**The Role of External Inspection in the Public Services:
The Case of the UK Training Market***

**Accepted for publication in *Public Administration*
September 2008**

Abstract

We consider two interpretations of the role of external inspection in the public services, in the context of publicly funded work-based training programmes for young people. The first is that inspection provides substantive information to buyers concerning training quality, thereby improving efficiency in the 'training market'. The second is that it provides procedurally-oriented reassurance concerning service quality to government and the public, irrespective of substantive quality. Evidence is drawn from the inspection procedures and reports of the Adult Learning Inspectorate between 2001 and 2005. The inspectors rated training providers on various attributes, some clearly procedural, others potentially substantive. We find that while inspectors took both procedural and substantive dimensions of training into account in judging the quality of a provider's services, they attached considerably more weight to procedural than to substantive attributes. In particular, they undervalued the trainee completion rate, despite its potential association with the substantive quality of training and the priority that the government attaches to raising it. These findings are interpreted as evidence of limited validity in inspection findings, which do little to resolve information asymmetries in the UK training market.

'Driving up achievement must remain the priority for work-based-learning.'

David Sherlock, Chief Inspector of Adult Learning (ALI, 2003a: 13)

1. INTRODUCTION

One of the main goals of public service reform in the UK is higher service quality. In that respect, public programmes of work-based training for young people – a service organised primarily through the (Modern) Apprenticeship programme – have caused recurrent concern, as instanced recently in the creation of the Quality Improvement Agency (QIA 2007). The Apprenticeship programme has displayed low completion rates, poor training standards, and adverse inspection results for many providers (Ryan and Unwin 2001; ALI 2002; Ryan, Gospel and Lewis 2007). The existence of a serious quality problem has been recognised repeatedly by the three leading public bodies with responsibility for work-based learning, namely the (former) Department for Education and Skills, the Learning and Skills Council and the Adult Learning Inspectorate (henceforth DfES, LSC and ALI, respectively) (LSC 2003a, 2003b; ALI 2002b, 2006; DfES 2005: 20).

The government has sought to raise the quality of training in a variety of ways, prominent amongst which has been an expansion of the external inspection of service providers, in particular through the creation in 2001 of the Adult Learning Inspectorate. One rationale for this policy is the belief that inspection increases the quality of the training on offer within the quasi-market, by reducing *informational asymmetries* between buyers and sellers. Inspection potentially informs buyers – whether a public agency or individual trainees – about the quality of particular providers' offerings, thereby making it possible for purchasers to divert their custom from low to high quality providers. As a result, low quality providers lose business, the prospect of which can give all providers an incentive to offer high quality, so that average quality rises (Laffont and Tirole 1993).

The scope for inspection to play such a role is, however, subject to two constraints. The first is the nature of service quality itself, as a multi-faceted, elusive and contested phenomenon. Second, inspection is itself potentially weakened by the very

informational problem that it is intended to redress, namely the difficulty of observing the quality of an organisation's services from the outside. In particular, according to Power (1999), it is easier for inspectors to observe (and remedy deficiencies in) the *procedural* characteristics of training providers – most notably their managerial processes in general and quality assurance systems in particular – than it is for them to observe (and improve) *substantive* dimensions of training, such as the quality of the teaching on offer and the amount of trainee learning actually achieved. As we shall see, such considerations lead to an alternative interpretation, according to which the principal function of inspection is not so much to overcome informational failure as it is to soothe public concern about the quality of the public services, by ensuring that (what are deemed to be) appropriate managerial systems and procedures are in place. The problem is that, while such procedural attributes may be related to substantive service quality, the links appear to be at best indirect and weak.

Drawing on evidence provided by the reports published by ALI between 2001 and 2005 on the operations of individual contractors to the Apprenticeships programme, this paper addresses these contrasting interpretations of inspection. Two sets of issues are investigated. First, to the extent that service quality is heterogeneous, how closely are the various dimensions related to each other? In particular, how much overlap is there between the procedural and the substantive dimensions of quality? Second, when inspection reaches an overall judgement on the quality of a provider's services, on which dimensions does it focus? Does it emphasise procedural rather than substantive attributes? In particular, what role do trainee completion rates play as a potential indicator of substantive quality, given in particular that the government has repeatedly expressed the wish that they should rise?

In assessing the extent to which quality assurance procedures have become divorced from substantive outcomes in one important sector, this paper responds to Power's remark, apropos the central hypothesis of his path-breaking 1999 book, that 'the empirical challenge of the audit explosion idea is to focus research and data-gathering effort on the evidence of negative consequences' (2003: 191). Viewed more broadly, our project responds to the 'need for more evaluations and other studies of quality programmes' in the public services, and to calls for more research on the regulation of the

public services (Øvretveit 2005: 553; Boyne, Day and Walker 2002). As the approach taken by inspection varies from service to service (Hughes, Mears and Winch 1997), and even within particular services (Day and Klein 1990), its operation in an activity as large as public training programmes is of some importance.

Our research methods are statistical and textual analysis. Statistical associations are calculated across ALI inspection reports between the overall verdict (Adequacy) and the detailed attributes rated by the inspectors. Multivariate probit regression is used to determine the contribution of those attributes to the inspectors' verdicts. Supplementary evidence is taken from the criteria and rules published by the Inspectorate.

The paper proceeds as follows. Section 2 discusses the analytical issues, followed by the context and content of the training quasi-market in section 3, and in section 4 the source and implications of the data. The results are reported in sections 5 and 6, and the conclusions in section 7.

2. ANALYTICAL FRAMEWORK

Training programmes operate in the UK within a quasi-market, wherein a variety of providers, including employers, specialist training companies, charities and public bodies, compete for publicly funding training contracts (Le Grand and Bartlett 1993; Lewis 2008). Recent years have seen quasi-markets play an increasing role in the delivery of a wide variety of public services. The quasi-market for training programmes has functioned for longer and with more similarity to regular markets than have its better-known counterparts in health services and compulsory education (Ryan and Unwin 2001). Competition for contracts is expected to encourage potential providers to minimise service costs, to tailor services more closely to users' requirements, and to develop new services and new ways of providing existing services. However, the commercial contracting involved in such markets may not generate an acceptable level of service quality. A central problem, the existence of which, at least on some interpretations, provides a rationale for inspection, is the difficulty that buyers have in observing service

quality, and the resulting scope for opportunistically-minded contractors to increase profits by reducing quality.

The role of inspection in a quasi-market for publicly-funded training services can be analysed along two lines. The first is economic: under asymmetric information, inspection potentially remedies the failure of market-based contracting to deliver an efficient level of service quality. The second is socio-political: put crudely, the role of inspection is to paper over the cracks, by focusing on organisational attributes of suppliers that are at most weakly related to, and possibly irrelevant or even damaging to, service quality. Inspection reassures the public that all is well, whether or not that is the case in practice. This section discusses the two interpretations in turn.

2.1 THE ECONOMICS OF INFORMATION

The output of a public service may be taken to have both quantity and quality dimensions. For training programmes, quantity is typically represented by the number of trainees or the total time spent in training. Human capital theory suggests that training quality can in principle be defined uni-dimensionally, as the amount learned or the amount of skill produced per unit of training quantity (e.g., trainee year) (Becker 1993; Ryan 1994). In practice, however, two informational problems affect the quality of training.

First, there is disagreement on the meaning of learning and skill, and on the aspects of learning and skill that matter in a particular programme, the upshot of which is that the definition of high quality training is intrinsically amorphous and contested (Kirkpatrick and Martinez-Lucio 1995: 8-9; Cullingford 1999). An example would be the dispute in the UK over the need for the separate teaching and assessment of underpinning knowledge, through Technical Certificates, in the training syllabi ('frameworks') recognised for the Apprenticeships programme. Second, even if the definition of quality were agreed, the actual quality of training offered by a programme might be difficult to observe and measure. In programmes that involve the certification of learning, the number and level of the qualification(s) attained by trainees in principle capture both the quantity and quality of training. Certainly, completion rates (i.e., qualifications attained)

feature regularly in policy documents as the principal indicator of training quality (e.g., ALI 2006: 1). However, much learning may not be certified, and the validity and reliability of the assessment of the learning that *is* certified may be open to question. Training quality tends therefore to be measured also by reference to other factors, including the attributes of teaching staff (e.g. qualifications of trainers), materials and equipment (LSC 2005; ONS 2006: 30-32). This is an issue to which we return in Section 5 below.

Economic theory concentrates on the measurement problem, implicitly assuming the existence of an agreed, uni-dimensional definition of service quality. Quality and quantity are assumed to be substitutable attributes of training in the sense that, for a given level of expenditure, more of one can be provided only by reducing the other. The buyers and sellers of public services are assumed to be motivated by economic self-interest. If it were possible to acquire information about the quality of the services on offer in the quasi-market at zero cost, so that the quality of service provided was not only observable by purchasers and providers but was also *verifiable*, in the sense that it could be unequivocally demonstrated to third parties such as a court of law, then purchasers could regulate service quality by contractual means, simply suing providers who failed to provide the quality of training specified in the contract. In such circumstances, a competitive quasi-market would in principle produce an efficient mix of service quantity and quality, the actual blend of the two depending on the buyer's preferences and the relative cost of the two dimensions of output (Barker *et al.* 1997; Chalkey and Malcomson 2000).

In reality, however, information about quality is likely to be costly and asymmetric. Providers are likely to have better information about quality than either purchasers or third parties such as tribunals and courts, so that quality is *not* verifiable. Contracts are then necessarily incomplete, being written contingent on the quantity but not the quality of the services provided. Under such circumstances, the efficient level of service quality will be forthcoming under a quasi-market regime only under specific, and restrictive, conditions. The central issue concerns the type of contract used. Fixed-price contracts – that is, ones in which payments to the supplier depend only on the *quantity* of training provided – do not reward providers for offering higher *quality*, so that producers

offer what from society's point of view is an inefficiently low level of service quality (Propper 1993: 37-38; Chalkey and Malcomson 2000).

