Assessment, Testing and Reporting 3-14 - Consultation on Partnership Commitments
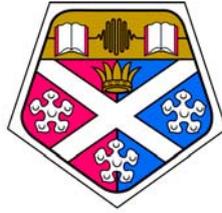
Final Report

March 2004

Dr Effie Maclellan

The Quality in Education Centre, University of Strathclyde,
Faculty of Education, Jordanhill Campus,
Crawfurd Building, 76 Southbrae Drive, Glasgow G13 1PP

Assessment, Testing and Reporting 3-14 – Consultation on Partnership Commitments

FINAL REPORT

March 2004

Dr Effie Maclellan
Quality in Education Centre
Faculty of Education, University of Strathclyde
76 Southbrae Drive, Glasgow G13 1PP

CONTENTS

List of tables

Executive summary

1   In recent years advances in our understandings of both learning and assessment challenge us all to try to find ways of improving assessment so that it can both enhance and measure pupil learning. The demand by policy makers, educators and the public that assessment should serve a variety of purposes finds expression in publications such as *A Review of Assessment in Pre-school and 5-14* (1999), *Improving Assessment in Scotland* (2000) and *Educating for Excellence* (2003). Just as for many countries in the developed world, the main purposes of effective assessment are seen to be:

- supporting learning, providing feedback to pupils, parents and other teachers, and identifying next steps in learning;
- providing information on which to monitor and evaluate provision and attainment at school, education authority and national levels.

2   Since 2001 the *Assessment is for Learning (AifL)* programme has been working to:

- provide a streamlined and coherent system of assessment; and
- ensure that parents, teachers and other professionals have the feedback they need on pupils' learning and development needs;

but to inform the AifL programme of assessment implications arising out of the more recent publication, *Educating for Excellence* (2003), the Scottish Executive sought views on the issues of whether to:

- replace reports with Annual Progress Plans
- replace the current provision of National Tests with a National Assessment Bank
- measure improvement in overall attainment through a Scottish Survey of Achievement rather than relying on the Annual 5-14 Survey.

3   Consultation questionnaires were returned by 1071 respondents, about three quarters of whom were professional school staff. The remainder comprised parents, local authority officers and persons from other educational organisations. In addition about 250 persons (again mostly professional school staff) engaged in seminar discussion on the issues.

4   The questionnaire consisted of 20 substantive questions requesting a closed response of YES or NO. In addition there were opportunities for respondents to make comments or suggest other options. As might be expected in a consultation of this size, the views expressed represented different, and sometimes even contradictory, points of view.

5   Views on the first issue, to replace reports with Annual Progress Plans, were almost evenly divided. While the current reports were seen to have disadvantages, they already had some of the alleged advantages of the proposed Annual Progress Plan. Furthermore, the proposal to change to an Annual Progress Plan was not well understood, largely because the implementation of the Annual Progress Plan was considered to be inter-related with implementation of Personal Learning Plans, the execution of which was still being piloted. In this context, 54% agreed with the proposal to develop Annual Progress Plans to a common framework with scope for local adaptation, while 46% disagreed. The alternative proposal, to produce a single national Annual Progress Plan/redesigned reports format that would be agreed and used by all schools, found support from 48% and resistance from 52% of the respondents.

6   The second issue of whether to replace the current provision of National Tests with a National Assessment Bank did not actually turn on a straightforward dichotomy. The overall conclusion was that respondents wanted assessment material to be available. This was confirmed in the statistic of 80% wanting the provision of National Test materials to continue. The issue for them was not whether the materials should be in the current or new form but whether the materials were available to teachers/schools as a resource or whether the availability of materials necessarily implied the constraints and stringencies of National Testing per se. Opposition to National Testing regimes was explicit.   This was further confirmed by 82% of the respondents who supported the introduction of a National Assessment Bank in the expressed assurance of the wording of the option that the materials in the bank could be used to confirm teachers' judgements. Views varied as to what should be included in/excluded from the new assessment bank, but almost all were premised on the belief that the professional decision-making of teachers should determine when and how assessment took place.

7   The third issue of whether to measure improvement in overall attainment through a Scottish Survey of Achievement rather than relying on the Annual 5-14 Survey triggered less response than did the other issues, partly because some of the questions were declared to be unclear, partly because the questions were not seen as pertinent and partly because the substance of respondents' views was declared to have been overtaken in the previous issue. Within this context, however, 68% of the respondents wished the Annual 5-14 Survey to end, an equivalent proportion favoured the introduction of the Scottish Survey of Achievement, but 79% were resistant to new subjects being included in the survey.

