The evolution of the ejournal
Disseminate your research immediately through a new online, open access, peer reviewed journal

Online Educational Research Journal

This paper was downloaded from

The Online Educational Research Journal

(OERJ)

www.oerj.org

OERJ is an entirely internet-based educational research journal. It is available to anyone who can access the web and all articles can be read and downloaded online. Anybody can submit articles as well as comment on and rate articles. Submissions are published immediately provided certain rules are followed.

All above average: expectations and the use of prior attainment data to estimate future attainment in pupil target-setting

Phil Taylor – Birmingham City University

Abstract

Continual demands for raised standards and expectations in English schools imply an impossible goal of above average attainment for all. To this end target-setting at pupil, subject and school levels has become well-established, often using predictive data such as that produced by Fischer Family Trust (FFT). This paper explores inherent issues in using prior attainment data to estimate future attainment and begins to analyse the effect on individual learners – an area that has not been extensively researched. With the relinquishing of statutory targets and proposed changes to Ofsted arrangements, the time is right to review target-setting in schools to ensure that available datasets are treated with critical awareness. Questions are raised about current practices observed in secondary schools, concluding that learning rather than performance should drive target-setting.

Background

The front page headline of the *Times Educational Supplement* on 10th June 2011 sets the scene:

> Thousands of schools could fall foul of new Ofsted regime. Those with below-average results will be failed – unless they can show evidence of 'steady' improvement. (Stewart 2011).

Schools are placed in an impossible situation, facing demands for ever-increasing performance amid calls for 'tougher' exams (Gove 2011). The coalition government since taking office in May 2010 has maintained its predecessor's focus on raising achievement, particularly among pupils from disadvantaged backgrounds, pointing to 'the soft bigotry of low expectations' (DfE 2010b p.4). Allied to this is the decision to abolish contextual-value added (CVA), a performance measure featuring prominently in Office for Standards in Education, Children's Services and Skills (Ofsted) inspections from 2005-9 (Ofsted 2005 p.6). The government has stated that CVA 'has the effect of expecting different levels of progress from different groups of pupils on the basis of their ethnic background, or family circumstances, which we think is wrong in principle' (DfE 2010b p.68).

National averages at age 16 have been steadily increasing (DfE 2011b) and the quest to raise expectations and narrow gaps for underperforming groups in schools is not new. Successive governments have demanded statutory targets that are 'challenging but also realistic' (DfE 2010a p.3). Target-setting processes are established and the current Ofsted evaluation schedule for schools stresses their importance in 'embedding ambition and driving improvement' (Ofsted 2011 p.38-40). However, in early 2011 the government announced 2012 as the last year for which statutory targets will be required, stating that 'schools should have the freedom to set their own priorities and be in control of their own improvement' (DfE 2011a). This change, along with the piloting of new Ofsted arrangements, provides an opportunity to review prevailing data-driven target-setting in schools.

Methodology

This paper considers the use of prior attainment to estimate future attainment for individual pupils, focusing on the age 11 to 16 secondary phase of schooling. A synthesis of extant research and policy literature is combined with analysis of data produced by the non-profit registered charity Fischer Family Trust (FFT). While statistical models to estimate future attainment for the purpose of assessing value-added are well documented and not considered in technical detail, there is less written about their use in informing target-setting. Anonymous FFT material utilised includes both real examples and exemplar cases from FFT's own training database. Fieldwork observations made in schools (*N*=10) concerning the use (and possible misuse) of future attainment estimate datasets are discussed alongside pupil examples. The construct of *assessment average* is explored, focusing on examples of pupils achieving the age-related expectation of Level 4 in primary school English.

Expected Attainment and Progress

Since its inception the National Curriculum in England has defined Level 4 as the **expected attainment** for most 11-year-olds by the end of Key Stage 2 (KS2) (DfEE/QCA 1999 p.17). **Expected progress** from KS2 through secondary school (Key Stages 3 and 4) is specified by a Department for Education grid, showing the anticipated General Certificate of Secondary Education (GCSE) grades based on KS2 levels (Fig. 1) (DfE 2010d and 2011c). A pupil reaching Level 4 in English (or Maths) at the end of KS2 is expected to achieve at least GCSE Grade C, usually five years later.

English 2010 **528,827** matched pupils		GCSE grade									
		no KS4 result	U	G	F	E	D	C	B	A	A*
KS2 test level	Other level or no prior available	Expected progress not made	Expected progress not made	Not included	Not included	Not included	Not included	Not included	Expected progress made	Expected progress made	Expected progress made
	B,N	Expected progress not made	Expected progress not made	Not included	Not included	Expected progress made	Expected progress made	Expected progress made	Expected progress made	Expected progress made	Expected progress made
	2	Expected progress not made	Expected progress not made	Expected progress not made	Expected progress not made	Expected progress made	Expected progress made	Expected progress made			Expected progress made
	3	Expected progress not made	Expected progress not made	Expected progress not made	Expected progress not made	Expected progress not made	Expected progress made	Expected progress made			Expected progress made
289,254 pupils	**Level 4 ➔**		**1%**	**1%**	**1%**	**5%**	**20%**	**43%**	**23%**	**6%**	**1%**
	5	Expected progress not made	Expected progress not made	Expected progress not made	Expected progress not made	Expected progress not made	Expected progress not made	Expected progress not made	Expected progress made	Expected progress made	Expected progress made

18,035 pupils

Fig. 1 Expected Progress, adapted from DfE (2010d) and Ofsted/DfE (2010).

The DfE/Ofsted (2010) *transition matrices* within *RAISEOnline*, enable the number and percentage of pupils in each position on the expected progress grid to be shown. In 2010, pupils achieving the expected Level 4 in KS2 English represented the majority of matched pupils (289,254 out of 528,827) (Fig. 1). The superimposed spread of English GCSE results achieved by these pupils shows C as the *most likely* or *median* grade, with 73% making the expected progress. But only 43% actually achieved Cs; 86% achieved the grades including and either side of the median, leaving a

further 14% who might be considered outliers. Six per cent or 18,035 pupils in this group exceeded the minimum expected grade C by two to achieve A, while slightly fewer did the opposite, gaining E. As well as representing past performance, this progress pattern provides a crude estimate of most likely future attainment – an *average* Level 4 pupil might be expected to achieve at least GCSE grade C. The transition matrices (DfE/Ofsted 2010) in Fig. 2, break levels 3, 4 and 5 into *sub-levels* (a, b and c) using raw marks from KS2 tests, to provide a sharper view of actual progress, from which expected progress can be derived. Pupils attaining Level 4a at the end of KS2 are more likely to gain higher GCSE grades than their peers who reach Level 4c or 4b.