An inefficient outcome is not inevitable, however, because, so long as the degree of informational asymmetry is not too severe, a high quality of service may be sustained by providers' concern for their reputation. The reason is that, although service quality may not be *verifiable*, precluding the use of enforceable contracts to maintain high quality, it may nevertheless be *observable* by purchasers; that is to say, once they have purchased a service from a particular provider, purchasers will learn about the quality of the service it offers. In that case, while the fact that service quality is not verifiable implies that low quality producers cannot be sued by dissatisfied clients, such producers may be penalised nonetheless, both because dissatisfied clients can take their custom elsewhere when next they purchase training, and also because providers of poor quality training acquire a poor reputation that deters new purchasers from using them (Nelson 1970). If providers' finances are sufficiently damaged by the acquisition of a poor reputation, and the consequent loss of both new and repeat business, then providing high quality services may be their long-run profit-maximising strategy, even though training quality is not verifiable (Barker *et al*. 1997: 91-94; Macleod 2007: sections 4-5).

The scope for reputational considerations to play a role in maintaining service quality seems likely to be greatest in those quasi-markets where third-party purchasers buy services from providers on behalf of final users. Such third party purchasers, notably the Learning and Skills Council, which purchases training on behalf of Apprentices in the UK's quasi-market in training, buy services on behalf of many trainees and are therefore likely to interact *repeatedly* with providers. Such repeat purchasers will be better placed than private individuals making a one-off purchase both to observe the quality of services provided and also to channel business away from those providers who, *ex post*, are observed to supply services of low quality. The feedback effect on demand requires the presence of a sufficient number of alternative providers for buyers to be able to act effectively on the information, an issue on which evidence is provided in section 4, below. When such conditions are met, there are *prima facie* grounds for believing that providers have an incentive to cultivate a reputation for providing good quality services.

And it is in enhancing this reputational mechanism for encouraging the provision of high quality services that, according to economic theory, the main role of inspection is to be found. The dissemination of the inspection findings is intended to improve buyers' information about the quality of services offered by particular suppliers, thereby strengthening the feedback between quality and demand required to sustain high service quality. 'Through inspection, learners receive information which enables them to make informed choices about learning programmes' (ALI 2002a: 4). More specifically, given the prominence officially accorded to completion rates as an indicator of (substantive) training quality (e.g. LSC 2006: 1), economic theory implies that inspectors should provide information that enables purchasers to identify and channel business to providers with high completion rates, thereby giving providers an incentive to raise quality (so defined). Thus, the regime through which the government has attempted to regulate service quality is a hybrid form of control that combines competition for contracts with the information (supposedly) provided by ALI inspections (Hood *et al*. 1999: 13-17, 45, 139; James 2000: 332).

2.2 THE POLITICS OF REASSURANCE

A less optimistic account of the contribution of inspection to service quality focuses on its socio-political functions. For Power (1999), the principal function of external inspection is to assuage public anxiety, not to provide valid evidence concerning training quality.

Like the economic analysis of inspection, the socio-political approach also begins with informational issues, in particular the possibility that asymmetric information about quality may handicap inspectors as well as buyers. Inspectors certainly enjoy privileged access to the provider's facilities, and inspection teams can contain experts on the relevant service. But inspectors are by definition outsiders to the targeted provider, and their ability to observe service quality may be limited by a number of factors: the requirement to announce their visits well in advance, which gives the supplier scope to implement cosmetic measures designed artificially to inflate inspectors' estimates of the quality of training on offer; the limited duration of their visit; their dependence on the

provider for information; and the scope for suppliers to conceal evidence of any failings (Kitchener *et al*. 1999; Sinkinson 2005).

In such circumstances, Power argues, inspectors will concentrate on those quality-related attributes of provision that they *can* readily observe, in particular the extent to which providers have installed and documented the existence of a quality assurance system (that is, a set of managerial procedures designed to regulate the quality of the services they provide).[1] Inspection therefore emphasises the procedural correlates of training quality at the expense of the substantive ones. The complex operations of a wide variety of organisations – companies, schools, hospitals, police forces, and training providers –become 'auditable', simply because being auditable is now defined, not in terms of actual (first-order) service quality, but rather in terms of the generic (second-order) managerial systems that are supposed to facilitate the monitoring and control of service quality. As Power has put it, 'Audits become possible in complex environments by abstracting from that complexity and by operating on a systems surface which in some cases has been designed with auditability in mind' (1999: 88).

The provision of documentary evidence is crucial in this regard because the paperwork in question testifies to the existence of the requisite internal controls and, in principle at least, records the organisation's compliance with them, thereby enabling the organisation to be held to account during inspections. On this view, therefore, organisations are made inspectable and auditable via an auditee-provided trail of documentary evidence, which is used to verify the existence and operation of the requisite management control system. In Power's words, 'audits work because organisations have literally been made auditable; audit demands the environment, in the form of systems, and performance measures, which makes a certain style of verification possible' (1999: 91; also see Power 2004: 770-71, 779).

This emphasis on procedural matters is reinforced by two further considerations. The first is the greater scope under procedurally-oriented inspection for government to pass the costs of regulation from the public purse to the organisations being regulated, who can be judged partly on their development of the capacity to assess themselves

---

[1] For example, it is easier to observe the presence of individual training plans for all trainees than the extent to which individuals' training needs have actually been met.

(Ayres and Braithwaite 1992: 103, 118; Parker 2003: 136). The second is the contemporary quest by the public sector, confronted with the difficulty of demonstrating performance in the face of intangible and contested notions of service quality, to bolster its legitimacy by imitating 'leading edge' audit techniques drawn from the private sector.

The upshot of inspection's emphasis on procedural issues is that the substantive dimensions of service quality, such as trainee outcomes, may become *decoupled* from the procedural dimensions of quality, including providers' quality assurance systems, in the sense that the former is to a great extent determined independently of the latter (Meyer and Rowan 1977; Power 1999: 95-96, 128ff; Parker 2003: 145-46). In extreme cases, inspection focuses entirely on the organisation's internal control mechanisms, such as the presence of a self-assessment report and associated quality assurance procedures, to the exclusion of substantive outcomes. The possibility therefore arises that inspection 'provide[s] assurance that the system works well even when the substantive performance is poor' and that 'being fit for auditing … says little about fitness for any other purpose' (ibid.: 60, 145). As Braithwaite (1984: 139) pithily puts it, public inspectors 'ensure the quality of your records, not the quality of your deeds'. Consequently, inspection may function less as a promoter of service quality than as a diversionary activity that has little or nothing to do with true quality, but which is central to the contemporary growth of 'an industry of empty comfort certificates' (Power 1999: 123). In the case of Apprentice training, for example, if inspectors focus on the procedural rather than the substantive dimensions of training quality, and if the former are largely decoupled from the latter, then inspection results will do little to provide purchasers with the information required to direct their custom towards those providers with high completion rates, as the government's emphasis on trainee outcomes would require.

For Power, then, inspection is best thought of, not as a device for improving operational efficiency, but rather as a means of reassuring the polity that all is well in the public services irrespective of the latter's substantive quality. Support for this claim can be found in previous research into inspection, which has frequently found evidence of proceduralist bias. The British schools inspectorate, OfSTED, has been found to emphasise procedural over substantive attributes (Thomas 1996; Webb *et* al. 1998; Case *et al*. 2000; Chapman 2002). Indeed, a recent major report on quality-adjusted measures

of public sector output considered the use of OfSTED inspection grades as a proxy for educational quality, but ultimately rejected doing so, at least in part because, '[s]chools prepare for OfSTED inspection and there may be an artificial climate' (Atkinson 2005: 134). Day and Klein (1990) found a similar tendency in inspections of social care services by the Audit Commission and the Social Services Inspectorate. Other examples of proceduralist bias include: the Investors in People (IiP) programme, which is intended to raise quality in employers' skills management and training practices but which is found to serve as little more than a 'plaque on the wall' and to do little to improve workers' skills (Watson and Watson 1999; Hoque, Taylor and Bell, 2005); the inspection of local authority children's homes, where the focus of inspection reports was limited mostly to 'easily quantifiable, measurable aspects of life in children's homes', such as the temperature of fridges and bath water, rather than 'less easily quantifiable issues such as the quality of education provision available to residents' (Kitchener *et al*. 1999: 345); and the former star rating system for NHS hospital trusts, in which no relationship has been found between patient outcomes and number of stars awarded, consistent with the marginal role played by outcomes relative to process indicators among performance measures (Rowan *et al*. 2004).

Those who design and operate inspection systems tend to defend the importance attached to procedural attributes by arguing that management systems are more robust and resilient than the individuals who administer them, and that, once installed and operational, they can ensure high quality whoever runs the training function. Appropriate managerial systems and procedures are implicitly viewed as both necessary and sufficient for high service quality: necessary, because training quality cannot be assured without suitable mechanisms to identify and correct problems when they arise; and sufficient, because once effective systems of quality assurance and self-assessment have been put in place, individuals can come and go without doing lasting damage to training quality (Weiner 2002). The most conceded to scepticism concerning the role of inspection in the training market is that procedural improvement may take time to show up in substantive improvement (ALI 2006: 6). This line of interpretation is discussed further below.

In sum, although the two interpretations of inspection overlap, they contrast sharply. In the former view, inspection is the oil that lubricates the quasi-market; in the latter one, the smoke that obscures its deficiencies.

3. CONTEXT

In order to study these issues, the key attributes of the UK training quasi-market during the period studied here, 2001-05, must first be outlined.[2]

3.1 ACTORS AND PROGRAMMES

Publicly funded work-based training for young people has been organised through what is now the Apprenticeship programme, and until 2004 was named Modern Apprenticeship (MA). The programme contained two streams: Foundation Modern Apprenticeship (FMA), comprising programmes pitched at Level 2 in the vocational part of the National Qualifications framework; and Advanced Modern Apprenticeship (AMA), covering Level 3 programmes. These Apprenticeship programmes have since 2002 contained three components: 'competence', which denotes practical skills, as assessed by a National Vocational Qualification (NVQ); applied literacy and numeracy, as assessed by Key Skills tests; and underpinning theoretical knowledge, as represented by a Technical Certificate. The 14 Areas of Learning (i.e., subject or occupational areas) recognised under Apprenticeships include not only those associated with sectors such as engineering and construction, where there is a history of apprentice training, but also those located primarily in sectors that had previously made little or no use of apprenticeship (e.g., retailing, care services) (ALI 2002a: 14-15).