8   The views expressed on the three issues are global ones and mask differences between class/subject teachers and groups of teachers on the one hand and senior management teams, parents and other organisations on the other hand. The class or subject teachers and the groups of teachers – in other words, those whose professional practice is exclusively taken up with minute-to-minute interactions with, and decision-making on the part of, learners – were at odds with the senior management team (and others) who make overall policy and strategic decisions but who are not necessarily involved in the day-to-day minutiae of curriculum delivery. They had significantly different views on:

- the introduction of a common framework for reporting achievement
- the use of assessment banks, the inclusion of additional materials in the assessment bank and the issue of local moderation
- the introduction of a Scottish Survey of Achievement, the inclusion of new subjects in the survey and the uses to which data from such a survey might be put.

9   Specifically, and most notably in Aberdeenshire, Edinburgh City, North Ayrshire and South Lanarkshire:

- senior management were supportive of Annual Progress Plans which had scope for local adaptation, whilst teachers anticipated unnecessary difficulties in the proposal and would prefer to retain the existing annual report
- although for both senior management and teachers the preference was for the introduction of the National Assessment Bank, significantly fewer of the senior management had reservations about the proposal, whilst the teachers more readily pointed up its human and financial resource implications
- most senior management were in favour of local moderation, while most teachers were not, on the grounds either that moderation was difficult to effect or that it negated the notion of a national standard

- to each of the suggestions for extending the new assessment bank the teachers were unequivocally resistant, for fear that this would intensify what they saw to be an already bureaucratic assessment culture, whilst senior managers were more relaxed on the issue
- the proposal to introduce the new Scottish Survey of Achievement was positively endorsed by majorities of both teachers and senior management but, just as for the introduction of the National Assessment Bank, greater proportions of teachers were wary of the proposal, either because of resource and workload implications or because it was perceived to exacerbate what were currently seen as unfortunate and deleterious consequences of current monitoring practices.

10   The findings suggest that full development and execution of the partnership commitments set out by the Scottish Executive are some way off because of the need for all stakeholders to understand what is meant by an effective system of assessment and to understand both the costs and benefits of such a system. In order to reconcile points of difference and tension in what an effective system might look like, stakeholders could usefully build on shared understandings of fundamental constructs (such as the basis for making comparative judgements and the notion of standards in education). These, however, have to be agreed through argument and discussion that takes account of the up-to-date research evidence on learning and measurement.

11   Overall, the number of responses, together with the detail in some of the responses, points to considerable interest in, and concern about, educational assessment. The desire to make educational assessment as good as it can be was evident in all responses. This is a most heartening message to emanate from the consultation exercise and one which provides a sound platform on which to work towards resolving the tensions that currently exist.

1 Introduction

1.1   In the last few years the findings from Scottish Executive initiatives, documented in publications such as *A Review of Assessment in Pre-school and 5-14* (1999), *Improving Assessment in Scotland* (2000) and *Educating for Excellence* (2003), suggest that effective assessment must contribute to:

- supporting learning, providing feedback to pupils, parents and other teachers, and identifying next steps in learning;
- providing information on which to monitor and evaluate provision and attainment at school, education authority and national levels.

1.2   One response to these declared purposes of assessment has been to establish, since 2001, the *Assessment is for Learning (AifL)* programme which aims to:

- provide a streamlined and coherent system of assessment; and
- ensure that parents, teachers and other professionals have the feedback they need on pupils' learning and development needs.

1.3   While the *AifL* programme is designed to enable improvement in the quality and use of assessment, a thoroughly coherent and integrated programme of assessment requires that parents and the wider community understand both the context and content of assessment in meaningful ways. Furthermore, teachers and other education professionals must take account of the views of parents and the wider community on how assessment should proceed in order that assessment practices develop dynamically to support learning and enable pupils to participate fully in an increasingly complex world. Cognisant of this need to encourage communication between and amongst the many stakeholders in Education, the Scottish Executive stated in *A Partnership for a Better Scotland: Partnership Agreement* (2003) a commitment to:

- provide more time for learning by simplifying and reducing assessment, ending the current system of National Tests for 5-14 year olds;
- promote assessment methods that support learning and teaching;
- measure improvement in overall attainment through broad surveys rather than relying on the National Tests;
- improve the transitions between nursery and primary and primary and secondary education so that the system fits the needs of the children;
- promote improved assessment of individual schools' progress as a better measure than national 'league tables'; and,
- strengthen the link between parents and schools through improving the quality of information that parents receive about their children's progress, and replacing reports with Annual Progress Plans.

1.4   Although work to translate the Scottish Executive's commitment into reality has already begun under the auspices of the *AifL* programme, the Executive was nevertheless keen to consult with the wider constituency and so in the consultation paper, *Assessment, Testing and Reporting 3-14 – Consultation on Partnership Commitments,* invited views on its intentions to:

- replace reports with Annual Progress Plans
- replace the current provision of National Tests with a National Assessment Bank
- measure improvement in overall attainment through a Scottish Survey of Achievement rather than relying on the Annual 5-14 Survey.

The remainder of this report describes, and reflects on, the findings of the consultation exercise.

