odds

with reality. Why do we have to persist with this farce? Are we not a mature enough profession to stand up and call this situation as it is? (Principal Primary).

In my experience as a teacher of 24 years, I have seen how teachers can distort assessment results to match parental expectations, class results and NI outcomes. (Teaching Principal Primary).

Teachers have been assessing the end of key stage 3 levels since 1993. The teachers' involvement does not invalidate the process. It is the pointlessness of the process that negates it. This is an ill-conceived process. (Key Stage 3 teacher)

When asked if teacher assessment should be used for assessment for learning purposes only and separated from accountability, 67% said yes, 11% said 'no' and 22% were not sure. The following is an independently selected sample of the views offered by 41 respondents:

Unreliable data is being used for accountability purposes at the minute. This was never the purpose of levels or teacher assessment. (Key Stage 3 teacher)

Teachers are being put under pressure to meet targets at KS3 and then show further improvement at KS4. Assessment should be about learning and not hitting targets. (Key Stage 3 teacher)

I believe the whole system of levels is of limited use...a level means very little to a pupil but a constructive comment with clear indicators of what needs to be done to bring about improvement can help. (Assessment Co-ordinator Post-primary)

Please just let us teach! Education is being ruined by all this accountability and limiting of children's experiences. (Key Stage 1 teacher)

Insights from research

The clear recommendation from assessment experts (*The Assessment Reform Group: Gipps; Tymms etc.*)⁸ is that the processes of teaching, learning and assessment should focus on improving learning and should not be over-burdened with bureaucracy or exposed to potential manipulation for accountability purposes. Virtually all of the research into the use of teacher assessment (*and levels of attainment in particular*) advises against the use of numerical assessment outcomes for target setting and accountability purposes. The following quotations give a flavour of researchers' views:

- *Reducing attainment to a single figure or grade while attractive to politicians and the public 'as a form of shorthand' in which to report performance masks complex nuances in ability and performance (Gipps, 1994)*.*
- *Trying to achieve multiple objectives with a single policy instrument is not feasible (Hanushek & Raymond, 2004)*.*
- *No single measure can fulfil both the formative and summative functions (Harris, 2010)*.*
- *Assessments should be treated as approximations, subject to unavoidable errors (Gardner, 2008)*.*
- *Use of assessment evidence for accountability is based on*

References *(and broader reading)*

- Alexander & Hargreaves, (2007) Community Soundings – the Primary Review
http://www.primaryreview.org.uk/downloads/Int_Reps/1.Com_Sdg/Primary_Review_Community_Soundings_report.pdf
- Assessment Reform Group (2002) Testing, Motivation and Learning
<http://www.aiaa.org.uk/content/uploads/2010/06/Testing-Motivation-and-Learning.pdf>
- Assessment Reform Group & TLRP (2009) Assessment in Schools: Fir for Purpose?
<http://www.tlrp.org/pub/documents/assessment.pdf>
- Bew, P. (2011) Independent Review of Key Stage 2 testing, assessment and accountability – Final Report. London: DfE
- Borooh, V. & Knox, C: Shuffling desks or improving education performance? Area planning in Northern Ireland, University of Ulster June 2013
- Brown M (2013) Deconstructing Evaluation in Education (The Case of Ireland) A thesis presented to Dublin City University for the Degree of Doctor of Philosophy
- Cassen, R. & Kingdon, G. (2007) Tackling Low Achievement, York, Joseph Rowntree Foundation
- CBI, (2012) First Steps Report
<http://www.cbi.org.uk/campaigns/education-campaign-ambition-for-all/first-steps-read-the-report-online/>
- Chevalier A., Dolton P. & Levacic R. (2005) School and Teacher Effectiveness, in Machin S. & Vignoles A. What is the Good of Education?, Princeton, Princeton University Press
- Clarke, C. and Ozga, J. (2011) Governing by Inspection? Comparing school inspection in Scotland and England, Paper for Social Policy Association conference, University of Lincoln, 4-6 July 2011.
http://www.social-policy.org.uk/lincoln2011/Clarke_Ozga.pdf
- Clotfelter, C.T; Ladd, HF., Vigdor, J.L., & Diaz, RA., (2004) “Do School Accountability Systems Make It More Difficult for Low-Performing Schools to Attract and Retain High-Quality Teachers?” *Journal of Policy Analysis and Management*. Vol. 23, No. 2 251–271
- Coleman J.S. et al (1966) Equality of Educational Opportunity, Washington D.C., National Centre for Educational Statistics
- Darr and McDowall, (2008) Standardised Testing: Dilemmas and Possibilities
- DENI (2008) Every School a Good School: A strategy for raising achievement in literacy and numeracy. Bangor, Department of Education
http://www.deni.gov.uk/literacy_and_numeracy_strategy_-_english.pdf