The Apprenticeship programme is administered by the Learning and Skills Council (LSC), a non-departmental public body directed by and accountable to the Department for Education and Skills. The LSC provides grants, which in 2004 varied between £1,000 and £15,000 according to trainee age and Area and Level of Learning,

---

[2] Additional details are provided in Ryan, Gospel and Lewis (2007).

for the provision of work-based training to young people. The LSC allocated training contracts and subsidies among competing potential providers, whose ranks include not only the employers and further education colleges who traditionally provided apprenticeships but also specialist training providers (both for-profit and non-profit).[3] While every trainee's programme must involve an employer, and while the LSC encourages employers to grant their Apprentices employee status, the employer's contribution can be limited to providing on-the-job training and work experience, in which case the employer acts simply as a sub-contractor to a specialist provider, who typically takes responsibility both for delivering the off-the-job training and also for assessing the Apprentices' learning.

The government's response to evidence of deficient quality in the Apprenticeship programme, outlined in section 1 above, has centred on the external inspection of training providers. It is to the nature of that inspection regime that we now turn.

3.2 INSPECTION METHODS

The body charged with inspecting all organisations holding an LSC contract for the supply of work-based training during the period investigated by this research project was the Adult Learning Inspectorate (ALI), a non-departmental government body set up alongside, but independent of, the LSC in 2000. ALI was to inform the government about 'the quality of education and training' and to 'help bring about improvement by identifying strengths and weaknesses and highlighting good and poor practice' (2002a: 4). It was to do so by providing information to participants in the training market.[4] Following Hood *et al*. (1999: 44-46), we shall outline the ALI's approach to inspection by reference to three key elements of systems of regulatory control, namely: the standards by reference to which providers are judged; the means by which information is gathered;

---

[3] In 2002-03, further education colleges accounted for 50 per cent of learners in the LSC-funded vocational preparation of relevance to this research, a category that we approximate by 'all trainees in 'work-based learning' plus part-time 16-24 year old college students'. The shares of other providers were: for-profit training companies, 24 per cent; employers training their own staff, 14 per cent; non-profit training providers, 7 per cent; employer groups, 5 per cent (Sources: LSC 2004, Tables 2, 6; ALI inspection database, below).
[4] In April 2007, ALI was merged into the schools inspectorate (OfSTED), as part of the government's efforts to reduce the regulatory burden.

and the ways in which providers whose performance is deemed to be unsatisfactory are encouraged to modify their behaviour.

3.2.1 Standards and assessment criteria

ALI began operating in 2001 and inspected providers of work-based learning on a four yearly cycle. Its inspections were to be guided by the Common Inspection Framework (CIF), a vehicle for the standardisation of assessment criteria across providers of post-16 learning whose central principle was that inspections should 'focus primarily on the experiences and expectations of individual learners' (ALI-OfSTED 2001: 3). The CIF stipulated, explicitly and in advance, seven aspects of provision that inspectors were to take into account in formulating their judgments about a provider's merits. The seven criteria, the application of which was non-negotiable so far as providers were concerned, were organised into three broad categories (ALI-OfSTED 2001: 6-13; ALI 2002a: 5). The first, headed 'Achievement and Standards', contained as its *sole* criterion the extent of trainees' achievements – in particular, the completion of their training programmes. The second category, labelled 'The Quality of Education and Training', encompassed five criteria: the effectiveness of teaching and learning; programme planning and resources; the effectiveness of the assessment of learning; responsiveness to the needs of learners; and guidance and support for learners. The third category, 'Leadership and Management', centred on the effectiveness of the provider's management in raising achievement and supporting all learners, and included measures designed to promote Equal Opportunities.

3.2.2 Detection: Audit and Inspection

ALI's procedures centred on self-assessment: 'The primary responsibility for improving the quality of training lies with the provider … it is vital that providers regularly evaluate all aspects of their provision and seek to improve its quality continuously' (ALI 2002a: 6). To that end, all providers were required to submit an annual self-assessment report to

their local LSC, in which they explicitly applied the CIF's criteria. When a provider was selected for inspection, its most recent report was to be passed on to ALI, both to assist the latter in planning their visit and also so that the rigour of the self-assessment process and the thoroughness of the report could be evaluated. In emphasising self-assessment by providers, ALI's approach to the regulation of training quality resembles the quality assurance systems developed by the British Standards Institute, which involve statements of objectives, the monitoring of performance against those objectives, the feedback of the results via written reports, and corrective action as required (ALI 2002: 6; Power 1999: 83-84).

The CIF stipulated that inspectors' judgements be based on sufficient evidence to justify their conclusions, drawing on observations of learning, interviews with learners and the provider's staff, and documentary evidence on training, assessment, and quality assurance. Providers were normally given between three and six months' notice of inspection. Inspection visits typically covered four or five working days. Inspection teams, which tended to be formed anew for each inspection, comprised between two and ten inspectors, with a full-time Lead Inspector, and possibly one other full-time inspector. Associate Inspectors, who comprised the remainder, were part-timers, chosen mostly for expertise in the relevant Areas of Learning. As the inspection proceeded, the inspection team was expected to inform the provider about its preliminary judgements, in order to permit the provider to offer additional evidence or to challenge its views, and finally to discuss its proposed findings with the provider (ALI 2002a: 8-9).

Inspectors awarded grades in order to summarise their judgments about the following aspects of provision: (i) Leadership & Management; (ii) Quality Assurance; (iii) Equality of Opportunity; (iv) Areas of Learning (Occupation/Curriculum); and (v) Learning Sessions. The first four of these attributes were graded on a five-point Likert scale, ranging from 'very weak' to 'outstanding'. A separate, seven-point scale was used to grade Learning Sessions. Details of these scales are set out in the Appendix.

These criteria formed a hierarchical structure. The grades for Quality Assurance and Equal Opportunity were required to contribute to the grade for Leadership and Management. Similarly, grades for Learning Sessions were apparently expected to contribute to that for Areas of Learning (O/C), though both the criteria on which the

former grades were to be awarded and their expected contribution to the latter ones remain obscure.

In addition, most inspection reports included data on trainee flows and outcomes – namely, numbers entering, staying in ('retained'), and completing training - broken down by Area and Level of Learning and by year of entry. Finally, many reports provided qualitative evidence of other quality-related attributes, such as the presence of a specialist training facility.

In the light of their findings, the inspectors were required to make an overall judgement as to whether, viewed as a whole, the training offered by a provider was 'adequate' or 'not adequate'. This is the key outcome, which we analyse in detail below. The inspection team was to reach its verdict, not by any formal aggregation algorithm, but rather through discussions held during its visit. Inspectors were, however, bound by one crucial rule, stated at the front of all inspection reports, namely that 'provision will normally be deemed inadequate where (i) more than one third of published grades for Areas of Learning (Occupation/Curriculum) or (ii) Leadership & Management are judged to be less than "satisfactory" '. Ultimately, a detailed inspection report was drafted by the lead inspector, using a format common for all inspections, checked for factual accuracy by the provider, edited and moderated by ALI headquarters (in order to promote clear and standardised terminology), and then published on the ALI website (ALI 2002a: 10).[5]

3.2.3 The Consequences of Inspection Results and Ways of Modifying Behaviour

The behaviour of providers was subject to modification through a variety of mechanisms. The latter ranged in severity, depending on the performance of the provider under inspection, forming a gradated series of sanctions or 'enforcement pyramid' through which improvements in quality were to be secured (Ayres and Braithwaite 1992: 35-36). In every case, the inspection team's findings were reported to the relevant local LSC, so that they could inform the latter's decisions about which providers should be awarded

---

[5] One of the authors participated in 2004 in a week long inspection of a commercial provider of engineering training. The inspectors stressed that their grades and overall verdict depended in particular on both the number of learners affected by the issue in question and its impact on those learners' experiences.

contracts to provide training. In other words, as noted above, the ALI's activities were intended to generate improvements in performance by informing prospective purchasers of the quality of the training offered by providers, thereby increasing the competitive pressure on the latter to offer high quality services (ALI 2004: 8). Moreover, although league tables of providers were not constructed, the Chief Inspector's annual reports did include lists of outstanding and poorly performing providers, with the prospect of such public praise or opprobrium potentially providing an additional incentive to enhance training quality (cf., e.g., ALI 2003a: 7-9, ALI 2004: 14-15; Hood *et al*. 1999: 153-54).

Those providers who were deemed to have performed especially poorly were further encouraged to improve the quality of their training by the prospect of re-inspection. Providers judged inadequate overall were to be re-inspected, normally within one year and at most within two. The scope of re-inspection was extended in September 2002 to include (i) the re-inspection, in the relevant area of provision only, of providers who received a grade less than 'satisfactory' for any aspect of provision, and (ii) short monitoring visits, subsequent to successful re-inspections, to detect and correct any backsliding (ALI 2002a: 11; 2004: 16, 17).

Re-inspection was assigned exclusively to the Lead Inspector, whose task it was, during a sequence of visits, to guide the provider in remedying its deficiencies, along the lines agreed in a post-inspection action plan, and culminating in a re-inspection report, also published on the ALI website (ALI 2002a: 11). Re-inspection reports included grades for the relevant Areas of Learning (O/C), and for Leadership and Management, Quality Assurance and Equality of Opportunity, but none provided fresh evidence on Learning Sessions or trainee outcomes. Re-inspection initially covered the provider's entire operations in cases in which either Leadership and Management or more than one-third of Areas of Learning (O/C) were graded less than satisfactory, but only the relevant Areas of Learning when less than one-third of Areas were so graded. Re-inspection methods will be considered further in section 5 below.

3.3. INSPECTION OUTCOMES: OVERVIEW

ALI's early inspections suggested that low quality was widespread in work-based training programmes. Only 42 per cent of the providers inspected during 2001-02 were rated 'adequate'. In the same year only 24 per cent of Apprentices completed their training programmes. Both statistics improved continuously during the subsequent four years. By 2005-06 the adequacy rate had risen to 88 per cent, and the trainee completion rate to 53 per cent (Figure 1).