1.5   The consultation paper, which included a questionnaire, was issued in autumn, 2003. It was distributed in paper form to every local authority school in Scotland, to professional bodies in the wider education community and to every main local authority library. It was also available electronically through the Scottish Executive website at www.scotland.gov.uk/consultations and the Parentzone website at www.parentzonescotland.gov.uk. As a result, 1071 responses were received. The consultation exercise also included seminars in Aberdeen (21st November, 2003), Edinburgh (18th November) and Glasgow (27th November) for the purposes of informing people about the issues and encouraging responses.   The responses at the seminars were arrived at through group discussions: from Aberdeen there were responses from 8 groups; from Edinburgh 10 groups and from Glasgow 12 groups. Since each group consisted of 8-10 persons, the views of about 250 persons attending the seminars are represented.

1.6   Of the 1071 questionnaires it was possible to identify the category of respondent in 936 cases, of which 810 (about 75%) came from professional teachers. Of the 936 responses:

- More than half (490) were from promoted staff in school: that is head teachers, assistant/depute head teachers or principal teachers.
- Just under a third (275) came from class/subject teachers responding as individuals.
- In a small number of instances groups of teachers submitted a collective response. There were 45 such submissions.
- A small number of parents (31) submitted questionnaires. In most cases the parents were representing the views of parent-teacher organisations or School Boards.
- Central/Local Authority Staff (44 responses in total) comprised educational advisors, directors of education responding on behalf of their own authorities, local authority councils and policymakers. In all of these cases responses represented the views of the corporate body concerned.
- It was anticipated that some pupils might have engaged in the consultation but none completed questionnaires.
- The remaining category, 'Other', comprised 52 responses from academics in higher education (some of whom represented a particular interest, such as teacher education), and from officials representing organisations such as the Educational Institute of Scotland, the GTC, the Association of Directors of Education, Sense Scotland and Learning and Teaching Scotland.

1.7   There were responses from all 32 of the Scottish Local Education Authorities and there were 2 responses from England. The proportions of responses from the Scottish Local Authorities varied significantly ($p<0.001$) with 8 of the authorities each represented by less than 10 respondents. On the other hand, 15 of the authorities were represented by between 21 and 131 respondents. Within this general overview, however, there were some statistics which point to the composition of the sample being skewed. Almost half of the local authorities (15) had no representation from parents. There is no obvious explanation for this lack, which was evident in both the two least and the two most populous authorities. In seven of the authorities there were no returns from individual teachers and, partially

overlapping, a further seven authorities were not represented by the views of central/local authority staff. Additionally, there were (statistically) disproportionate numbers of returns from promoted teachers as distinct from class/subject teachers. While, of course, there is no compulsion on persons to participate in a consultative exercise, it is perhaps disappointing to note that, in the current political climate of initiatives to foster inclusion and full democratic participation, the analysis of the responses to the consultation exercise is from a sample that is less representative of the Scottish community than would be ideal.

1.8   The questionnaire consisted of 20 substantive questions on whether to:

- replace reports with Annual Progress Plans
- replace the current provision of National Tests with a National Assessment Bank
- measure improvement in overall attainment through a Scottish Survey of Achievement rather than relying on the Annual 5-14 Survey.

1.9   Each of the questions requested a closed response of YES or NO. In addition there were six opportunities for respondents to make comments or suggest other options. Respondents were selective in their attention to the questionnaire. No respondent completed the questionnaire in its entirety: some attended only to the closed questions; some discounted the closed questions but wrote free responses and some responded to both closed and open questions on the issues they perceived as pertinent.

1.10   Since the level of measurement of the data was nominal, non-parametric statistics were used to analyse the data. References in the report to response differences being significant or non-significant mean that statistical probabilities (using either chi-squares or Mann-Whitney Tests) were established.

1.11   As might be expected in a consultation of this size, the views expressed represented different, and sometimes even contradictory, points of view. The next two sections will try to capture the views expressed.

2   Findings from the questionnaire survey

This section is organised round the three issues that were the subject of the consultation, namely to:

- replace reports with Annual Progress Plans
- replace the current provision of National Tests with a National Assessment Bank
- measure improvement in overall attainment through a Scottish Survey of Achievement rather than relying on the Annual 5-14 Survey.

Italicised script captures text from the responses.

2.1   Replace reports with Annual Progress Plans

2.1.1   The first option to be considered here was whether to develop Annual Progress Plans to a common framework but with scope for local adaptation. Difference on this option was not significant (54% saying YES and 46% saying NO). Reasons for maintaining scope for local adaptation typically included the need for schools and local authorities to be able to reflect 'local issues' in their reports, to evidence the very genuine curricular differences between 3 and 14, and to capture, responsively, the additional support needs of pupils. Flexibility in reporting formats was also deemed to be a logical consequence to earlier, unsuccessful attempts to develop a common national framework, and to accommodate to the possibility that achievements **outside** (emphasis added) of school should be celebrated in any progress review.