Attainment at the End of KS4 in English

Sub-level Achieved in KS2 English	Number of Pupils	U	G	F	E	D	C	B	A	A*	TOTAL
	B	681	1,617	3,487	3,521	2,055	677	84	17	0	12,139
	N	340	879	2,158	2,295	998	264	17	1	0	6,952
	2	112	288	949	1,401	801	207	16	3	0	3,777
	3c	421	912	2,945	5,850	4,822	1,588	130	8	0	16,676
	3b	483	783	2,734	7,455	9,415	4,567	394	26	1	25,858
	3a	398	611	1,934	6,456	11,779	8,403	922	64	5	30,572
	4c	795	1,028	2,524	9,650	29,305	38,108	8,480	1,058	59	91,007
	4b	487	502	1,029	3,861	18,825	47,351	21,615	4,415	375	98,460
	4a	308	185	340	1,342	8,845	38,752	35,693	12,562	1,760	99,787
	5c	180	93	112	454	2,901	22,244	43,342	33,394	9,657	112,377
	5b	26	9	8	18	148	1,610	6,838	12,236	7,906	28,799
	5a	0	0	0	2	2	30	214	830	1,345	2,423

Sub-level Achieved in KS2 English	Percentage of Pupils	U	G	F	E	D	C	B	A	A*	TOTAL
	B	6%	13%	29%	29%	17%	6%	1%	0%	0%	100%
	N	5%	13%	31%	33%	14%	4%	0%	0%	0%	100%
	2	3%	8%	25%	37%	21%	5%	0%	0%	0%	100%
	3c	3%	5%	18%	35%	29%	10%	1%	0%	0%	100%
	3b	2%	3%	11%	29%	36%	18%	2%	0%	0%	100%
	3a	1%	2%	6%	21%	39%	27%	3%	0%	0%	100%
	4c	1%	1%	3%	11%	32%	42%	9%	1%	0%	100%
	4b	0%	1%	1%	4%	19%	48%	22%	4%	0%	100%
	4a	0%	0%	0%	1%	9%	39%	36%	13%	2%	100%
	5c	0%	0%	0%	0%	3%	20%	39%	30%	9%	100%
	5b	0%	0%	0%	0%	1%	6%	24%	42%	27%	100%
	5a	0%	0%	0%	0%	0%	1%	9%	34%	56%	100%

Fig. 2 2010 Transition Matrices, from Ofsted/DfE (2010).

Adjusting Expectations with Statistical Models

Estimating future attainment from prior attainment underpins CVA measures, such as those used in *RAISEOnline* and also FFT's statistical models. The complex multi-level modelling involved has been thoroughly documented (FFT 2005, Ray 2006, Kelly and Downie 2010 and DfE 2010c). To summarise briefly, KS2-KS4 CVA is derived by comparing expected attainment to actual attainment for every pupil and averaging across a cohort. Calculating expected attainment from past performance of pupils with similar profiles is where the complexity of the statistical model lies; prior attainment is fine graded and further adjusted for pupil and school contextual factors. The DfE (2010c p.5) points out: 'even when we include contextual factors we find that prior attainment is by far the strongest predictor of outcomes'.

CVA was designed principally to *look back* at past pupil and school performance, but because expected attainment must be calculated to do this, the model can also *look forward* and inform

target-setting. Alongside their 'analyses to support self-evaluation', which include information similar to CVA measures, FFT produces data specifically designed to look forward. Over 90% of secondary schools regularly use FFT's main data source, *FFT Live* (Treadaway 2009). This produces **estimates** of pupils' likely KS2 levels, KS3 levels or GCSE grades, often expressed as percentage chances or probabilities, built on two statistical models – **prior attainment** (PA) and **socio economic** (SE). Both models, by taking into account additional factors, attempt to arrive at more accurate estimates for each pupil. These are aggregated to support target-setting at subject and whole-school levels, as well as for individuals.

The prior attainment (PA) model incorporates decimalised levels using raw marks from national tests (a further refinement to sub-levelling), subject differences in test scores and teacher assessments (TAs). It also uses two basic pupil contextual factors (gender and month of birth) to adjust estimates. FFT's Training Resources (FFT 2011a) exemplify the effect of these extra factors on pupils' percentage chances. Eight pupils achieving Level 4 in KS2 English, Maths and Science tests each have a 23% chance of attaining Level 6 or above at the end of Key Stage 3 in English, using only the national median. When the FFT PA model is applied their Level 6+ chances vary from 4% to 48%, with decimalised levels having the greatest influence on this change. Perhaps counter-intuitively, taking into account month of birth favours younger pupils, because they have higher prior attainment relative to their age. In gender terms, two pupils with identical decimalised levels have very different chances of Level 6+ in KS3 English; girl – 26%; boy – 10%.

The socio-economic (SE) model adds *school* contextual factors to the PA model, including the mean and spread of cohort prior attainment on entry, mobility, free school meals (FSM) ranking and other socio-economic indicators from the school locality. FFT has a third, **school extended** (SX) model, which incorporates further *pupil* contextual factors (such as ethnicity, FSM eligibility etc.). It is important to note that this model is **not** used to look forward. It exists as part of FFT's 'analyses to support self-evaluation', helping schools look back, and is similar to the Ofsted/government CVA model. Fig. 3 summarises the factors used within all three FFT models.

Pupil Factors	PA	SE	SX
Mean Test Level (FG)	✓	✓	✓
Mean TA Level	✓	✓	✓
Subject Variations	✓	✓	✓
Gender	✓	✓	✓
Month of Birth	✓	✓	✓
EAL			✓
FSM			✓
SEN Stage, Statemented			✓
Ethnicity			✓
Mobility (joined late / time in school)			✓
School Factors	**PA**	**SE**	**SX**
Mean Intake Test Level		✓	✓
Spread of Intake Test Level		✓	✓
FSM Entitlement (Percentile Rank)		✓	✓
Geodemographic Data (Percentile Rank)		✓	✓

Fig. 3 Factors included in FFT models (FFT 2011a).