INSERT FIGURE 1 ABOUT HERE

4. DATA SOURCES AND ATTRIBUTES

Our data are taken from the 949 inspection reports that ALI completed and published during 2001-05 concerning all non-college prime contractors for LSC-funded work-based training for young people, including specialist training companies and charities, companies and public bodies who train their own staff, and employer organisations. (Further education colleges are excluded because the inspection of the work-based learning they organised involved different arrangements, namely joint ALI-OfSTED teams, and reporting formats.) We excluded in addition providers who did not offer at least one Modern Apprenticeship (MA) programme, at Advanced or Foundation level, in one or more of the five Areas of Learning considered here: construction, engineering, business administration, information and communications technology (ICT), and retailing. We also excluded reports that lacked complete information for at least one Area-cum-Level on the seven quality-related attributes graded by ALI or – typically – for trainee completion rates. The result is a set of 442 training providers. Further details are provided in the Appendix.

The training market in which these providers compete for business is potentially competitive. Taking 2002-05 as a whole, the largest provider accounted for only 5.5 per cent of all trainees (in the full set of 949 providers); the ten largest, all of whom operated across the country, for 28.2 per cent. Competition is undoubtedly limited by the confinement of most smaller providers to particular localities, and by specialisation in

particular Areas of Learning by many providers. In many localities and Areas of Learning, however, there should be considerable scope to switch contracts from low quality to other providers.

Returning to the 442 retained providers, their median inspection year was 2002-03. The earliest cohort of trainees for which usable data on outcomes were available started training in 1997-98 (in engineering, Level 3); the modal cohort, in 1998-99; and the last cohort, in 2003-04 (in retailing, Level 2). Trainee outcomes are calculated for an average of almost two (1.9) entry cohorts per provider, for between one and ten Area-Level categories.

Table 1 summarises the distributions of training providers and trainees in our data. A majority of providers offer programmes in business administration; large minorities do so in retailing and engineering. Most offer programmes at both Level 2 and Level 3. In both construction and engineering, one half or more specialise in that Area, and only five operate across all ten Areas and Levels. The data cover almost 100,000 trainees, 58 per cent of whom entered retailing or business administration programmes, and slightly more of whom entered Level 3 than Level 2 programmes.

INSERT TABLE 1 ABOUT HERE

Our selection of reports according to the availability of completion rates may affect the content of our data. First, it removes all providers who started operations only recently in the five Areas of Learning covered here, leaving only those with longer track records. Second, and of greater concern, it may disproportionately remove low quality providers. If the absence of data on trainee outcomes from an inspection report is associated with defective record keeping by the provider, and if lower quality providers less frequently keep records of trainee progress, then the distribution of training quality will be truncated from below. The threat is plausible: some discarded reports contain qualitative evidence suggestive of low quality. Casual inspection suggests, however, that any association between sample selection and provider quality is less than close, as other discarded reports contain evidence of high quality. Moreover, the incidence and time

pattern of Adequacy in our sample is similar to that in ALI inspections as a whole (Figure 1, above). Nevertheless, our results may be affected by uncontrolled selection effects.

The coverage and timing of the evidence differ by variable. In brief, the scope of the provider's operations covered varies from variable to variable: e.g., completion rates apply to the five chosen Areas of Learning only, whereas Leadership & Management grades apply to the entire operation, including other Areas of Learning and non-Apprenticeship programmes. Further details are provided in the Appendix.

In terms of timing, organisational attributes refer to the period of the inspection itself, completion rates to trainees who had typically entered four years previously, i.e., for the earlier entry cohorts, under the aegis of ALI's predecessor, the Training Standards Council (Table 2). The discrepancy in timing is potentially important: the Chief Inspector has suggested that organisational improvement raises completion rates, but only with a lag (ALI 2006: 6). Such a discrepancy is however unavoidable if training outcomes are to be included in measures of service quality, and even acceptable, in that these data on trainee retention and completion represent the most up-to-date information on outcomes that was available to the inspectors.

Overall grades and trainee outcomes are shown in Table 2. Trainee outcomes are led by the completion rate, which we have calculated from the raw data in inspection reports on trainee inflows and outflows, excluding trainees still in learning at the time of the inspection. The retention rate, the inverse of the drop-out rate, is defined as the share who remain in training for the standard duration or who completed before its end. As data on retention are available for only a subset of retained reports, we focus primarily on completion.[6]

INSERT TABLE 2 ABOUT HERE

The median grade for Areas of Learning (O/C) and Learning Sessions is 'satisfactory' or better in all five Areas. By contrast, average trainee outcomes appear less than satisfactory. The drop-out rate averaged 54 per cent: only 46 per cent of entering

---

[6] We do not analyse the share of trainees who completed only part of their programme (e.g., NVQ2 only on a Level 3 programme), as data on that are provided by only a minority of reports, and collected on a basis that appears to vary from report to report.

trainees remained on their programmes until the scheduled end or left before then because they had already completed. Only 37 per cent of entrants actually completed their programmes.

Grades and outcomes also differ in terms of their dispersion across Areas of Learning. Grades for Areas of Learning (O/C) and Learning Sessions varied little across the five Areas, while retention and completion rates were much higher in the traditional Areas (telecommunications, engineering, and construction) than in the new ones (business administration and retailing). Completion rates were around one-half for Level 3 programmes in construction, engineering, and ICT, but only around one-third in business administration (35 per cent) and less than one quarter in retailing (22 per cent).

5. RESULTS

This section considers the two questions posed in section 1, namely how closely associated with each other are the various aspects of training quality, and how important are substantive and procedural aspects in the inspectors' overall verdict on the quality of a provider's services? Evidence is taken from the findings of ALI's inspection reports, and also from its formal inspection criteria.

5.1. DIMENSIONS OF TRAINING QUALITY

The view of training quality as heterogeneous is consistent with the variety of attributes on which the inspectors report. Five summary attributes feature in all reports: for the provider as a whole, grades for Leadership & Management, Quality Assurance, and Equal Opportunity; and for each Area of Learning, grades for Areas of Learning (Occupational/Curriculum) and Learning Sessions. To these we add our calculations of retention and completion rates, by Area and Level of Learning and year. The reports omit quality-related criteria that would be considered both measurable and important in some countries, notably the extent to which trainees receive part-time education, general as well as vocational, and the qualifications held by trainers (Ryan and Unwin 2001).

Table 3 shows the pairwise associations (rank correlations) among the eight quality-related attributes on which ALI inspections provide information, including the overall Adequacy verdict. The view of quality as comprising distinct but overlapping dimensions is consistent with the finding that all correlations between the detailed attributes are significantly less than unity and greater than zero, respectively.

INSERT TABLE 3 ABOUT HERE

We group the criteria into procedural and substantive categories by assigning Leadership & Management, Quality Assurance, and Equality of Opportunity (EO) to the former, and Learning Sessions, retention rates and completion rates to the latter.[7] Areas of Learning (O/C) is assigned provisionally to the procedural category. This categorisation reflects the pattern in the correlations in Table 3: the closest associations (at least 0.67) are found *within* the two categories, that is between Leadership & Management, Quality Assurance, and Areas of Learning (O/C), on the one hand, and between completion and retention rates, on the other. By contrast, all pair-wise correlations between criteria across the two groups are at most moderate (0.30 to 0.51). The pattern is also consistent both with the CIF's stipulation that the grade awarded for Quality Assurance should contribute to the Leadership grade, and also with the intrinsic overlap between retention and completion (by definition, only trainees who are retained can be, or become, completers).

The fit between *a priori* classification and *ex posteriori* findings is not perfect. First, and perhaps unsurprisingly, Equal Opportunity (EO) is associated only moderately strongly with Leadership & Management (to which it is required to contribute) within the procedural category. Second, notwithstanding the substantive connotation of its label and its specification in such terms by the CIF, Areas of Learning (O/C) proves more closely associated with Leadership & Management and with Quality Assurance than with

---

[7] Equal Opportunity could be oriented either to procedural attributes (e.g., provision of a training session covering EO issues) or substantive ones (e.g., actual shares of minority and female trainees) or both. The Common Inspection Framework refers to EO only as a subsidiary aspect of Leadership and Management: 'how well EO is promoted and discrimination tackled', and suggests that the inspectors check that 'there are effective measures to eliminate oppressive behaviour, including all forms of harassment' (ALI-OfSTED 2001: 13). It does not suggest that actual outcomes (e.g. the shares of minorities and females among all trainees) are to be considered. We therefore treat EO as a primarily, if not entirely, procedural attribute.

retention, completion or Learning Sessions, a finding that favours a primarily procedural view of its status. Third, Learning Sessions, whose assignment to the substantive category might be considered plausible given that the inspectors report on their observations of actual learning sessions, shows low correlations (below 0.35) with all the other criteria, including trainee outcomes, suggesting heterogeneity amongst the potentially substantive dimensions of training quality.

More important for present purposes is the validity of retention and completion rates as proxies for substantive attributes of training quality. We note here that, although the data from which we calculate completion and retention rates are conveniently *reported* in ALI inspection reports, those data are *not* the result of calculations or judgements made by the inspectors. The latter's role extended merely to reporting inflows and outflows of trainees taken from the provider's administrative records. The inspectors play a part neither in deciding whether particular trainees pass the various components of the Apprenticeship programme, nor in recording in the first instance the number of Apprentices retained and/or completing. This is important because it suggests that completion and retention data are potentially not subject to the same procedural biases as inspection grades, and may therefore provide a valid indicator of substantive quality.

Of course, the degree to which retention and completion rates indicate substantive training quality cannot be established definitively from the data considered here. Doubts about the extent to which retention and completion rates are truly informative about substantive training quality arise for a number of reasons. For example, whether trainees drop out of or complete their programmes undoubtedly depends not only on the quality of training they receive but also on other factors (e.g. the availability of a job during or after training). Moreover, the standards by reference to which trainees' success in completing (the various components of) their Apprenticeships may well have varied over time. For example, the increase in completion rates in recent years may well have been aided, to an unknown extent, both by any increase in the scope afforded providers, given the dominance of internal assessment of the NVQ component of Apprenticeships, to increase completion rates without improving the quality of their training (Power 2004: 775), and also by the relaxation – to an extent that remains unknown – of the 'requirement' for a Technical Certificate in some Apprenticeship frameworks (Ryan *et al*. 2006). Our point,

though, is that while such factors imply that the completion rate cannot provide a perfect indicator of the substantive quality of training, they are unlikely to destroy its validity entirely, and thereby to prevent us from assessing with these data the hypothesised proceduralist bias of ALI inspections.