DENI (2012) Vision for Education 2012-15 www.deni.gov.uk.

DfE (2011) The Framework for the National Curriculum: A report by the Expert Panel for the National Curriculum review London
<https://www.education.gov.uk/publications/standard/publicationDetail/Page1/DFE-00135-2011>

Doyle, S. (2013) Schools with more 'poor' pupils get worse results, Irish News of 26 February 2013.

Education Scotland (2013) Senior Phase Benchmarking Tool
<http://www.educationscotland.gov.uk/thecurriculum/howisprogressassessed/qualifications/benchmarking/index.asp>

Ehren, M. Tymms, P. Jones, K. Gustafsson, J. Myrberg, E. McNamara, G. O'Hara, J. Conyngham, G. Altricher, H. Kemethofer, D. and Schmidinger, E. Technical report EU-project 'Impact of School Inspections on Teaching and Learning'

<http://schoolinspections.eu/wp-content/uploads/downloads/2012/06/Technical-report-ISI-TL-2011-1.pdf>

- Dr. D. Greger (Charles University in Prague, Faculty of Education, Czech Republic) Foley, B. & Goldstein, H., (2012) Measuring Success: League tables in the public sector. A British Academy Policy Centre report.

Fullan, M. (2011). *Choosing the Wrong Drivers for Whole System Reform*. Centre for Strategic Education, Seminar Series 204.

Fullan, M. (2012). *Stratosphere: Integrating technology, pedagogy, and change knowledge*. Toronto: Pearson.

Fullan, M. and Langworthy, M. (2013) *Towards a New End: New Pedagogies for Deep Learning*, The global partnership www.newpedagogies.org/Pages/assets/new-pedagogies-for-deep-learning---an-invitation-to-partner-2013-19-06.pdf

Fullan, M. & Quinn, (2010) Capacity Building for Whole System Reform

Gallagher C, (2012) *Curriculum Past, Present and Future: A study of the development of the Northern Ireland Curriculum and its assessment in a UK and international context*, unpublished thesis, University of Ulster.

Gallagher, C. Hipkins R. and Zohar A. (2012) Positioning thinking within national curriculum and assessment systems: Perspectives from Israel, New Zealand and Northern Ireland, *Journal of Thinking Skills and Creativity Elsevier* <http://www.elsevier.com/locate/tsc>

Gardner J, Harlen W, Hayward L and Stobart G (2008): Changing Assessment Practice: Process, Principles and Standards, Assessment Reform Group Pamphlet

Gipps C (1994) *Beyond Testing*, Routledge Falmer, London

Gipps, C. and Goldstein, H. (1983), *Monitoring Children* (London, Heinemann).

Goldstein, H. and Spiegelhalter, D. J. (1996), 'League tables and their limitations: statistical-issues in comparisons of institutional performance' in *Journal of the Royal Statistical Society*, A. 159: 385–443

Goldstein, H. and Leckie, G. (2008) 'School League Tables: What can they really tell us?' in *Significance*, 5(2).

Goldstein, H. and Myers, K. (1996), 'Freedom of Information: Towards a code of ethics for performance indicators' in *Research Intelligence*, 57.

Goldstein, H., Burgess, S. and McConnell, B. (2007), 'Modelling the impact of pupil mobility on school differences in educational achievement' in *J. Royal Statistical Society*, A. 170: 941– 954.