2.1.2   The second option to be considered was whether to produce a single national Annual Progress Plan/redesigned reports format that would be agreed and used by all schools. Again, difference on this option was not significant (48% saying YES and 52% saying NO). Reasons for having a single national plan/format (often with the condition that *schools should be consulted on its format*) were to facilitate communication across the country, to be economical of both parents' and teachers' time in understanding pupils' needs, and to assist in monitoring pupil achievement.

2.1.3   Many more reasons were offered for considering the option to be unsuitable. Some reasons were on the theme that extant arrangements were at least preferable to the proposals, as in:

- *We are disappointed to note the Scottish Executive's implication that current provision "rarely" provides information about future needs. Our reports, at all stages, refer to development need and identify next steps.*
- *The view that pupils currently receive an end-of-year report betrays a considerable lack of awareness of what actually happens. In many schools reports are issued at various times throughout the session.*
- *Members of staff have now become familiar with the present pupil report form which has been improved annually.*
- *Our reports encourage choice and diversity rather than Annual Progress Plans being the same for every pupil.*

Other reasons referred to the workload implications of introducing Annual Progress Plans, as in:

- *Annual Progress Plans would be cumbersome and unwieldy and would add to the amount of paper in circulation.*

- *The extra time required to complete Annual Progress Plans is unlikely to yield sufficient benefit to pupils/parents to compensate for the time diverted from more immediate preparation for teaching.*
- *Any Annual Progress Plan will need to be easy to complete so that teacher workload is kept to a minimum.*
- *Yet another redesigned report adds to teacher workload.*

2.1.4   Concerns about the feasibility of Annual Progress Plans were offered, as in:

- *More details and/or an example of what an Annual Progress Plan looks like might help. As it is we are expected to comment in a vacuum.*
- *There are differences between primary and secondary schools which make the proposal for common reporting frameworks quite unworkable.*
- *It is implied that the Annual Progress Plan would be issued at the end of the year but this is too late.*
- *An Individual Education Plan at present takes considerable time to complete for a small number of pupils. It is difficult to imagine that Annual Progress Plans would be tenable for larger numbers of pupils.*

2.1.5   There were concerns that Annual Progress Plans were essentially cosmetic, as in:

- *As a paper exercise, the Annual Progress Plans will not improve the quality of the pupils' learning experience.*
- *The redesigned report will not change the current partnership arrangement.*
- *Annual Progress Plans would simply be a repeat of the information which is already with parents and pupils.*
- *It is unclear in what ways Annual Progress Plans are different from extended reports.*

2.1.6   A final set of reasons focused on the relationship between Annual Progress Plans and Personal Learning Plans. The general tenor of the comments acknowledged that Personal Learning Plans were probably a good idea but that their implementation in primary and secondary schools presented different workload and pedagogical issues which were, as yet, not altogether clear. These comments are perhaps best summarised in the remarks of one organisation:

The proposal to introduce Annual Progress Plans is set firmly in the context of the development of Personal Learning Plans. Given that Personal Learning Plans are still at the stage of being trialled, it would seem premature to move forward on this proposal until there is clear evidence not only of the feasibility but also of the usefulness of Personal Learning Plans.

2.1.7   Views as to the form of Annual Progress Plans were fairly evenly divided. Embedded within this, however, was a difference between teachers and senior management teams on the proposal that there be a common framework for Annual Progress Plans.   Senior management teams were significantly more enthusiastic than were teachers about the proposal, with teachers' cautious comments both anticipating that the proposal would invoke unnecessary difficulties and pointing out that the proposal had either resource or workload implications causing them to prefer the retention of the existing Annual Report.

2.1.8   Overall, the lack of a clear view on the question of replacing of reports with Annual Progress Plans is perhaps unhelpful to the Scottish Executive. However, given that many of the respondents stated that they did not know what Annual Progress Plans looked like or how they worked, and given that the consultation documentation contained no exemplar

material to illuminate points of change to/progress from extant reporting arrangements, the overall ambivalence towards the proposal is perhaps not surprising.

2.2   Replace the current provision of National Tests with a National Assessment Bank

2.2.1   The first option to be considered here was to end the provision of materials for National Testing. With all categories of respondent being of the view that the provision of National Test materials should not end, difference on this option was highly significant (*p*<0.001) with 80% saying NO and 20% saying YES.

2.2.2   Reasons for wishing the provision of National Test materials to end fall into one of two categories. One category referred to the ineffectual nature of the test materials themselves, as in:

- *They do not improve attainment.*
- *National tests do not give an accurate view of a child's progress. Many children panic and freeze as soon as they hear the word, test. It would be much better only to consider the written and oral work produced in class.*
- *National tests have not been very helpful in confirming teachers' assessments. There has been a lack of consistency between tests and the content is still often culturally biased.*

2.2.3   That so many of the test materials were of the language-bound, paper-and-pencil type was often added as a supplementary reason for wishing the provision of National Test materials to end. Such tests could not tease out language comprehension facility from conceptual attainment and were seen to be invalid and discriminatory, particularly for second language learners and pupils with special needs.