Based on the PA and SE models, four main *types* of FFT estimates have been used in schools:

- **FFT Type A** – using the prior attainment (PA) model, in line with all schools nationally.
- **FFT Type B** – using the socio-economic, i.e. adjusted for similar schools.
- **FFT Type C** – as FFT B, adjusted to meet locally or nationally agreed targets.
- **FFT Type D** – as FFT B, in line with similar schools at the 25th percentile.

In addition to these four types, *FFT Live* enables an 'estimate rank' from 50 to 5 to be applied to PA or SE estimates, adding further challenge. For example, FFT Type D has an estimate rank of 25, while 10 produces estimates in line with the performance of schools at the 10th percentile.

The impact on a pupil's estimated English GCSE grades of applying four estimate types (A, B, D and D+) are shown in the following examples (Fig. 4), using FFT's training database. This particular pupil has an *average* prior attainment profile, having achieved Level 4s at the end of KS2 and Level 5s at the end of KS3 in all three core subjects. Comparing the spread of FFT Type A chances from KS2 for this pupil to the actual 2010 outcomes shown in Fig. 1 reveals a slight downward shift, reflecting low Level 4s (when decimalised) and other adjustments. Also apparent are the relatively small upward shifts resulting from the supposedly more challenging FFT D and D+ estimate ranks.

FFT Type A

KS2/3 Results							KS4 Estimates														
	Test			TA			Est	Subject	% chance of achieving KS4 Grade								% chance		Grades		
KS	EN	MA	SC	EN	MA	SC	Basis	Group	G	F	E	D	C	B	A	A*	A*-C	Pass	GM	HGM	GA
2	4.0	4.2	4.5	4	4	4	PA	English	1 %	2 %	8 %	27 %	44 %	16 %	2 %	1 %	62 %	99 %	C	18 %	B
3				5	5	5	PA	English	1 %	1 %	5 %	20 %	45 %	23 %	5 %	1 %	73 %	99 %	C	28 %	B

The first row of estimates is based on end of KS2 prior attainment, with decimalised levels and TAs shown. The second row is based on end of KS3 prior attainment, with only TAs shown, as this post-dates KS3 SATs. The highlighted Cs show what FFT call the GM grade – 'a 'median' grade or even a 'minimum' grade' – the chance of this grade or higher is at least 50%. In this case it is also the most likely grade for this pupil. The highlighted Bs are one grade above GM, known as GA grades.

FFT Type B

KS2/3 Results							KS4 Estimates														
	Test			TA			Est	Subject	% chance of achieving KS4 Grade								% chance		Grades		
KS	EN	MA	SC	EN	MA	SC	Basis	Group	G	F	E	D	C	B	A	A*	A*-C	Pass	GM	HGM	GA
2	4.0	4.2	4.5	4	4	4	SE	English	1 %	3 %	12 %	35 %	40 %	8 %	1 %	1 %	48 %	99 %	D	48 %	C
3				5	5	5	SE	English	1 %	1 %	5 %	24 %	47 %	19 %	3 %	1 %	69 %	99 %	C	22 %	B

Taking school socio-economic factors into account lowers chances for this pupil, suggesting that the school context is less favourable than average – for example the free school meals ranking might be high. Based on KS2 prior attainment, the GM and GA grades are one less than before. The reason for this is that the chance of grade C or higher has dropped below 50%, even though C is still the most likely grade.

FFT Type D

KS2/3 Results							KS4 Estimates														
	Test			TA			Est	Subject	% chance of achieving KS4 Grade								% chance		Grades		
KS	EN	MA	SC	EN	MA	SC	Basis	Group	G	F	E	D	C	B	A	A*	A*-C	Pass	GM	HGM	GA
2	4.0	4.2	4.5	4	4	4	SE	English	1 %	2 %	10 %	32 %	43 %	11 %	1 %	1 %	55 %	99 %	C	12 %	B
3				5	5	5	SE	English	1 %	1 %	4 %	21 %	47 %	23 %	4 %	1 %	74 %	99 %	C	27 %	B

Adjusting Type B estimates in line with similar schools at the 25th percentile raises the chances, but they are still below the Type A estimates for this particular pupil and school. The GM and GA grades are the same as those for Type A.

FFT Type D+

KS2/3 Results							KS4 Estimates														
	Test			TA			Est	Subject	% chance of achieving KS4 Grade								% chance		Grades		
KS	EN	MA	SC	EN	MA	SC	Basis	Group	G	F	E	D	C	B	A	A*	A*-C	Pass	GM	HGM	GA
2	4.0	4.2	4.5	4	4	4	SE	English	1 %	2 %	8 %	29 %	46 %	14 %	2 %	1 %	61 %	99 %	C	15 %	B
3				5	5	5	SE	English	1 %	1 %	3 %	18 %	46 %	26 %	6 %	1 %	78 %	99 %	C	32 %	B

Adjusting estimates in line with similar schools at the 10th percentile, raises the chances slightly further but not enough to change the GM and GA grades.

Fig. 4 FFT Type A, B, D and D+ estimates in English GCSE for an example pupil (FFTLive 2011).

If an English target for the pupil in Fig. 4 is to be set, a reasonable minimum might be a C, with a more aspirational target of a B and recognition that A is unlikely but not impossible. However, should decisions about targets be based on this data alone? Fischer Family trust has provided clear guidance on school use of estimates, acknowledging the potential for misuse (FFT 2009a):

> FFT estimated grades are estimates, NOT targets and should NOT be imported into school tracking systems without modification by teachers. They should be used to SUPPORT target setting for young people in conjunction with other data and your own professional judgement.

Evidence from observation suggests that some schools' internal target-setting practices follow FFT's guidance. Some go further, arriving at agreed targets owned by all involved, through dialogue between pupils, teachers and sometimes parents, informed by **all** available data. Such schools recognise that conversations need sensitive handling when prior attainment is low and information points to likely future attainment below age-related expectations. Where guidance is overlooked schools simply import median grade estimates into tracking systems, which effectively become targets. In fact, *FFT Live* provides schools with an 'export' feature to 'select and download datafiles for local use, including import into Management Information Systems', though warnings that estimates 'should not be used unmoderated as targets' pervade FFT's documentation.