Certainly, completion rates (i.e. qualifications attained) feature regularly as evidence of quality in education and training services both in policy documents (e.g. Atkinson 2005: 128-33, 194; ONS 2006: 21-25; QIA 2007) and also in the academic literature (e.g. Rosenthal 2004; Simpson 2007). Moreover, moving on to practical policy-making, the liability of 'failing' schools to 'special measures' and closure has in recent years depended on dropout rates and examination performance – i.e., on pupil retention and completion. Furthermore, in the training context, the LSC's recent reworking of public subsidies across the Apprenticeship programme sought validity by using data from 'providers who had long experience of delivering high quality Apprenticeships *and who have high completion rates'* (emphasis inserted; LSC 2006: 1).

We therefore interpret retention and completion rates as measures of substantive training quality, and their low correlation with the other attributes in Table 3 as our initial evidence of a low association between the procedural and substantive dimensions of service quality. Moreover, by calculating retention and completion rates, and using them as indicators of substantive training quality, we have at least been able to remedy the neglect by inspection of the information it has itself collected on trainee outcomes.

5.2. MULTI-VARIATE ANALYSIS

The second set of issues concerns the weight inspection attaches to particular quality-related attributes. The inspectors are required to reach a verdict on the adequacy or otherwise of the provider's services as a whole. In doing so, how much importance do they accord to particular attributes, procedural and substantive?

A preliminary answer is provided by the bivariate associations in Table 3. While the overall verdict (Adequacy) is positively and significantly associated with all seven detailed attributes, it is more closely related to procedural attributes than to substantive ones. The association of Adequacy with Learning & Management, Quality Assurance,

26

and Areas of Learning (O/C) is much stronger (correlations between 0.70 and 0.89) than that with Learning Sessions, or with completion and retention rates (correlations between 0.29 and 0.37). The one exception to the greater importance of procedural criteria is the relatively low correlation of Adequacy with Equality of Opportunity (0.44). Given that Equal Opportunity has only low correlations with the other criteria, we conclude that it captures a relatively separate aspect of the provider's operations, as might be expected given that in principle it concerns social justice rather than training quality *per se*.

A clearer view of the importance of procedural relative to substantive criteria requires a multivariate analysis, in order to estimate the contribution of each attribute holding the others constant. Estimation problems arise here, given that dependent variable (Adequacy) is binary (0, 1) and that most of the independent variables are ordinal (grades). The appropriate method is semi-parametric probit regression, estimated by maximum likelihood and using for each of the graded variables a set of piece-wise indicator ('dummy') variables that captures the attainment of successive levels of the grade scales.[8] In view of the secular increase in Adequacy ratings and completion rates during the period in ALI inspections as a whole (Figure 1, above), the analysis includes time fixed effects, with 2001 as base year, against which subsequent years' effects are estimated.

Tables 4 and 5 report the results of the probit analysis of Adequacy. Two stages are involved. In the first, every attribute for which the attainment of a particular grade is associated with a particular adequacy score across all providers is removed, as are all of the providers receiving that grade for that attribute.[9] The second stage sees the multivariate regression of the remaining observations, using the remaining independent

---

[8] Probit regression, using maximum likelihood methods, handles the intrinsic non-normality of the distribution of the stochastic component of a binary dependent variable, which invalidates ordinary least squares regression. A piece-wise specification of a graded variable comprising N grades uses N-1 indicator variables to capture the marginal effect on Adequacy of an increase of a single grade (e.g., a five grade scale generates four indicator variables, the first of which is unity when the provider is given any grade greater than the lowest grade and zero otherwise, and the fourth of which is unity for the highest grade only) (Conover 1999: 344ff; Greene 2003: 120ff).
[9] When a particular value of one independent variable is unfailingly associated with a particular value of the outcome, the estimation procedure in principle results in an infinitely large value for both the relevant regression coefficient and its standard error, and, *en route* to that result, the non-convergence of the iteration procedure or, more prosaically, its premature termination by the statistical package. The difficulty is avoided statistically by removing the variable and observations in question, and restricting the regression to the remaining variables and observations.

variables. The removal of the independent variables and particular observations in the first stage of the analysis is necessary for the identification of the effect of the remaining variables.

INSERT TABLE 4 AROUND HERE

The first stage sees an extensive removal of both variables and observations (Table 4). Fully 361 of the 442 observations are removed as a result of an invariant relationship between Adequacy and the attainment (or non-attainment) of a particular grade on the attribute in question. Leadership & Management alone generates the removal of 290 observations: 199 of those removals reflect the non-attainment of grade 3 ('satisfactory'), which is unfailingly associated with an overall rating of 'inadequate'; the other 91 are associated with the attainment of either grade 4 ('good' or better) or grade 5 ('outstanding'), both of which are unfailingly associated with a rating of 'adequate'. Lesser contributions come from Learning Sessions (45 observations removed, primarily for attaining at least 'satisfactory', which invariably means a rating of 'adequate'); from Equality of Opportunity (23 observations, for the attainment of 'satisfactory' or better); and from Quality Assurance (3 observations, for the attainment of 'good' or better). The indicator variable for 2005 removes two further observations.

The pattern of removals reflects in part the ground rules laid down by ALI. The finding that not reaching 'satisfactory' on Leadership & Management invariably means an overall rating of 'inadequate' shows that the inspectors followed the relevant ALI rule, viz. that provision will normally be deemed inadequate when Leadership & Management is judged to be less than satisfactory. The same does not however apply to the other removals. For example, as the rules do not stipulate that Leadership grades of 'good' or 'outstanding' must lead to an overall rating of 'adequate', irrespective of other attributes, this finding must be attributed to the priorities adopted by the inspectors themselves.

The scale of removals reflects the informational limitations of a dataset in which the dependent variable and most of the independent variables are binary (i.e., zero, one). Informational limitations are also indicated by the automatic discarding of further

independent variables, marked 'D' in Table 4, each of which is perfectly collinear with one or more of the retained variables.

Extensive removals do not however mean that information is discarded: on the contrary, the first stage indicates the importance of particular variables, notably Leadership and Management, in that a Leadership & Management grade of less than 'satisfactory' invariably means a verdict of 'inadequate', while a Leadership & Management grade of 'good or better' invariably means a verdict of 'adequate', and that settles the matter nearly two-thirds of the time. More generally, the importance of the variables removed at the first stage is indicated by the number of observations they cause to be removed.

Nor of course do the removals imply any perfect correlation between the 'parent' variables themselves. Thus, while particular Leadership grades (less than 'satisfactory' and more than 'satisfactory') are perfectly associated with particular adequacy grades, the adequacy verdict does vary within the 'satisfactory' category, which leaves scope for the multivariate analysis of the subset of providers given that grade.

INSERT TABLE 5 AROUND HERE

The second stage of the analysis is the regression analysis of the remaining 79 observations, all of them providers whose Leadership was rated no more and no less than 'satisfactory' (Table 5). The attributes that prove significant are Areas of Learning (O/C) at grade 3 (i.e., at least 'satisfactory') and the completion rate, both in the expected direction.[10] The coefficients in Table 5 represent the effect on the probability of an 'adequate' verdict of a one unit increase in the relevant independent variable, evaluated at the latter's mean. Thus a one grade rise in Areas of Learning (O/C) starting at grade 2 (i.e., from 'unsatisfactory' to 'satisfactory') increases the probability of an 'adequate' verdict by fully 84 percentage points, while a one percentage point increase in the completion rate (from its mean value) raises the probability of a verdict of 'adequate' by

---

[10] The retention rate is discarded as it proves insignificant in all specifications that include the completion rate and its inclusion removes the 71 observations for which it is not available.

0.0003. The latter effect is small but significantly greater than zero. The coefficients on the first three variables have the wrong sign, but none is statistically significant.

Surprisingly, the time indicator variables are negative as well as significantly non-zero: the probability of attaining an 'adequate' verdict, for given inspection grades and a given trainee completion rate, fell significantly after 2001-02. The upward trend in overall inspection results, as shown in Figure 1 above, reflects therefore the presence of even stronger upward trends in inspection grades and completion rates.[11]

5.3 DISCUSSION

These results suggest that the probability that a provider is rated 'adequate' depends primarily on attaining a grade of at least 'satisfactory' for Leadership & Management, with secondary contributions from reaching at least the same verdict for Learning Sessions (grade 4) and Equality of Opportunity. In addition, attaining an overall grade for Areas of Learning (O/C) of 'satisfactory' or better helps providers whose adequacy is not immediately determined by the criterion/grade pairs in Table 4. The completion rate is also associated positively and significantly with Adequacy among that subset of providers.

The evidence provides further support for the socio-political hypothesis of a primacy of procedural over substantive criteria in inspection practice. Three aspects are salient in that regard. The first is the statistical predominance of Leadership & Management; if the provider fails to reach the grade of 'satisfactory' or is graded 'very good' or 'outstanding', that settles the Adequacy verdict in two-thirds of the cases.

Second, there is the limited contribution of trainee outcomes to the inspection verdict, even though the Common Inspection Framework places 'how well do learners achieve?' first among seven 'key questions', and among its three 'evaluation requirements'. Moreover, an evaluation of learners' 'success in achieving challenging

---

[11] Any non-randomness in selection potentially distorts the time pattern of the adequacy rate. For example, were providers whose quality is more suspect inspected sooner, the rate would rise over time even were the distribution of training quality to remain unchanged. ALI has suggested that its selection of providers for inspection was random, or at least close to random: 'each quarter we select a representative sample of providers' (ALI 2002c: 2).

targets, including qualifications and learning goals, and trends over time' is the first on its list of detailed issues for inspectors to consider. The CIF also calls on inspectors to 'consider, where applicable, the extent to which … results and retention rates compare well with local and national averages' and 'performance over time show[s] continuous improvement or the maintenance of very high standards' (ALI-OfSTED 2001: 6-7).

Things turned out differently in practice. A large minority of inspection reports provide no data on trainee outcomes. In some cases, the lacuna can be attributed to the impossibility of providing such information (e.g., because the provider had only recently begun operations). More commonly, however, the provider had been in business long enough for trainee outcomes to be measurable in principle, but the data remained unavailable. It is striking in those cases to see that the inspection reports do not explicitly infer any quality-related failing from the non-availability of data on outcomes – a failing that is, ironically, likely to be procedural, insofar as the absence of data on outcomes suggests a defective management system. Moreover, even in the majority of reports that do provide data on trainee outcomes, only absolute flows are reported: numbers entering, retained, completing and still in training, by Area and Level of Learning and year of entry. No report contains calculations of rates of retention and completion, despite the ease of doing so, let alone compares those rates to local or national averages – the task indicated by the CIF, which ALI undertook only in the aggregate, providing breakdowns of retention and completion by Area of Learning and region only (ALI 2003b).