2.2.4   The other category of reasons for wishing the provision of National Test materials to end drew attention to the importance of the teacher's professional judgement, as in:

- *Teachers' professionalism is at the heart of McCrone. We should be trusted to produce assessments that are set at the appropriate level and the results should be accepted without having to use up further time on a National Test.*
- *Given the Scottish Executive's commitment to provide more time for learning, it seems unnecessary to expose pupils to even more assessment through National Tests if school-based evidence and satisfactory moderation processes confirm achievements.*
- *Teachers' judgements in all subjects other than English and Mathematics are accepted without the need to confirm by National Testing/Assessment. English and Maths should come into line with this.*

2.2.5   Reasons for wishing the provision of National Test materials to continue were more varied. Some distinguished, at least implicitly, between the materials as a resource and the materials as a process, as in:

- *Tests are a good diagnostic tool.*
- *National Test materials are useful in context and within an informal setting*.
- *There is nothing wrong with National Tests per se. It's what's been done with them that has been so damaging.*

2.2.6   Other reasons (largely from teachers) for maintaining National Test materials seemed to be more to do with feeling overwhelmed by the scale and rapidity of change in educational assessment and the need now to consolidate a 'comfort zone', as in:

- *We have only just become comfortable with this process. There is no need to change it.*
- *I am confident, as are the children, in the use of the present National Testing materials. There is no need to change.*
- *Everyone, including parents, has got to grips with it. Why change?*

2.2.7   A further clutch of reasons (almost exclusively from teachers) saw no point in ending the national provision of materials because of the belief that such action would cause local authorities to *provide their own alternatives*.

2.2.8   Finally, there were reasons to do with the need to have some benchmarks, as in:

- *The use of National Tests regularises, to some extent the information transferred from primary to secondary school.*
- *National Tests are not perfect but the standardisation they allow within schools/clusters is important.*
- *Since there are national guidelines in place there must also be National Tests to assess levels of attainment.*

2.2.9   The need for benchmarks to be able to describe where pupils were within the school, within similar demographies and within the national perspective was a value strongly felt by respondents from all categories. However, there was concern about the extent to which current National Tests could evidence this, since they had not been standardised in the fully technical sense. Elaborating on this issue, (a few) respondents were also concerned about the stereotypical demonisation that there had been of scientifically standardised tests, pleading the case that because such tests can provide high degrees of reliability and can be used effectively for diagnostic and formative purposes as well as for measurement of attainment, consideration should be given to the use of appropriate, rigorously standardised test materials within the context of formative assessment.

2.2.10 The second option was on the introduction and scope of the National Assessment Bank and comprised 7 questions.

Table 1: Proposal for National Assessment Bank

Proposal	YES	NO	Difference
1. Use assessment banks to confirm teachers' judgements	82%	18%	*p*<0.001
2. Support schools to introduce local moderation	58%	42%	*p*<0.001
3. Include materials for science	47%	53%	n/s
4. Include materials for social subjects	42%	58%	*p*<0.001
6. Include materials for modern languages	40%	60%	*p*<0.001
6. Include materials for practical assessment	39%	61%	*p*<0.001
7. Include materials for core skills	52%	48%	n/s

2.2.11 On the question of whether to introduce the National Assessment Bank for use in the same way as before to confirm teachers' judgements, all categories of respondent were

significantly in agreement with the proposal. This is not at all a surprising finding since it would resonate with views expressed above about the importance of teachers' professional decision making. A qualification to the support for the introduction of the National Assessment Bank was the repeated point that if materials were to be made available electronically, the increased financial burden on schools of accessing National Assessment Bank materials, together with the costs of purchasing and maintaining up-to-date ICT hardware and software, would have to be factored in at local education authority level.

2.2.12 A minority of respondents did not agree with the introduction of the National Assessment Bank. Essentially they questioned what the difference was between the provision of National Test materials and the introduction of a National Assessment Bank (although National Test materials were perceived to have a slight advantage since they were free!). It was pointed out that either form of provision was open to the same abuses of target-setting, teaching-to-the-test and successive re-testing to achieve mastery (but only of the test!). Rather, sceptics of the National Assessment Bank suggested that it would be preferable to develop the professionalism of teachers through programmes of staff development which included induction in moderation, engagement in professional debate and the reflection on exemplars of practice and pupil work.

2.2.13 Although for both senior management and for teachers the preference was for the introduction of the National Assessment Bank, significantly fewer of the senior management had reservations about the proposal whilst the teachers more readily pointed up its human and financial resource implications.