Observations also confirm FFT's view that school leaders do not always look at the full range of possible grades and their associated chances (Treadaway 2009). Some temper this by explicitly treating median estimates as **minimum** target grades, as suggested by FFT. Some believe that FFT D estimates represent demanding targets, being in line with similar schools at the 25th percentile; others increase estimate ranks further, or set targets at FFT D +1 to incorporate further challenge. In some schools, pupil targets from FFT estimates are given to teachers by senior leaders with little or no consultation, resulting in anxiety about their provenance and the use to which they might be put (Kelley et al 2010 p.36). Pupil targets, whether generated mechanistically or negotiated and informed by professional judgement, can become immutable within schools' tracking systems, irrespective of the on-going progress of pupils for whom they have been set. Some schools observed avoid this through regular target reviews in the light of assessments, increasing targets if pupils are on track to meet or exceed them and intervening if they are falling behind.

Further Examples of Pupil FFT Estimates

Fieldwork observations and FFT's training database provide further examples of pupil estimates, which illustrate their application to target setting and chart modifications made by FFT to minimise misuse.

Pupil Example 1

Fig. 5 reinforces the foregoing point that decimalising levels has a pronounced effect on estimates. It also demonstrates the impact of gender – the boy's GM and GA grades in English are one lower than those of the girl with the same decimalised English KS2 level.

		KS2 Results							KS4 Estimates														
			Test			TA			Est	Subject	% chance of achieving KS4 Grade								% chance		Grades		
DoB	Sex	KS	EN	MA	SC	EN	MA	SC	Basis	Group	G	F	E	D	C	B	A	A*	A*-C	Pass	GM	HGM	GA
28/03/95	F	2	4.1	4.0	4.0	4	4	4	PA	English	1 %	3 %	11 %	31 %	41 %	11 %	1 %	1 %	54 %	99 %	C	13 %	B

Girl achieving a low Level 4 in the KS2 English SAT.

		KS2 Results							KS4 Estimates														
			Test			TA			Est	Subject	% chance of achieving KS4 Grade								% chance		Grades		
DoB	Sex	KS	EN	MA	SC	EN	MA	SC	Basis	Group	G	F	E	D	C	B	A	A*	A*-C	Pass	GM	HGM	GA
14/05/95	F	2	4.9	4.4	4.8	4	4	4	PA	English	1 %	1 %	1 %	6 %	30 %	39 %	20 %	3 %	93 %	99 %	B	23 %	A

Girl achieving a high Level 4 in the KS2 English SAT.

		KS2 Results							KS4 Estimates														
			Test			TA			Est	Subject	% chance of achieving KS4 Grade								% chance		Grades		
DoB	Sex	KS	EN	MA	SC	EN	MA	SC	Basis	Group	G	F	E	D	C	B	A	A*	A*-C	Pass	GM	HGM	GA
24/12/94	M	2	4.9	5.2	4.6	4	4	4	PA	English	1 %	1 %	2 %	12 %	39 %	33 %	11 %	2 %	86 %	99 %	C	46 %	B

Boy achieving a high Level 4 in the KS2 English SAT.

Fig. 5 FFT Type A estimates in English GCSE for example pupils (FFT Live 2011).

Pupil Example 2

Fig. 6 shows two different FFT estimate formats dating back to 2007, based on Y9 (i.e. end of KS3) tests and TAs. The first format, provided as an electronic document, shows only most likely estimated grades for each subject group, with no indication of percentage chances across grades. The second format from *FFT Live*, shows estimates for the same pupil including the full range of chances. It is apparent that some of the median grade estimates differ between the two examples – some of the Ds have become Cs and vice versa. This is because the first format was produced earlier; FFT regularly refine the statistical models and update underlying data sets, reflected in the *live* data. While the intention behind these updates is presumably to improve accuracy, Kelley et al (2010 p.16) point to frustration this can cause school leaders who perceive such changes as 'moving goalposts'. This reinforces the need to review targets based on FFT estimates, both when FFT updates are made and when new school-based data becomes available.

Comparing the second table in Fig. 6 (below) to those shown in Fig. 5 (above) reveals further presentational refinements made by FFT. GM and GA grades and their associated shading, not present in the Fig. 6 example, are a recent development clearly designed to encourage schools to take notice of not just the median grade but the next grade too. This example also illustrates the risk of adopting estimates as targets without moderation, with the case of History most striking; the chances of D and C are the same. It is inconceivable that this pupil should be set a History target of D, but this could happen if only median grades are used. In a school where pupils' estimates/targets are used to inform ability sets and tiers of exam entry, an unnecessary ceiling might be placed on future attainment.

Y9 Test / TA Levels			Y11 Estimates				Total	Estimated Grades (in subject groups) .... Note : + signifies A*											
EN	MA	SC	CS	5AC	5AC(EM)	5AG	Points	AD	CA	EN	GE	HI	HU	IT	LA	MA	PE	SC	TE
5.5	6.1 5	5.4	22%	69%	33%	98%	370	C	C	C	C	D	C	D	D	C	C	C	C

Decimalised KS3 test levels are shown, along with a TA for Maths, because at Level 5 this differs from the test Level 6; for English and Science the test and TA levels (as whole levels) are the same. CS = chances of achieving C+ in English, Maths and Science; 5AC = chances of 5+ A-C GCSE or equivalent grades; 5AC(EM) = chances of 5+ A*-C GCSE or equivalent including English and Maths; 5AG = chances of 5+ A*-G GCSE or equivalent grades; AD = art and design; CA = creative arts; EN =*

English; GE = geography; HI = history; HU = humanities; IT = information technology; LA = languages; MA = maths; PE = physical education; SC = science; TE = technology.