Third, the association of Adequacy with the completion rate (as calculated here), while positive and statistically significant, is small. A provider whose completion rate rose by fully four standard deviations (around the mean), from zero to 82 per cent, other factors held constant, would experience an increase in the probability of being rated 'adequate' of only 2.5 percentage points. The marginal role of completion rates in inspection outcomes is underlined by the presence of several providers who are rated 'adequate' despite having a completion rate of less than one quarter – and, in the extreme case (Horizon Training Ltd.), of only 6 per cent.

These findings suggest that a major reordering of inspection criteria occurred when ALI moved from inspection *principles* to inspection *practice*. The priority given in the Inspection Framework to trainee attainments was repeated in the guidance given to

training providers (ALI 2002a: 5). However, inspection practice gave priority, both formal and actual, to procedural attributes in general, and to Leadership & Management in particular, even though that criterion was placed last among the CIF's seven 'key questions'.

A purely proceduralist view of inspection is not however tenable. Adequacy is significantly associated with potentially substantive attributes as well, in the shape of the completion rate, Learning Sessions, and, more ambiguously, Areas of Learning (O/C). Concerning the latter, ALI's operational rules state that Areas of Learning (O/C) should enjoy equal status with Leadership & Management in determining the overall verdict. Although the criterion does not prove consistently decisive at the first stage in the way that Leadership does, it does exert a strong influence among marginal providers.[12] But should it be treated as part of the procedural or the substantive category, or as straddling the two? The latter interpretation is favoured by two considerations. First, there is the criterion's close association with both Leadership & Management and Quality Assurance (Table 3). Second, there is its moderately close association with trainee retention and completion, particularly in engineering and construction.[13]

The substantive attributes of training appear therefore to influence the overall verdict not only directly, through the completion rate and the Learning Sessions grade (given the invariant relationship between a grade of 'satisfactory' and an overall verdict of 'adequate' in Table 4), but also indirectly, through their association with the Areas of Learning (O/C) grade.

Overall, then, the multivariate analysis of inspection data indicates that ALI inspection reports do contain *some* information about the substantive quality of training offered by providers. However, the amount of substantive information that they provide, let alone process into usable form – and, therefore, the extent to which those reports can remedy the problem of asymmetric information about substantive quality in the training market by informing purchasers about the actual quality of a provider's services - is

---

[12] The weaker performance of Areas of Learning (O/C) than of Leadership of Management in our results may partly reflect the fact that we measure it as the median grade across only ten Areas and Levels of MA learning, whereas the ALI criterion requires a 'satisfactory' grade in at least one-third of all of the Areas of Learning offered by the provider.
[13] The simple correlation between the Areas of Learning (O/C) grade and the completion rate is .61 and .53 in engineering and construction, respectively, but less than .50 in the other three Areas of Learning, and only 0.37 in retailing.

severely limited. In keeping with Power's 'decoupling' hypothesis, inspection results appear to be driven less by the substantive dimensions of training quality and more by procedural attributes. The rationale for inspection must be sought more in its socio-political than in its economic contributions.

5.4 RE-INSPECTION

Further evidence comes from re-inspection reports. Re-inspection was automatically visited on providers who were rated 'inadequate'. It involved the re-grading of Leadership & Management, Quality Assurance, Equality of Opportunity, and Areas of Learning (O/C). No new grades were given for Learning Sessions. Nor were any additional data collected on trainee outcomes, even though on average another year of data had become available, given a mean interval between inspection and re-inspection in our data of fifteen months. Re-inspection reports are available for 154 of the 207 providers who failed their first inspection during the period we investigate. The vast majority (87%) of this set of providers managed to raise their game sufficiently by re-inspection to be rated 'adequate'. The pattern of success rates in our data is similar to that in ALI re-inspections as a whole (Figure 1, above).

The re-grading of procedural attributes, along with the absence of fresh grades for Learning Sessions or updated evidence on trainee outcomes – any failings in which might be expected to require longer than fifteen months to turn around – provides additional evidence of the primacy of procedural attributes in inspection practice as a whole.

6. CONCLUSIONS

The content of the Adult Learning Inspectorate's reports on providers of publicly funded work-based training for young people is consistent with a view of training quality as a heterogeneous phenomenon with overlapping dimensions. The distinction between

procedural and substantive dimensions proves important. Procedurally oriented dimensions are taken to comprise grades given for Leadership & Management and for Quality Assurance, and also inspectors' remarks on, *inter alia*, details such as the preparation of individual learning plans. Substantively oriented dimensions are taken to comprise retention and completion rates, and grades for Learning Sessions and, in part, Areas of Learning (Occupation/Curriculum).

We find that the inspectors' overall verdict on a provider's adequacy gives priority to procedural over substantive dimensions of quality. This is apparent in: ALI's operational guidelines, which give pride of place to Leadership & Management; the closer statistical association of the Adequacy rating with the procedurally oriented than with the substantively oriented attributes; and re-inspection practice, which largely ignores substantive issues. However, the pattern is not black and white: holding procedural attributes statistically constant, the probability that a provider with a middling grade for Leadership and Management is rated 'adequate' overall is significantly associated with substantive attributes, in the shape of completion rates in particular.[14]

To the extent that inspection prioritises procedural over substantive attributes, does that constitute a distortion, or does it simply represent a valid set of inspection priorities? The official view implicitly treats management systems and records as both necessary and sufficient for desirable trainee outcomes. The issue cannot be settled by the evidence presented here, but any outright denial of bias lacks plausibility: partly because it ignores the possibility that compliance with procedurally-oriented quality assurance procedures can be formal but not real, using paperwork and procedures to gloss over unqualified staff, poor training, insufficient learning, and inadequate trainee outcomes; and also because it ignores the likelihood that inspection procedures themselves encourage the development of just such a situation in the first place, particularly among specialist training companies, whose principal motive is presumably to maximise the profits to be had from public programmes.

Overall, then, our evidence suggests that the contribution of inspection to informational feedback in the training market is, while positive, strictly limited. It is

---

[14] The priority of procedural over substantive attributes has a potentially important side effect, namely the under-rating of employers relative to other providers, given that completion rates are much higher for the former than the latter (Lewis and Ryan 2007).

positive, in that procedural attributes are indeed likely to constitute part of overall training quality, and substantive attributes do affect statistically the verdict on some providers' adequacy. But it is limited, in that the contribution of procedural attributes to quality is itself intrinsically limited, and in that the statistical influence of substantive attributes on the overall inspection verdict is itself modest. Inspection might well be able to contribute more to decision-making in the training market but, as currently practised, it does little to reduce informational asymmetries. The alternative interpretation, as papering over the cracks, and even as generating spurious reassurance, gains the stronger support from our evidence.

Finally, the bias of inspection toward procedural attributes makes it difficult to interpret the apparently impressive increase in the share of inspected providers that is rated 'adequate', from less than one-half in 2001 to nearly nine-tenths in 2005 (Figure 1, above). The increase may indeed constitute an 'extraordinary achievement for work-based learning providers' (ALI 2006: 6). Alternatively, it may reflect the greater ease of raising a low quality provider's game in procedural matters than in substantive ones, such as trainee completions – which, as acknowledged by ALI, have not risen commensurately and which still leave much to be desired: 'Apprenticeship success rates lower than 50 per cent in many areas of learning … are wholly unacceptable' (ALI 2005: 6, 7).

ALI's preferred explanation of the failure of completion rates to rise as strongly as procedural attributes is itself open to doubt. The causes are described as 'persistent problems in some of the processes associated with work-based learning – initial assessment, ongoing assessment of competence, and the planning and management of learning …' (ALI 2005: 23). But these are primarily *procedural* deficiencies, which, given ALI's inspection priorities, should have debarred the relevant providers from being deemed adequate in the first place. Indeed, were ALI's explanation correct, it would call into question the validity of its inspectors' grades themselves.

References

ALI. 2002a. *A Guide to Inspection*. Coventry: Adult Learning Inspectorate.

ALI. 2002b. *Annual Report of the Chief Inspector, 2001-2002*. Coventry: Adult Learning Inspectorate.

ALI. 2002c. *Notification: Being Selected for Inspection*. Coventry: Adult Learning Inspectorate.

ALI. 2003a. *Annual Report of the Chief Inspector, 2002-2003*. Coventry: Adult Learning Inspectorate.

ALI. 2003b. *Retention and Achievement Data from Work-Based Learning Inspections*. Coventry: ALI.

ALI. 2004. *Annual Report of the Chief Inspector, 2003-2004*. Coventry: Adult Learning Inspectorate.

ALI. 2005. *Annual Report of the Chief Inspector, 2004-2005*. Coventry: Adult Learning Inspectorate.

ALI. 2006. *The Final Annual Report of the Chief Inspector, 2005-2006*. Coventry: Adult Learning Inspectorate.

ALI-OfSTED. 2001. *The Common Inspection Framework for Inspecting Post-16 Education and Training*. Coventry: Adult Learning Inspectorate.

ALI-OfSTED. 2005. *The Common Inspection Framework: for Inspecting Education and Training*. Coventry: Adult Learning Inspectorate.

Atkinson, T. 2005. *Atkinson Review Final Report: Measurement of Government Output and Productivity for the National Accounts*. London: HMSO. Available online at http://www.statistics.gov.uk/about/data/methodology/specific/PublicSector/Atkinson/final_report.asp Accessed 21/05/2008.

Ayres, I. and J. Braithwaite. 1992. *Responsive Regulation: Transcending the Deregulation Debate*. New York: Oxford University Press.

Barker, K., M. Chalkley, J. Malcolmson and J. Montgomery. 1997. 'Contracting in the National Health Service: Legal and Economic Issues', in R. Flynn and G. Williams (eds.), *Contracting for Health: Quasi-Markets and the National Health Service*. Oxford: Oxford University Press.

Becker, G. (1993). *Human Capital*. Third edition. Chicago: University of Chicago Press.