Hallgarten, J. (2001) 'School League Tables: Have they outlived their usefulness?' in *New Economy* 8(4).

Hanushek, E.A. (1992) The trade-off between child quantity and quality, *Journal of Political Economy*, 100, 84-117

Hanushek & Raymond, 2004, Does School Accountability Lead to Improved Student Performance? NBER Working Paper No. 10591 Issued in June 2004

<http://www.nber.org/papers/w10591>

Hargreaves, A., & Fullan, M. (2012). *Professional Capital: Transforming Teaching in Every School*. New York City: Teachers College Press.

Hattie, J. (2011). *Visible Learning for Teachers: Maximizing Impact on Learning*. New York: Routledge.

Harland et al, (2002) Is the Curriculum Working? The Key Stage 3 Phase of the NI Curriculum Cohort Study, Slough: NFER

Harris, D.N. Value-Added Measures of Education Performance: Clearing Away the Smoke and Mirrors. *Policy Analysis for California Education*. 2010.

Hood, C. (2007), 'Public Service Management by Numbers: Why does it vary? Where has it come from? What are the gaps and puzzles?' in *Public Money and Management*, Vol. 27, No.2.

Husband, C. (2012) *Accountability: just what do we want to measure?* Institute of Education, London

<http://ioelondonblog.wordpress.com/2012/12/27/accountability-just-what-do-we-want-to-measure/>

Hayward, L. et al, (2012) Assessment at Transition, Draft Advice to Scottish Government, University of Glasgow.

Hooge, E. Burns, T. Wilkoszewski, H. (2012) Looking Beyond the Numbers: Stakeholders and Multiple School Accountability, OECD

http://www.oecd-ilibrary.org/education/looking-beyond-the-numbers-stakeholders-and-multiple-school-accountability_5k91dl7ct6q6-en

Johnson, M. (2007) Reforming Urban Education Systems, in Pink W.T. and Noblit G.W. (Eds) *International Handbook of Urban Education*, Dordrecht: Springer

Kounali, D. Robinson, T. Goldstein, H. and Lauder, H., (2012) The probity of free school meals as a proxy measure for disadvantage, Centre for Multilevel Modelling, Graduate School of Education, University of Bristol.

<http://www.bristol.ac.uk/cmm/publications/fsm.pdf>

Ladd, H. F. and Walsh, R. P. (2002), 'Implementing Value-Added Measures of School Effectiveness: Getting the incentives right' in *Economics of Education Review*, 21(1).

Leckie, G. and Goldstein, H. (2011), 'Understanding uncertainty in school league tables', in *Fiscal Studies* (Forthcoming).

Leckie, G. and Goldstein, H. (2009), 'The limitations of using school league tables to inform school choice' in *Journal of the Royal Statistical Society*, A 172: 835–851.

Laurillard, D. (2012). *Teaching as a Design Science: Building Pedagogical Patterns for Learning and Technology*. New York City: Routledge. 30

Light, D., Price, J., & Pierson, E. (2011). *Using classroom assessment to promote 21st century learning in emerging market countries*. Center for Children & Technology.

Levin, B. (2010) How to change 5000 Schools: *A Practical and Positive Approach for Leading Change on Every Level*, University of Auckland Centre for Educational Leadership

Mc Beath, J. 2012: *The Future of the Teaching Profession* University of Cambridge Faculty of Education and Education International Research Institute

McGuinness, C., Curry, C., Greer, B., Daly, P., & Salters, M. (1996). Final report on the ACTS project: Phase 1 (p. 80). Belfast: Northern Ireland Council for Curriculum Examination and Assessment.

McGuinness, C., Curry, C., Greer, B., Daly, P., & Salters, M. (1997). Final report on the ACTS project: Phase 2 (p. 28). Belfast: Northern Ireland Council for Curriculum Examination and Assessment.

McGuinness, C.(1999). From thinking skills to thinking classrooms: A review and evaluation of approaches for developing pupils' thinking. DfEE research report no. 115. Norwich: HMSO.

McGuinness, C. (2005a).Meta-cognition in primary classrooms: Apro-ACTive learning effect for children .In Paper presented at the ESRC TLRP annual conference Warwick, 28–30 November.