KS3 Results							KS4 Estimates											
	Test			TA			Est	Subject	% chance of achieving KS4 Grade								% chance	
KS	EN	MA	SC	EN	MA	SC	Basis	Group	G	F	E	D	C	B	A	A*	A*-C	Pass
3	5.5	6.1	5.4	5	5	5	SE	ArtDes	1 %	2 %	6 %	13 %	37 %	27 %	13 %	2 %	79 %	99 %
3	5.5	6.1	5.4	5	5	5	SE	CreatArt	2 %	5 %	10 %	18 %	32 %	25 %	8 %	1 %	66 %	98 %
3	5.5	6.1	5.4	5	5	5	SE	D&T	1 %	3 %	9 %	24 %	39 %	16 %	6 %	1 %	62 %	98 %
3	5.5	6.1	5.4	5	5	5	SE	English	1 %	1 %	6 %	25 %	48 %	18 %	2 %	1 %	69 %	99 %
3	5.5	6.1	5.4	5	5	5	SE	Geog	1 %	4 %	14 %	31 %	40 %	8 %	1 %	1 %	50 %	99 %
3	5.5	6.1	5.4	5	5	5	SE	History	3 %	9 %	19 %	27 %	27 %	13 %	2 %	1 %	43 %	99 %
3	5.5	6.1	5.4	5	5	5	SE	Hu'ties	2 %	4 %	11 %	21 %	33 %	22 %	7 %	1 %	63 %	99 %
3	5.5	6.1	5.4	5	5	5	SE	ICT	2 %	6 %	11 %	21 %	36 %	18 %	5 %	1 %	60 %	97 %
3	5.5	6.1	5.4	5	5	5	SE	Maths	1 %	1 %	12 %	41 %	41 %	5 %	1 %	1 %	48 %	98 %
3	5.5	6.1	5.4	5	5	5	SE	MFL	1 %	4 %	13 %	31 %	36 %	12 %	3 %	1 %	52 %	99 %
3	5.5	6.1	5.4	5	5	5	SE	PE	1 %	2 %	11 %	32 %	26 %	19 %	8 %	1 %	55 %	99 %
3	5.5	6.1	5.4	5	5	5	SE	Science	1 %	5 %	21 %	43 %	29 %	1 %	1 %	1 %	33 %	99 %

Shaded percentage chances indicate the median grades and the grades completely within the top 25% of possible likely outcomes. Cumulative chances of A-C grades and passes (A*-G) are shown.*

Fig. 6 Two formats of FFT Type D Pupil Estimates from KS3-4 (FFT, FFT Live 2007).

Pupil Example 3

Fig. 7 shows a pupil who made considerable progress during Key Stage 3 in all three core subjects, particularly Maths. Based on the end of Key Stage tests (pre-dating end of KS3 SATs) the pupil made four levels progress in Maths during the three years and based on teacher assessments, three levels progress. This better than expected progress in Maths has had a profound impact on their estimated chances of success at GCSE, the 'most likely' grade shifting from a D to a B and the chances of an A*-C grade increasing from 35% to 95%.

This example highlights potential risks in using only end of KS2 national test results to estimate future attainment five years on; the progress made in the early secondary years is ignored. This is particularly pertinent, with national KS3 testing ended and KS3 in many schools becoming compressed to allow earlier starts to KS4 courses. The data suggests that this pupil's strong progress through KS3 should inform a far more ambitious target than their primary school level alone would indicate.

KS2/3 Results							KS4 Estimates											
	Test			TA			Est	Subject	% chance of achieving KS4 Grade								% chance	
KS	EN	MA	SC	EN	MA	SC	Basis	Group	G	F	E	D	C	B	A	A*	A*-C	Pass
2	4.7	3.8	4.5	4	4	4	SE	Maths	2 %	8 %	21 %	34 %	31 %	4 %	1 %	1 %	35 %	99 %
3	6.1	7.1	6.0	5	7	6	SE	Maths	1 %	1 %	1 %	4 %	39 %	45 %	11 %	1 %	95 %	99 %

Fig. 7 FFT Type D Pupil Estimates (FFT Live 2009).

Pupil Example 4

Fig. 8 shows a pupil with median grades close to the highly significant C-D borderline. Some schools observed make effective use of such data to motivate pupils working below most likely estimates, who are therefore at risk of underachievement. Sensible and sensitive sharing of FFT chances with pupils unsure of their own potential has been observed to boost self-esteem and re-energise effort, especially when they understand that past national results underpin the data.

This example also allows comparison of English estimates with those of arts and technology subject groups. Teachers of foundation subjects often query the validity of prior attainment in core subjects as a basis for estimates in their disciplines (Mansell 2008). The chances generated for arts and technology subjects in Fig. 8 are slightly more spread out than the English estimates, as might be expected. The proportion of grades including and either side of the median for the three groups of foundation subjects shown ranges from 74% to 81% (compared to 88%/93% for English). The actual subjects included in each group demand closer consideration, with the full range documented in FFT's Training Resources (FFT June 2011a):

- **ArtDes** includes Art, Art and Design (many variants), Fine Art, History of Art and Craft.
- **CreatArt** includes Creative Arts, Music, Music Technology, History of Music, Dress and Jewellery.
- **D&T** includes Design and Technology (many variants), Business Studies, Motor Vehicle Studies, Engineering, Graphics.
- **English** includes English Language, English Literature, Drama, Theatre Studies, Media Studies, Film Studies.

FFT point out that KS4 subject groups are used because there are too many individual subjects to produce separate estimates and pupils' individual subject choices are unknown. They also acknowledge that pupils may have 'specific aptitudes or skills' in some subjects, that 'estimates may be less relevant' and 'professional judgement and common sense should be used alongside the data' (FFT 2011b).

KS2/3 Results							KS4 Estimates											
	Test			TA			Est	Subject	% chance of achieving KS4 Grade								% chance	
KS	EN	MA	SC	EN	MA	SC	Basis	Group	G	F	E	D	C	B	A	A*	A*-C	Pass
2	4.4	4.1	4.9	3	4	4	SE	ArtDes	1 %	2 %	6 %	13 %	37 %	26 %	13 %	3 %	79 %	99 %
3				5	5	5	SE	ArtDes	1 %	2 %	6 %	14 %	40 %	25 %	10 %	2 %	77 %	99 %
2	4.4	4.1	4.9	3	4	4	SE	CreatArt	2 %	4 %	10 %	19 %	31 %	24 %	9 %	1 %	65 %	99 %
3				5	5	5	SE	CreatArt	2 %	5 %	12 %	22 %	33 %	21 %	5 %	1 %	59 %	99 %
2	4.4	4.1	4.9	3	4	4	SE	D&T	2 %	3 %	9 %	21 %	36 %	19 %	9 %	1 %	65 %	98 %
3				5	5	5	SE	D&T	1 %	3 %	9 %	24 %	40 %	17 %	6 %	1 %	62 %	98 %
2	4.4	4.1	4.9	3	4	4	SE	English	1 %	1 %	5 %	20 %	43 %	25 %	5 %	1 %	74 %	99 %
3				5	5	5	SE	English	1 %	1 %	4 %	23 %	50 %	20 %	2 %	1 %	72 %	99 %

Fig. 8 FFT Type D Pupil Estimates (FFT Live 2010).