2.2.14 On the question of whether to support schools and authorities in the introduction of local moderation, significantly more (58%) were in agreement than were not (42%). Reasons for favouring the inception of local moderation included:

- *It would allow reference to a national standard while still acknowledging the professionalism of staff.*
- *If given greater prominence than at present, local moderation could enable teachers' judgements to be made with more confidence.*
- *Local moderation is useful to ensure consistency in relation to aspects of assessment which are not paper-based.*

2.2.15 Although there was support for local moderation, this was often qualified:

- *Authorities must address supply cover issues to allow moderation groups to meet.*
- *To develop authority-wide moderation would require significant management and funding since it is hard to see how any system - which makes claims for reliable results and the way the data can be used for comparison - can be anything other than costly if it is to be rigorous.*
- *Localised marking may be fraught with inconsistencies and would not thus allow standards to be established or compared.*

2.2.16 While the dominant view was in favour of local moderation, class/subject teachers and groups of teachers, significantly, did not share this view. In dissenting from the proposal they offered a number of reasons of which the following are typical:

- *Moderation would add complexity. National moderation didn't work so there is no hope for local moderation.*
- *Local moderation has already proved difficult in quality assurance exercises. Primary teachers are much more lenient in their interpretation of the criteria than secondary*

*teachers. If moderation is to be introduced it must include a **trained** outsider (original emphasis).*

- *The levels of quality assurance and organisation needed to effect local moderation make the whole process bureaucratic.*

2.2.17 The difference in view between senior management and teachers on the introduction of local moderation is best summarised by teachers' concerns that either moderation is difficult to effect or that it (allegedly) negates the notion of a national standard.

2.2.18 On the questions of whether to include, in the assessment bank, materials for science, social subjects, modern languages, practical assessments and core skills, there was resistance to this proposal for all items except core skills, a term that was frequently commented on as ambiguous. Resistance to the inclusion of materials for science was less clear-cut, perhaps because respondents noted science to be a national priority in the curriculum and therefore could justifiably merit the gravitas of a 'National Test'. Reasons for not extending the materials in the National Assessment Bank seemed to be either pedagogical:

- *We will be spending much if not all of our time assessing. When will we have time to teach?*
- *The extension to the Assessment Bank seems to point to an over-emphasis on summative assessment at the expense of formative assessment which has greatest impact on learning and teaching.*
- *There is evidence that current testing arrangements have narrowed the curriculum. The inclusion of more subject areas in the National Assessment might narrow it even further.*

or logistical:

- *Let's walk before we run! Give the new Assessment Bank time to settle in before we add to it.*
- *There is enough assessment already without adding to it. Listen to teachers for once!*
- *The extension of the assessment bank would not be the "little adjustment" claimed but an extra pressure; a pressure that would be too much for many.*

2.2.19 An afterthought often included in resisting the extension of the assessment bank was the perceived incongruity of such a proposal with the declared intention of reducing the assessment load. While the dominant view was that the assessment banks should not be extended to contain materials for a range of curricular areas, this view was most keenly held by the class/subject teachers and by the groups of teachers. All other constituent groups were of the view that extension of the assessment banks was beneficial because:

- *It signals that other curricular areas are as important as Maths and English Language.*
- *Practical and core skills deserve a special emphasis in assessment.*
- *It would allow teachers to assign 5-14 levels to pupils in subjects beyond English and Maths.*

2.2.20 Furthermore, such extension was seen as desirable, provided:

- *There were marking schemes and training provided.*

- *The materials were not used as National Tests but as a useful resource for teachers to dip into using their own judgement and discretion.*
- *Schools helped to contribute materials to the Bank thereby indicating that they were happy to use them.*

2.2.21 At a general level, there was appreciation that the National Assessment Bank would generate materials randomly, to avoid teaching-to-the-test. However, this was countered by concerns that the demographic diversity in Scotland might mean that some pupils could be disadvantaged through contextual and cultural issues. Another general comment that was made from time to time in respect of the extension of the Assessment Bank materials was that any decision was premature until a review of the curriculum had been completed.

2.2.20 The question of whether to replace the current provision of National Tests with a National Assessment Bank generated a confusing response. The majority of respondents were of the view that the current provision of test materials should not end but at the same time most respondents were also in favour of the introduction of a new National Assessment Bank. What seems to be the common denominator is the desire to have a supply of test materials available. Resistance to any provision of test materials was a minority view. However, the diverse views on the use to which the materials were to be put flagged up the inherent tension in trying to have an assessment system which serves both the support-for-learning purpose and the accountability purpose.