McGuinness, C. (2005b).Teaching thinking: Theory and practice. British Journal of Educational Psychology Monograph Series II Number3–Pedagogy–Teaching for Learning, 1(1), 107–112.Ministry of Education Israel.

McGuinness, C. et al. (2005) 'Metacognition in primary classrooms: a pro-ACTive learning effect for children'. Paper presented at the TLRP Annual Conference, University of Warwick, November.

McKinsey, (2007): How the world's best performing school systems came out on top <http://www.shift-learning.co.uk/useful-links/153-mckinsey-report-2007-how-the-worlds-best-performing-school-systems-came-out-on-top.html>

McKinsey, (2010) Mourshed, M. Chijioke, C. and Barber, M: How the world's most improved school systems keep getting better

Mansell, W. (2007) Education by Numbers: The Tyranny of Testing, Politico Publishing Ltd

Mehta, J., Schwartz, R. B., & Hess, F. M. (2012). *The Futures of School Reform*. Cambridge: Harvard Education Press.

(MET) Bill & Melinda Gates Foundation. (2012). *Measures of Effective Teaching*. Retrieved from <http://www.metproject.org/>

Miller, R., Looney, J., & Siemens, G. (2011). *Assessment Competency: Knowing What You Know and Learning Analytics*. Promethean.

Mortimore P (1999): *Understanding pedagogy and its impact on learning*, Chapman Publications

Mortimore P. (1998) The Road to Improvement: Reflections on School Effectiveness, Lisse, Swets & Zeitlinger.

National Economic and Social Council - NESC (2012) Quality and Standards in Human Services in Ireland: The School System No. 129 August 2012, National Economic and Social Development Office (NESDO)

OECD – Scotland Report, 2007

<http://www.oecd.org/unitedkingdom/reviewsnationalpoliciesforeducation-qualityandequityofschoolinginscotland.htm>

OECD (2011) Education at a Glance 2011 Indicators, Paris: Organisation for Economic Cooperation and Development, (Accessed 2 February 2012)

<http://www.oecd.org/dataoecd/61/2/48631582.pdf>

OECD Reviews of Evaluation and Assessment in Education NEW ZEALAND (2012):

(Deborah Nusche, Dany Laveault, John MacBeath and Paulo Santiago)

<http://www.oecd.org/edu/school/49681441.pdf>

OECD (2013a), "School evaluation: From compliancy to quality", in Synergies for Better Learning: An International Perspective on Evaluation and Assessment, OECD Publishing. <http://dx.doi.org/10.1787/9789264190658-10-en>

OECD (2013b) 'Synergies for Better Learning An International Perspective on Evaluation and Assessment Pointers for Policy Development'

www.oecd.org/edu/evaluationpolicy

Ozga, Jennifer et al (2009). ECRP05: Governing by Numbers: data & education governance in Scotland & England: Full Research Report ESRC End of Award Report, RES-000-23-1385. Swindon: ESR REFERENCE No. RES 00 231385

Penzer, G and Allen, V. (2011) School Inspection: what happens next? CfBT

Perry, C. (2012) *School inspection - An overview of approaches to inspection in Northern Ireland, England, Scotland, Wales*, Ireland and internationally, Research and Information Service, Northern Ireland Assembly.

Purvis D (2011) Educational Disadvantage and the Protestant Working Class

A Call to Action

http://www.nicva.org/sites/default/files/A-Call-to-Action-FINAL-March2011_0.pdf

Ravitch, D. 2010. First, let's fire all the teachers! The Huffington Post [Online], 03 February.

Available from: http://www.huffingtonpost.com/diane-ravitch/first-lets-fire-all-the-t_b_483074.html

Reeves, D. (2005). *Accountability in action: A blueprint for learning organizations*. Englewood, CO: Advanced learning Press.