Pupil Example 5

Fig. 9 represents a lower attaining pupil, whose end of KS2 and end of KS3 assessments were below age-related expectations. Two versions of the same set of estimates are shown for a selection of subject groups, the first with percentage chances of each grade, the second with

cumulative chances of that grade or higher. The low probability of achieving A*-C grades in most of the subjects shown has profound implications for this pupil. While the focus on C or above (i.e. Level 2 within the National Qualifications Framework) has dominated efforts to raise GCSE standards in England, grades D to G (i.e. NQF Level 1) remain GCSE passes. Failure to achieve grade C or above is often perceived as just that – failure. Yet for the stage of development of this pupil, Level 1 is the most probable KS4 outcome. Arguably, this pupil requires a KS4 curriculum that leads to Level 1 success, as opposed to Level 2 failure, followed by post-16 opportunities to progress to Level 2, Level 3 and beyond. However, the estimates indicate that in some subjects attaining grade C or higher is improbable, but not impossible.

KS2/3 Results							KS4 Estimates														
	Test			TA			Est	Subject	% chance of achieving KS4 Grade								% chance		Grades		
KS	EN	MA	SC	EN	MA	SC	Basis	Group	G	F	E	D	C	B	A	A*	A*-C	Pass	GM	HGM	GA
2	3.5	3.6	3.6	3	3	3	SE	Hu'ties	24 %	28 %	23 %	16 %	8 %	1 %	1 %	1 %	9 %	85 %	F	48 %	E
3				4	4	4	SE	Hu'ties	18 %	24 %	24 %	19 %	12 %	3 %	1 %	1 %	15 %	90 %	E	34 %	D
2	3.5	3.6	3.6	3	3	3	SE	ICT	13 %	19 %	21 %	23 %	21 %	3 %	1 %	1 %	24 %	86 %	E	47 %	D
3				4	4	4	SE	ICT	11 %	17 %	20 %	24 %	24 %	4 %	1 %	1 %	28 %	89 %	D	28 %	C
2	3.5	3.6	3.6	3	3	3	SE	Maths	19 %	33 %	28 %	15 %	5 %	1 %	1 %	1 %	5 %	95 %	F	48 %	E
3				4	4	4	SE	Maths	16 %	32 %	29 %	17 %	6 %	1 %	1 %	1 %	6 %	96 %	E	23 %	D
2	3.5	3.6	3.6	3	3	3	SE	MFL	17 %	29 %	28 %	19 %	7 %	1 %	1 %	1 %	7 %	96 %	E	26 %	D
3				4	4	4	SE	MFL	13 %	25 %	28 %	23 %	9 %	1 %	1 %	1 %	11 %	97 %	E	34 %	D
2	3.5	3.6	3.6	3	3	3	SE	PE	3 %	13 %	30 %	35 %	13 %	4 %	1 %	1 %	18 %	99 %	D	18 %	C
3				4	4	4	SE	PE	2 %	10 %	27 %	37 %	17 %	6 %	1 %	1 %	24 %	99 %	D	24 %	C
2	3.5	3.6	3.6	3	3	3	SE	Science	14 %	27 %	31 %	20 %	8 %	1 %	1 %	1 %	8 %	96 %	E	28 %	D
3				4	4	4	SE	Science	10 %	22 %	31 %	25 %	12 %	1 %	1 %	1 %	12 %	97 %	E	38 %	D

Individual chances of each grade.

KS2/3 Results							KS4 Estimates														
	Test			TA			Est	Subject	% chance of achieving KS4 Grade								% chance		Grades		
KS	EN	MA	SC	EN	MA	SC	Basis	Group	G+	F+	E+	D+	C+	B+	A+	A*	A*-C	Pass	GM	HGM	GA
2	3.5	3.6	3.6	3	3	3	SE	Hu'ties	99 %	76 %	48 %	25 %	9 %	2 %	1 %	1 %	9 %	85 %	F	48 %	E
3				4	4	4	SE	Hu'ties	99 %	82 %	58 %	34 %	15 %	3 %	1 %	1 %	15 %	90 %	E	34 %	D
2	3.5	3.6	3.6	3	3	3	SE	ICT	99 %	87 %	68 %	47 %	24 %	3 %	1 %	1 %	24 %	86 %	E	47 %	D
3				4	4	4	SE	ICT	99 %	89 %	72 %	52 %	28 %	4 %	1 %	1 %	28 %	89 %	D	28 %	C
2	3.5	3.6	3.6	3	3	3	SE	Maths	99 %	81 %	48 %	20 %	5 %	1 %	1 %	1 %	5 %	95 %	F	48 %	E
3				4	4	4	SE	Maths	99 %	84 %	52 %	23 %	6 %	1 %	1 %	1 %	6 %	96 %	E	23 %	D
2	3.5	3.6	3.6	3	3	3	SE	MFL	99 %	83 %	54 %	26 %	7 %	1 %	1 %	1 %	7 %	96 %	E	26 %	D
3				4	4	4	SE	MFL	99 %	87 %	62 %	34 %	11 %	1 %	1 %	1 %	11 %	97 %	E	34 %	D
2	3.5	3.6	3.6	3	3	3	SE	PE	99 %	97 %	83 %	53 %	18 %	5 %	1 %	1 %	18 %	99 %	D	18 %	C
3				4	4	4	SE	PE	99 %	98 %	88 %	61 %	24 %	7 %	1 %	1 %	24 %	99 %	D	24 %	C
2	3.5	3.6	3.6	3	3	3	SE	Science	99 %	86 %	59 %	28 %	8 %	1 %	1 %	1 %	8 %	96 %	E	28 %	D
3				4	4	4	SE	Science	99 %	90 %	68 %	38 %	12 %	1 %	1 %	1 %	12 %	97 %	E	38 %	D

Cumulative chances of each grade plus all those higher.