2.3   Measure improvement in overall attainment through a Scottish Survey of Achievement rather than relying on the Annual 5-14 Survey

2.3.1   Views on how to measure improvement were accessed through three options. The first option to be considered was whether to continue the current Annual 5-14 Survey. With all categories of respondent being of the view that the current Annual 5-14 Survey of Achievement should not continue, difference on this option was highly significant ($p<0.001$) with 68% saying NO and 32% saying YES. While, typically, reasons for either view were not given, the following were offered. Reasons to discontinue the Survey included:

- *The Annual Survey has resulted in the inappropriate use of assessment data both by politicians and the community at large.*
- *The publication of school-level and authority-level data puts pressure on teachers.*
- *As there is no moderation of 5-14 testing the results are meaningless.*

2.3.2   Respondents wishing to continue with the Survey believed:

- *It is appropriate to have publicly declared standards of national attainment.*
- *The survey is not intrusive and it gives information on the success/failure of current educational policy on a national basis.*

2.3.3   However, in wishing to continue with the Survey, respondents considered that:

- *National standards should be used only for formative purposes that are internal or local to the school.*
- *There is a need to improve the consistency of test administration and marking to enable the survey to reflect pupil performance accurately.*
- *Baseline assessment to characterise what learners bring into the educational system is also needed.*

2.3.4   The second option was on the introduction, in a variety of manifestations, of the new Scottish Survey of Achievement. It comprised 5 questions. In perusing the table below it should be noted that apart from question 1, where the response rate was much the same as the response rates for all previous items in the questionnaire (with a range of 80% to 90%), anything up to a third of the respondents did not attend to questions 2-5. This is partly explained, at least, by those respondents who explicitly made the point that they were insufficiently informed about extant monitoring arrangements to make any comment.

Table 2: Scottish Survey of Achievement

Proposal	YES	NO	Difference
1. Introduce Scottish Survey of Achievement	68%	32%	*p*<0.001
2. Extend the SSA to sample from S4	39%	61%	*p*<0.001
3. Link SSA data to census data for each sampled pupil	52%	48%	n/s
4. Link SSA data to schools' annual attainment data	49%	51%	n/s
5. Extend SSA samples to include special groups	46%	54%	n/s

2.3.5   In response to the first question, as can be seen from Table 2, the preferred option was in favour of a new survey being introduced, with the main justification being that it would be preferable to the current Annual 5-14 Survey now perceived as discredited. However, teachers (as distinct from senior management in school) were more evenly divided on the introduction of a Scottish Survey of Achievement, often positing that the new survey would be of no greater use than the system it was purporting to replace. While the need for national monitoring was conceded in some responses, it was suggested that an externally determined and administered instrument, with results collated on an anonymous basis, might be fairer and more workable.

2.3.6   The question of whether to extend the Scottish Survey of Achievement in various ways (questions 2-5 above) was seen as necessarily invoking further cost. This was seen to be an inappropriate use of limited resources which would be better spent on the improvement/enhancement of teaching and learning. It was also seen to have the possibility of generating data which could so easily be misread and misused, particularly since it was alleged that current arrangements for the accurate return of attainment data were not sufficiently refined. These views, resisting the range of uses to which data from the Scottish Survey of Achievement might be put, were held largely by teachers. The other categories of respondent seemed more open to ways of extending the survey. Notwithstanding the concerns of local authorities and senior management about the need to manage the introduction of these extensions in ways that did not make unreasonable demands of people, their belief that such extensions would yield additional information and that such information was ethically appropriate to derive, remained firm and unquestioned.

2.3.7   Within this overall context there was a significant preference not to include S4 pupils in the sample, largely because of the other assessment tasks that the S4 population typically faced. The question of whether to include groups of pupils with special characteristics did not yield a definitive response, perhaps because respondents noted that they were unclear as to what was meant by 'special characteristics'.

2.3.8   Reasons for the proposal, however, included:

- *This would enable effective monitoring of the inclusion agenda.*
- *The identification of pupils with special/individual needs would help explain why targets and attainment results might be 'lower than average'.*

while reasons against identifying pupils with special characteristics focused on data protection concerns and the methodologies for sampling.

2.3.9   The third option was on what the cycle of subjects should be in the new survey and comprised 4 questions. Again the response rate to these questions never exceeded 69% of the respondents.

Table 3: Cycle of subjects

Proposal	YES	NO	Difference
1. Include new subjects in the survey	21%	79%	$p<0.001$
2. Survey only English and Maths every year	25%	75%	$p<0.001$
3. Survey English or Maths and one other subject every year	28%	72%	$p<0.001$
4. Include embedded core skills in each survey	49%	51%	n/s

2.3.10 There was a clear preference by all categories of respondent that no new subjects should be included in the survey (indeed teachers, as distinct from other groups, were almost unanimous in this view) and that the cycle of Maths and English surveying should be no more frequent than obtains in the current Assessment of Achievement Programme. Such justification as was given was that English/Maths should retain their principal 'status' since they were basic to almost all other forms of assessment. Views on whether to assess core skills were much less clear-cut, with some responses suggesting that performance on core skills should influence what subjects are surveyed, while others emphasised the earlier comment that what was meant by core skills had not been defined (even although a definitional note in the consultation documentation adumbrated core skills to comprise communication, numeracy, problem-solving, information technology and working with others). A final comment, which is not inconsistent with earlier points, was that it was premature to consider what the cycle of subjects should be until a review of the curriculum was complete, given that the 5-14 Guidelines were seen to be a bit jaded, if not outdated.