Boyne, G., P. Day and R. Walker. 2002. 'The evaluation of public service inspection: a theoretical framework', *Urban Studies*, 39, 1197-212.

Braithwaite, J. 1984. *Corporate Crime in the Pharmaceutical Industry*. London: Routledge.

Case, P., S. Case and S. Catling. 2000. 'Please Show You're Working: A Critical Assessment of the Impact of OfSTED Inspection on Primary Teachers', *British Journal of Sociology of Education,* 21, 605-621

Chalkey, M. and J.M. Malcomson. 2000. 'Government Purchasing of Health Services', in A.J. Culyer and J.P. Newhouse (eds), *Handbook of Health Economics,* vol 1a. Elsevier: Amsterdam.

Chapman, C. 2002. 'Ofsted and School Improvement: Teachers' Perceptions of the Inspection Process in Schools Facing Challenging Circumstances', *School Leadership and Management*, 22, 257-72.

Conover, W.J. 1999. *Practical Nonparametric Statistics*. 3rd edition. New York: Wiley.

Cullingford, C. (ed). 1999. *An Inspector Calls: OfSTED and its Effect on School Standards*. London: Routledge and Kogan Page.

Day, P. and R. Klein. 1990. *Inspecting the Inspectorates: Services for the Elderly*. London: Joseph Rowntree Memorial Trust.

DfES. 2007. *Further Education and work-based learning for young people – learner outcomes in England 2005/06*, Statistical First Release, ILR/SFR13, April (http://www.dfes.gov.uk/rsgateway/DB/SFR/s000723/index.shtml).

Greene, W. H. 2003. *Econometric Analysis*. Upper Saddle River, NJ: Pearson Education.

Hood, C., O. James, G. Jones, C. Scott and T. Travers. 1999. *Regulation inside Government: Waste-Watchers, Quality Police, and Sleazebusters*. Oxford: Oxford University Press.

Hoque, K., S. Taylor and E. Bell. 2005. 'Investors in People: Market-led Voluntarism in Vocational Education and Training', *British Journal of Industrial Relations*, 43, 135-53.

Hughes, G., R. Mears and C. Winch. 1997. 'An Inspector Calls? Regulation and Accountability in Three Public Services', *Policy and Politics,* 25, 299-313.

James, O. 2000. 'Regulation inside Government: Public Interest Justifications and Regulatory Failures', *Public Administration*, 78, 327-43.

Kirkpatrick, I. and M. Martinez Lucio. 1995. 'Introduction', in I. Kirkpatrick and M. Martinez Lucio (eds), *The Politics of Quality in the Public Sector*. Routledge: London.

Kitchener, M., I Kirkpatrick and R. Whipp. 1999. 'Decoupling Managerial Audit: Evidence from the Local Authority Children's Homes Sector', *International Journal of Public Sector Management*, 12, 338-50.

Laffont, J.-J. and J. Tirole. 1993. *A Theory of Incentives in Procurement and Regulation*. Cambridge, MA: MIT Press.

Le Grand, J. and W. Bartlett (eds). 1993. *Quasi-Markets and Social Policy*. Macmillan, London.

Lewis, P.A. 2008. 'Quasi-markets', forthcoming in N.F.B. Allington and J.S.L. McCombie (eds), *The Cambridge Student Handbook in Economics*. Cambridge: Cambridge University Press.

Lewis, P.A. and P. Ryan, P. 2007. 'External Inspection and the Role of Employers in Publicly-Funded Vocational Training: Evidence from the UK', unpublished paper, Department of Management, King's College London.

LSC. 2001. *Quality Improvement: Standards Fund 2001/02*. Circular. Coventry: LSC.

LSC. 2003a. *Quality Improvement Annual Report 2002/2003*. Coventry: LSC.

LSC. 2003b. *Work-Based Learning Provision: Factors Affecting Quality and Performance. Evidence from the West Midlands*. Coventry: LSC.

LSC. 2003c. *Government Supported Further Education and Work Based Learning for Young People on 1 November 2002, Volumes*, Statistical First Release ILR/SFR01, 31 March. Coventry: Learning and Skills Council.

LSC. 2003d. *Requirements for Funding Work-Based Learning for Young People 2003/04*. Coventry: Learning and Skills Council.

LSC. 2004. *Further Education, Work Based Learning for Young People and Adult and Community Learning – Learner Numbers in England 2003/04*, Statistical First Release ILR/SFR05, 14 December. Coventry: Learning and Skills Council.

LSC. 2005. *Planning for Success – a Framework for Planning and Quality*. Coventry: Learning and Skills Council.

LSC 2006. *LSC News Release*, 23rd February.

Macleod, W.B. 2007. 'Reputations, Relationships and Contract Enforcement', *Journal of Economic Literature*, 45, 595-628.

Meyer, J.W. and B. Rowan. 1977. 'Institutionalised Organizations: Formal Structure as Myth and Ceremony', *American Journal of Sociology*, 83, 340-63.

Nelson, P. 1970. 'Information and Consumer Behaviour', *Journal of Political Economy*, 78, 729-54.

ONS. 2006. 'Public Service Productivity: Education.' *Economic Trends*, 626, January, 13-37. Available online at http://www.statistics.gov.uk/articles/economic_trends/ET626_CeMGA.pdf Accessed 23rd May 2008.

Øvretveit, J. 2005. 'Public Service Quality Improvement', in E. Ferlie, L.E. Lynn, Jr. and C. Pollitt (eds), *Oxford Handbook of Public Management*. Oxford: Oxford University Press.

Parker, C. 2003. *The Open Corporation: Effective Self-Regulation and Democracy*. Cambridge: Cambridge University Press.

Power, M. 1999. *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.

Power, M. 2003. 'Evaluating the Audit Explosion', *Law and Policy*, 25, 185-202

Power, M. 2004. 'Counting, Control and Calculation: Reflections on Measuring and Management', *Human Relations*, 57, 765-83.

Propper, C. 1993. 'Quasi-Markets, Contracts and Quality in Health and Social Care: The US Experience.' In Le Grand and Bartlett (ed., 1993).

QIA. 2007. *Pursuing Excellence – the National Improvement Strategy for the Further Education System*. Coventry: Quality Improvement Agency.

Rosenthal, L. 2004. 'Do School Inspections Improve School Quality? Ofsted Inspections and School Examination Results in the UK', *Economics of Education Review*, 23, 143-51.

Rowan, K, D. Harrison, A. Brady and N. Black. 2004. 'Hospitals' Star Ratings and Clinical Outcomes: Ecological Study', *British Medical Journal*, 328, 924-25. Available online at http://www.bmj.com/cgi/content/full/328/7445/924. Accessed 8/05/07.

Ryan, P. 1994. 'Training Quality and Trainee Exploitation.' In R. Layard, K. Mayhew and G. Owen (eds), *Britain's Training Deficit*. Aldershot: Avebury.

Ryan, P. 2003. 'Evaluating Vocationalism', *European Journal of Education*, 38, 147-62.

Ryan, P., H. Gospel, and P. Lewis. 2006. 'Educational and Contractual Attributes of the Apprenticeship Programmes of Large British Employers', *Journal of Vocational Education and Training*, 58, 359-83.

Ryan, P., H. Gospel, and P. Lewis. 2007. 'Large Employers and Apprenticeship Training in Britain', *British Journal of Industrial Relations*, 45, 127-53.

Ryan. P. and L. Unwin. 2001. 'Apprenticeship in the British "Training Market" ', *National Institute Economic Review*, 178, 99-114.

Simpson, H. 2007. 'Productivity in Public Services.' University of Bristol CMPO working paper series no. 07/164. Available online at: http://www.bris.ac.uk/Depts/CMPO/workingpapers/wp164.pdf Accessed 22/05/08.

Sinkinson, A. 2005. 'Aspects of Consistency and Inconsistency in Secondary Humanities PGCE Courses with 'Perfect' Ofsted Grades', *International Journal of Method and Research in Education*, 28, 33-54.

Thomas, G. 1996. 'The New Schools' Inspection System: Some Problems and Possible Solutions', *Educational Management and Administration*, 24, 55-69.

Watson, T and D. Watson. 1999. 'Human Resourcing in Practice: Managing Employment Issues in the University', *Journal of Management Studies*, 36, 483-504.

Webb, R., G. Vulliamy, K. Hakkinen and S. Hamalainen. 1998. 'External Inspection or School Self-Evaluation? A Comparative Analysis of Policy and Practice in Primary Schools in England and Finland', *British Educational Research Journal*, 24, 539-56.

Weiner, G. 2002. 'Auditing Failure: Moral Competence and School Effectiveness', *British Educational Research Journal*, 28, 789-803.

Appendix

Data are taken from inspection reports on individual training providers, as completed by ALI between April 2001 and May 2005 and published on its website by July 2005. The organisation inspected is in all cases the holder of the prime contract with the LSC.

Of the 949 reports, we retain the 442 that satisfy the following criteria. First, the provider offered at some point during the period at least one Modern Apprenticeship programme, at either Level (Advanced or Foundation), in at least one of the five Areas of Learning chosen here: Construction; Engineering, Technology and Manufacturing; Business Administration, Management and Professional; Information and Communications Technology, and Retailing, Customer Service and Transportation. Second, the provider trained at least nine entrants, summed across entry cohorts, Levels and (five) Areas of Learning. Third, the inspection report: used the standard five and seven point grade scales stipulated by ALI for use through spring 2005; reported numbers of trainees entering, completing and remaining in learning (at the time of the inspection), by year of entry; and did so for at least one cohort that started training sufficiently far in advance of the inspection for its members to have been able to complete their programmes in that Area/Level within the LSC's standard length of stay (i.e., duration of grant payments) for 16-19 entrants to Advanced Modern Apprenticeship in that category – i.e., typically within two years, and within three years for Advanced MAs in construction and engineering (LSC 2003d, Annex B, pp. 39, 41).

The reasons for excluding further, otherwise eligible providers were: no report could be located (15 discards); no training was offered in any of the five Areas (292); only non-MA training was offered in the five Areas (39); data on trainee outcomes were absent or inadequate (161).