Fig. 9 FFT Type D Pupil Estimates (FFT Live 2010).

Conclusions

Statistical models principally designed for looking back at past performance and providing fairer measures of school effectiveness have the inherent capacity to look forward, predicting future outcomes. The risk that this dual purpose might lower expectations has led to the abolition of CVA, despite the view of some that it is 'the best measure of the quality of education provided by a

school' (Wiliam 2008). It is likely that Fischer Family Trust, while reviewing their approach, will continue to provide data for schools supporting both purposes (FFT 2011c). Of the additional pupil and school information deployed to provide sharper estimates of expected attainment than those possible from broad past performance patterns, most influential are test marks used to finely grade prior attainment. The fact that certain contextual factors are also applied to FFT estimates is highly relevant to their use in pupil target setting. Three fundamental questions arise:

1. Are fine graded levels derived from end of Key Stage tests sufficiently reliable to justify their use in refining estimates of expected attainment?
2. Is it right to change expectations of pupil attainment based on gender and date of birth, even if past performance suggests that these play a part?
3. Should expectations of pupil attainment be influenced by the social context of the school, for example the level of free school meals eligibility?

Reliability of national tests, defined as the extent to which they 'can be trusted to give consistent information' (Mansell et al 2006 p.12), has been debated extensively. One way to determine reliability of tests where pupils' marks are to be categorised into levels is to assess accuracies of classification, which Oppos and He (2011 p.13) state 'are generally about 90% for the mathematics tests, 87% for the science tests, and 85% for the English tests'. This suggests that as many as 15% of pupils each year (approximately 87,000 individuals) might be awarded an incorrect level in KS2 English tests. Against this background, using sub-levels or decimalised levels to adjust estimates looks dubious. Guidance has urged 'caution' with sub-levels, which carry no agreed or approved definitions as National Curriculum levels were intended to 'indicate representative attainment' (NS/DCSF 2011). Newton (2009) points to complex limitations and inherent uncertainty in standardised tests, concluding that 'although results from national curriculum testing are somewhat unreliable, they are not necessarily too unreliable'. This surely depends on the purposes to which KS2 test results are to be put. Bew (2011 p.10-11) recommends just three: holding schools to account; informing parents and secondary schools of pupil progress; benchmarking between schools and monitoring performance nationally and locally. Doyle and Godfrey (2005 p.42) have questioned their predictive use by secondary teachers.

Related to reliability, the **validity** of KS2 tests is also open to question in terms of 'being clear about what exactly we are measuring' (Wiliam 2001 p.17). In the case of English, only two of the three National Curriculum Attainment Targets are assessed by the tests – *reading* and *writing*, but not *speaking and listening* (DfEE/QCA 1999, Stobart 2009 p.168). Kelly & Downie's (2010 p.196) conclusion that CVA scores 'suggest a degree of precision in the measurement of school performance that is not justified' seems equally applicable to FFT estimates, particularly as they have similar origins. Furthermore, questionable precision appears to be applied unevenly, with the decimalising of levels at one extreme, and the blurring of subject differences through grouping at the other. The clear correlation between fine-graded levels and GCSE grades seen in past performance patterns may be more a consequence of the tests themselves than the National Curriculum Level Descriptors and GCSE course specifications they partially assess.

The rationale for further refining estimates of pupil performance using school and pupil contextual factors is more plausible when merely looking back. In relation to CVA, Ofsted (2008 p.7) have warned that 'mis-using the data to predict future performance could depress expectations of groups of pupils that have performed less well in previous years'. Although *RAISEOnline* can generate pupil targets from transition matrices (DfE/Ofsted 2010), no contextual factors are included. As shown above, some FFT estimate types incorporate a subset of the available pupil

and school level factors and it is clear that these can affect individual pupil chances. Lowering expectations of a pupil because of their gender or school context is surely unjustifiable. Narrowing attainment gaps among pupil groups, for example by gender or free school schools eligibility, has been a consistent policy aim (NS/DCSF 2009, DfE 2010b), yet allowing such factors to influence targets risks perpetuating these gaps.

Observations, supported by Kelley et al (2010), suggest that staff in schools do not always have sufficient knowledge of how expected attainment estimates are derived. FFT has recognised that data misuse is 'sufficiently common to be a concern' (Treadaway 2009). Limitations need to be realised and understood, to prevent mechanistic generation of fixed pupil targets and their aggregation to subject and school targets. A particular problem is the tendency to base targets on median estimated grades, potentially **lowering** expectations by introducing false ceilings and self-fulfilling prophecies. Hattie (2009 p.44) warns that 'expectations of success (which are sometimes lower than students could attain) may become a barrier for some students as they may only perform to whatever expectations they already have of their ability'. This can be avoided through an 'understanding of the 'provisional' rather than 'literal' nature of data, from all stakeholders within and outside the school' (Kelley et al 2010 p.22). Used carefully, predictive data can help to raise the aspirations of certain pupils, but the refinements applied by FFT are unnecessary for this purpose and in some cases counterproductive.

Most importantly, Mansell et al (2009 p.9) offer a reminder that 'marks, levels, judgmental comments or the setting of targets, cannot, on their own, be formative. Pupils may need help to know **how** they can improve.' The termination of statutory school targets in 2012 provides an opportunity to revisit Harlen and Deakin Crick's (2002 p.8) recommendation of ten years earlier to 'end the practice of basing targets only on test results'. The effects on pupils, teachers and schools of commonplace target-setting cultures warrant further research, though the concomitant prevalence of data-led assessment practice is well documented (for example ARG 2002, Mansell et al 2009, Watkins 2010). It is likely that over-emphasising quantitative pupil targets expressed as levels or grades, at the expense of qualitative targets couched as success criteria, promotes performance over learning. Watkins (2010 p.3-4) concludes that '…a focus on learning can enhance performance, whereas a focus on performance (alone) can depress performance'. Above average performance – exceeding today's average so that next year's is higher – might be better realised by extending expectations across the full range of possible outcomes and learning from the experiences of positive (and negative) outliers.