Sahlberg, P. (2010) Finnish Lessons: What can the world learn from educational change in Finland? The series on School Reform, University of Washington

Sahlberg, P. (2012) Quality and Equity in Finnish Schools, School Administrator Magazine P 27-30

http://pasisahlberg.com/wp-content/uploads/2013/01/Qualit_and_Equity_SA_2012.pdf

Sahlberg, P. 2012. Finland: A non-competitive education for competitive economy. In OECD: Strong performers and successful reformers – Lessons from PISA for Japan. Paris: OECD, pp. 93-111.

Salisbury, R. 'An Independent Review of the Common Funding Scheme

Sammons P (2007) School effectiveness and equity: making connections, CfBT.

SICI European Inspectorates' Profiles: Spain, 2009

Stothard, Snowling, Bishop, Chipchase & Kaplan, (1998) Language-impaired preschoolers: a follow-up into adolescence) <http://www.ncbi.nlm.nih.gov/pubmed/9570592>

Sturman, L., Twist, L., Burge, B., Sizmur, J., Bartlett, S., Cook, R., Lynn, L. and Weaving, H. (2012) PIRLS and TIMSS 2011 in Northern Ireland: reading, mathematics and science. Slough: NFER. www.nfer.ac.uk/

Teddlie C. & Reynolds D. (2000) *The International Handbook of School Effectiveness Research*, London, Falmer Press

TES magazine, 'GCSE deflation blamed on tactical exam entry' 23 August, 2013
<http://www.tes.co.uk/article.aspx?storyCode=6351210>

Trilling, B., & Fidel, C. (2012). *21st century skills: Learning for life in our times*. New York: Jossey-Bass.

Tymms P & Fitz-Gibbon C (2001) Standards, achievement and educational performance in Philips, R and Furlong, J *Education, Reform and the State: Politics, policy and practice, 1976–2001*, London: Routledge

Tymms P (2004) Are standards rising in English primary schools? *British Educational Research Journal* 30:4, August 2004, <http://eppe.ioe.ac.uk/eppe3-11/eppe3-11%20pdfs/The%20Impact%20of%20Research%20on%20Policy.pdf>

Tymms P Jones P Albone S and Henderson B (2007) *The First Six Years at School*, Paper to British Educational Research Conference

Tymms P & Merrell C (2007) Standards and Quality in English Primary Schools Over Time: the national evidence (Primary Review Research Survey 4/1), Cambridge: University of Cambridge Faculty of Education

Van Horn, C. (2011, May 25). Their disappointment is justified. *New York Times*. Retrieved from <http://www.nytimes.com/roomfordebate/2011/05/24/the-downsized-college-graduate/graduates-disappointment-is-justified>

Vander Ark, T. (2011). *Getting smart: How digital learning is changing the world*. New York: Jossey-Bass.

Visscher, A. J. (2001), 'Public School Performance Indicators: Problems and recommendations' in *Studies in Educational Evaluation*, 27 (3).

Vieluf, S., Kaplan, D., Klieme, E., & Bayer, S. (2012). *Teaching practices and pedagogical innovations: Evidence from talis*. OECD Publishing. Retrieved from http://www.oecd-ilibrary.org/education/teaching-practices-and-pedagogical-innovations_9789264123540-en

Voogt, J., & Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. (2012. *Journal of Curriculum Studies*, 44 (3, 299-321).

WBEE/EBT Education Inspection and Performance Systems in Europe: A Comparative Handbook http://wbee-euproject.org/images/WBEE_handbook_combined_digital_version_29-05-12.pdf

Wiggins, A. and Tymms, P. (2002), 'Dysfunctional Effects of League Tables: A comparison between English and Scottish primary schools' in *Public Money and Management*, 22 (1).

William D (2001) Level Best? Levels of attainment in national curriculum assessment, ATL publication

Wilson, D. (2004), 'Which Ranking? The impact of a 'value-added' measure on secondary school performance' in *Public Money and Management*, 24(1).

Wilson, D., Croxson, B. and Atkinson, A. (2006), 'What Gets Measured Gets Done' in *Policy Studies*, 27(2).

General Teaching Council for Northern Ireland

3rd – 4th Floor Albany House

73 - 75 Great Victoria Street

Belfast BT2 7AF

Tel: (028) 9033 3390

Fax: (028) 9034 8787