The variables are defined and scaled as follows:

Attribute	Scale[a]	Definition/criterion/notes
Overall adequacy	0 Inadequate 1 Adequate	Adequacy requires that (i) Leadership and Management be graded at least 'satisfactory' and (ii) not more than one third of grades for Areas of Learning (Occupation/ Curriculum) be less than 'satisfactory'
Leadership & Management Quality Assurance Equality of Opportunity	1 Very weak 2 Unsatisfactory 3 Satisfactory 4 Good 5 Outstanding	Grades for Quality Assurance and Equality of Opportunity contribute to this grade
Areas of Learning (O/C)	″	Median grade across all Areas of Learning; ties broken according to higher number of trainees
Learning Sessions	1 Very poor 2 Poor 3 Unsatisfactory 4 Satisfactory 5 Good 6 Very good 7 Excellent	″
Retention rate	0-100%	*Number of entrants retained / Number of entrants*[b]
Completion rate	0-100%	*Number of entrants completing / (Number of entrants minus number of entrants still in learning)*[c]

Notes: authors' definitions in italics

a. ALI scales inverted so that larger numbers represent higher quality

b. 'Numbers retained': entrants who either remain in training for the standard length of stay or complete before the end of that period.

c. 'Numbers completing': entrants who finish all elements of their training frameworks, including, since the 2001-02 entry cohort, Key Skills certification at Level 2 in language and number skills and a Technical Certificate representing underpinning technical knowledge, in addition to the original NVQ3 certificate of competence

The scope of the assessment within the provider's overall operation differs from variable to variable. Completion and retention rates refer only to Modern Apprenticeship

programmes in the five chosen Areas of Learning. Grades for Areas of Learning (Occupation/Curriculum) and Learning Sessions also refer only to programmes in the same five Areas of Learning, but include any non-Apprenticeship training also provided therein. Grades for Leadership & Management, Quality Assurance, and Equality of Opportunity refer to *all* of the provider's work-based learning, including both other Areas of Learning in Modern Apprenticeship and also other non-Apprenticeship training programmes (e.g. 'NVQ only' and the New Deal for Young People). The effect of these discrepancies should be limited: Apprenticeship programmes in the five Areas on which we focus account for the great majority of training supplied by the organisations in our data. Trainees in those five Areas accounted nationally in November 2002 for 68.0 per cent of the trainees in all fourteen Areas (LSC 2003c: Table 2), and in our data at the time of inspection for 79.2 per cent of all trainees.

Table 1. Number of providers and entrants to training programmes by Area and Level of Learning

	Area of Learning	Level 2	Level 3	All (both)	Any (either)
Providers	Construction	50	49	31	68
	Engineering	123	161	95	189
	Business administration	249	256	219	285
	ICT	59	42	36	65
	Retailing	177	170	149	198
	All	7	5	5	7
	Any	353	403	309	442
Trainee entrants	Construction	3,404	13,930	17,334	
	Engineering	5,416	13,072	18,488	
	Business administration	14,804	12,516	27,320	
	ICT	4,987	1,326	6,313	
	Retailing	18,857	11,539	30,396	
	All	47,468	52,383	99,851	

Notes

'Trainee entrants': numbers starting training, aggregated across all entry cohorts used.
The earliest cohort is 1997-98; the last, 2003-04. An average of 1.9 cohorts is used across Area-Level categories.

Table 2. Median grades and trainee outcomes by Area and Level of Learning

Area of Learning	Median grade		Retention rate (%)[a]			Completion rate (%)[a]		
	Area of Learning (O/C) (scale 1-5)	Learning Sessions (scale 1-7)	Level 2	Level 3	Both	Level 2	Level 3	Both
Construction	3	5	50.9	56.6	55.6	45.1	50.6	49.5
Engineering	3	5	37.3	58.6	52.7	29.5	53.0	46.3
Business administration	3	4, 5[b]	44.5	44.1	44.3	35.4	35.2	35.4
ICT	3	5	69.7	61.5	68.0	57.6	50.3	56.0
Retailing	3	5	32.9	32.3	32.7	21.5	23.8	22.3
All	3	5	42.5	49.2	46.1	32.2	41.7	37.2

Source: as Table 1
Note: Scales are five (Area of Learning) or seven (Learning Sessions) points long (Table 1)
Definitions. The completion rate is defined as (C/(S-L)), the retention rate as (R/S), where
   S: number of entrants (starts) in the cohorts for which adequate data are available
   R: number of entrants retained (i.e., remained in training for the LSC's standard programme duration or completed before then)
   C: number of entrants who completed their training programme at any point before the inspection
   L: number of entrants still in learning at time of inspection
a. arithmetic mean, weighted by number of entrants to training
b. equal numbers of trainees in the two grades

Table 3. Rank correlations between inspection grades and outcomes

	Adeq-uacy	Leader-ship & Manage-ment	Quality Assurance	Equality of Opportun-ity	Areas of Learning (O/C)	Learning Sessions	Complet-ion rate	Retention rate
Adequacy	1.00							
Leadership & Management	.89	1.00						
Quality Assurance	.70	.82	1.00					
Equality of Opportunity	.44	.56	.42	1.00				
Area of Learning (O/C)[a]	.73	.78	.67	.37	1.00			
Learning Sessions[a]	.32	.34	.28	.21	.34	1.00		
Completion rate[a]	.37	.41	.35	.13	.51	.15	1.00	
Retention rate[a]	.29	.36	.33	.12[#]	.48	.05[#]	.80	1.00

Notes: definitions and scales are shown in Table A2
n = 442 for all pairings except those involving Learning Sessions and retention rate (n=325);
all coefficients significantly different from zero (p=.01) except for those marked # (significant at p=.05
a. Aggregated across all Areas and Levels of Learning in which the provider operates (see Appendix)

Table 4. Probit analysis of Adequacy, first stage: removal of observations and variables, by attribute and grade

Grade awarded	Leadership & Management	Quality Assurance	Equality of Opportunity	Area of Learning (O/C)[a]	Learning Sessions[a]	Year 2005-06
2	D	D	D	D	n.a.	
3	199 (-)	retained	23 (+)	retained	retained	
4	83 (+)	3 (+)	retained	retained	35 (+)	
5	8 (+)	D	D	D	retained	
6	n.a.	n.a.	n.a.	n.a.	9 (+)	
7	n.a.	n.a.	n.a.	n.a.	1 (+)	
Observations removed	290	3	23	0	45	2

Notes
Cells containing a number indicate the number of observations removed before the final iteration, because of perfect correlation between particular values of the independent variable and the dependent variable. The sign in parentheses indicates whether the criterion is, respectively, the attainment of that grade (positive) or the non-attainment of that grade or a higher one (negative).
D: variable dropped owing to perfect correlation (multi-collinearity) with a retained independent variable
Grade scales: see Appendix (omitted grade is 1 in all cases)
a. Five selected Areas of Learning

Table 5. Probit analysis of Adequacy rating, second stage: regression results for retained variables and observations

Retained variables		Estimated coefficient	Standard error	z
Quality Assurance	(grade 3)	-.0040	.0047	-0.97
Equality of Opportunity	(grade 4)	-.0049	.0074	-1.09
Areas of Learning (O/C)	(grade 4)	-.0478	.084	1.28
Areas of Learning (O/C)	(grade 3)	.842	.188	4.18*
Learning Sessions	(grade 5)	.010	.012	1.26
Completion rate (/100)		.031	.021	3.04*
Year 2002-03		-.542	.202	-5.56*
Year 2003-04		-.425	.208	-5.52*
Year 2004-05		-.702	.110	n.d.
N		79		
Pseudo R^2		.52		

Notes: * reject at p=.01 the hypothesis that the parameter is zero; n.d. signifies 'not defined'
The coefficients are incremental probabilities, evaluated at sample means, estimated by the dprobit routine of Stata 8 package, with Huber-White robust standard errors
Significance tests refer to the underlying probit estimates

Figure 1. Inspection results and Apprentices' outcomes, England, 2001-2005
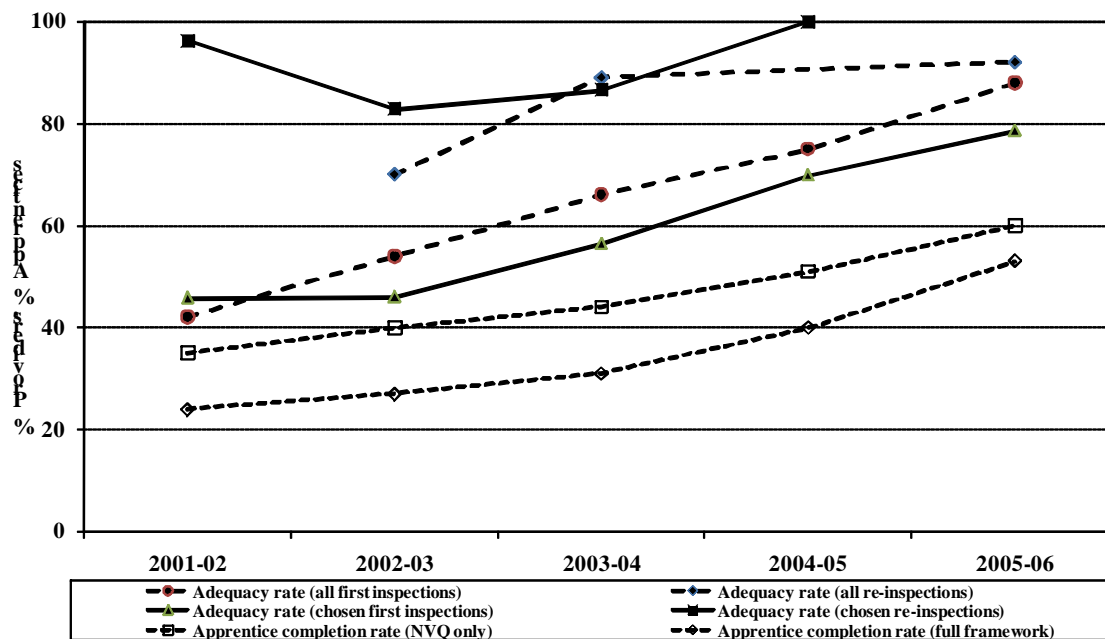
Source: ALI inspection reports (see Appendix); ALI (2002b, 2003a, 2004, 2005, 2006).
Note: Adequacy rate for all re-inspections in 2004-05 estimated by linear interpolation
'Chosen': 442 selected providers only