References

Assessment Reform Group (2002) *Testing, Motivation and Learning*. Cambridge: ARG.

Bew, P. (2011) *Independent Review of Key Stage 2 testing, assessment and accountability - final report*. London: Crown.

Department for Education and Employment / Qualifications and Curriculum Authority (1999) *The National Curriculum – Handbook for primary teachers in England.* London: Crown, DfEE & QCA.

Department for Education (2010a) *Guidance for Local Authorities and Schools on Setting Education Performance Targets for 2012.* http://www.education.gov.uk/schools/adminandfinance/schooladmin/target-setting/a0066387/2012-target-setting-guidance (Accessed May 2011).

Department for Education (2010b) *The Importance of Teaching – The Schools White Paper 2010*. Norwich: The Stationery Office.

Department for Education (2010c) *A Technical Guide to Contextual Value Added (Including English & Maths) Key Stage 2 to 4 - 2010 Model.* http://www.education.gov.uk/performancetables/schools_10/cvacalc.pdf (Accessed May 2011)

Department for Education (2010d) *Performance Tables 2010*. http://www.education.gov.uk/performancetables/ (Accessed July 2011)

Department for Education (2011a) *Ending of statutory school targets*. http://www.education.gov.uk/schools/adminandfinance/schooladmin/target-setting/a0074763/ending-of-statutory-school-targets (Accessed May 2011)

Department for Education (2011b) *Statistical First Release 01/2011: GCSE and Equivalent Results in England 2009/10 (Revised)*. London: DfE.

Department for Education (2011c) *Statistical First Release 10/2011: Percentage of Pupils Making Expected Progress in English And Mathematics Between Key Stage 2 And Key Stage 4 In England: 2007/08 – 2009/10*. London: DfE.

Doyle, L. & Godfrey, R. (2005) *Investigating the reliability of the Key Stage 2 test results for assessing individual pupil achievement and progress in England*. London Review of Education, 3:1, 29-45.

Fischer Family Trust Live – www.fftlive.org.uk (Accessed 2007-11).

Fischer Family Trust (2005) *Data Analysis Project – Methodology*. FFT. http://www.fischertrust.org/ (Accessed June 2011)

Fischer Family Trust (2007) *Making best use of FFT estimates*. FFT. http://www.fischertrust.org/ (Accessed June 2011)

Fischer Family Trust (2009a) *FFT Live Development Reports: Key Features*. FFT. http://www.fischertrust.org/ (Accessed June 2011)

Fischer Family Trust (2009b) *Using FFT Live 3.0 – How FFT data can help your school improve*. FFT. http://www.fischertrust.org/ (Accessed June 2011)

Fischer Family Trust (2011a) *Summary Secondary Training & Support Materials.* FFT. http://www.fischertrust.org/ (Accessed June 2011)

Fischer Family Trust (2011b) *FFT Live: An Interactive Guide to KS4 Subject Estimates*. FFT. http://www.fischertrust.org/ (Accessed June 2011)

Fischer Family Trust (2011c) *Materials from Regional Meetings*. FFT. http://www.fischertrust.org/ (Accessed June 2011)

Gove, M. (2011) interviewed by Hurst, G., Sylvester, R. and Thomson, A. in The Times 18th June 2011. London: The Times.

Harlen, W. & Deakin Crick, R. (2002) *A systematic review of the impact of summative assessment and tests on students' motivation for learning* (EPPI-Centre Review, version 1.1). In: Research Evidence in Education Library. Issue 1. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Hattie, J. (2009) *Visible Learning – A synthesis of over 800 meta-analyses relating to achievement.* London: Routledge

Kelley, A. & Downie, C. (2010) *Value-added measures for schools in England looking inside the 'black box' of complex metrics*. Educational Assessment, Evaluation and Accountability 22:181–198.

Kelley, A., Downie, C. & Rietdijk, W. (2010) *Data dictatorship and data democracy: understanding professional attitudes to the use of pupil performance data in English secondary schools*. CfBT.

Mansell, W. (2008) *Fischer data under growing scrutiny*. Times Educational Supplement 12th December 2008. London: TSL Education Ltd.

Mansell, W., James, M. & the Assessment Reform Group (2009) *Assessment in schools. Fit for purpose? A Commentary by the Teaching and Learning Research Programme.* London: Economic and Social Research Council, Teaching and Learning Research Programme.

National Strategies/Department for Children, Families and Schools (2009) *Narrowing the Gaps: from data analysis to impact – A practical guide*. London: Crown, DCSF.

National Strategies (2011) *Sublevels*. http://nationalstrategies.standards.dcsf.gov.uk (Accessed May 2011, archived).

Newton, P. E. (2009) *The reliability of results from national curriculum testing in England*. Educational Research, 51:2, 181-212.

Ofsted (2005) *Using the evaluation schedule.* London: Ofsted.

Ofsted (2008) *Using data, improving schools.* London: Ofsted.

Ofsted (2011) *Evaluation Schedule for Schools. April 2011, No. 090098*. London: Ofsted.

Ofsted/DfE (2010) *Transition matrices for Key Stage 2 to 4 – 2010*. https://www.raiseonline.org/documentlibrary (Accessed July 2011)

Opposs, D. & He, Q. (2011) *The Reliability Programme – Final report*. Crown, Ofqual.

Ray, A. (2006) *School Value Added Measures in England*. London: DfES.

Stewart, W. (2011) *Thousands of schools could fall foul of new Ofsted regime*. Times Educational Supplement 10th June 2011. London: TSL Education Ltd.

Stobart, G. (2009) *Determining validity in national curriculum assessments*. Educational Research, 51:2, 161-179.

Treadaway, M. (2009) *Data analysis*. Leader: The Education Leader Magazine, ASCL.

Watkins, C. (2010) *Learning, Performance and Improvement*. London: INSI-IoE.

Wiliam, D. (2001) *Reliability, validity, and all that jazz*. Education 3-13, 29:3, 17-21.

Wiliam, D. (2008) *Context beats exam tables in search for best schools*. Education Guardian, Wednesday 20 February 2008. Guardian News and Media Limited.