2.3.11 As with the proposal to introduce a National Assessment Bank, the proposal to introduce the new Scottish Survey of Achievement was positively endorsed by majorities of both teachers and senior management but, just as for the introduction of the National Assessment Bank, greater proportions of teachers were wary of the proposal, either because of resource and workload implications or because it was perceived to exacerbate what were currently seen as unfortunate and deleterious consequences of current monitoring practices.

2.3.12 Overall, the question of whether to measure improvement in overall attainment through a Scottish Survey of Achievement rather than relying on the Annual 5-14 Survey was addressed by a smaller number of respondents than had attended to the previous two questions. While perceived relevance of the options within the question may have been a reason for either discounting the question altogether or offering little in the way of extended comment, it is also possible to discern from the responses that the formative purposes of assessment (with which class and subject teachers are primarily concerned) and the summative purposes of assessment (with which senior management and local authorities have necessarily got to be concerned in order to target resources appropriately) are not yet perceived to be compatibly managed through one single assessment system.

3   Views from the seminars

In addition to the results reported in Section 2, the consultation exercise included three seminars: one in each of Aberdeen, Edinburgh and Glasgow. Their purpose was to inform delegates both about the effects of assessment (on learning, learners and teachers) and the ways in which the proposed system of assessment in Scotland would support learning. The delegates then discussed (in groups of 8-10 persons) what they saw to be the issues surrounding the Scottish Executive's proposals to:

- replace reports with Annual Progress Plans
- replace the current provision of National Tests with a National Assessment Bank
- measure improvement in overall attainment through a Scottish Survey of Achievement rather than relying on the Annual 5-14 Survey.

In total there were responses from 30 groups. Perusal of the lists of delegates suggests that most represented schools, with only a sprinkling of representation from other interested persons. This section will try to represent the views of approximately 250 seminar delegates and will be concerned to reflect, as faithfully as possible, the views of the respondents. *Italicised script* captures text in the responses.

3.1   Replace reports with Annual Progress Plans

3.1.1   Discussion on this proposal included the perceived advantages and disadvantages of the current reports.

Advantages of reports

- *They are now well established and so understood by the profession.*
- *They have been customised to suit local and authority needs and to reflect more sophisticated understandings of assessment. For example, the Reports frequently draw attention to the pupil's strengths and point to how such strengths can be built upon. They are more than a summative account of achievement.*
- *They can be used as the basis for face-to-face discussion with parents.*
- *Parents seem to want an-end-of-year report that summarises achievement.*

Disadvantages of reports

- *The amount of time used by teachers in their completion is significant. For example it is not unreasonable to devote an hour and a half to one report. When this is multiplied by the number of pupils on whom one is reporting, a vast amount of time is being used, which may be of questionable effect.*
- *The information contained in the Report may be of limited use to parents, either because there is simply too much of it or because it is written in language that parents do not readily understand.*
- *The comments that are available in the data bank may not best reflect the pupil's needs/achievements. The availability of a comments data bank is not as helpful as some might believe.*
- *Annual reports issued at the end of the year do not help parents to work with their children during the year.*

3.1.2   While an exploration of the advantages and disadvantages of current reports might provide a baseline against which to examine the proposal for Annual Progress Plans, a discussion of its pros and cons did not really develop, largely because many of the delegates

felt unable to comment appropriately on Annual Progress Plans (APPs). A number remarked that since they hadn't seen an example of what an Annual Progress Plan looked like they couldn't comment. However, this comment may actually imply deeper confusions. The proposal to replace reports with Annual Progress Plans assumes that Personal Learning Plans are well developed in all schools. However, this is not the case. Indeed, Personal Learning Plans themselves, together with the management of Personal Learning Plans are still at the stage of being piloted in a small number of schools and local authorities. Furthermore, the evaluation of both of these projects is not yet complete. Because Personal Learning Plans are as yet not well understood, they themselves generated a number of concerns.

3.1.3   One type of concern was to do with the nature of Personal Learning Plans:

- *What are PLPs?*
- *What goes into a PLP?*
- *How much time will they need to prepare?*

3.1.4   Another type of concern was the anticipation of pedagogical developments:

- *The use of PLPs involves a culture change in how we work.*
- *It is a mistake to foist PLPs on pupils, teachers and parents before formative assessment is fully integrated into the curriculum.*
- *Any system for reporting should join up with the curriculum review.*

3.1.5   A third type of concern was the relationship between PLPs and other records:

- *If PLPs are well developed we might not actually need APPs.*
- *Will APPs just be summaries of the PLPs?*
- *PLPs could join up with Individual Education Plans (IEPs).*

3.1.6   A fourth type of concern was the resource implication:

- *It would take huge amounts of time to do a PLP and an APP with each pupil. Furthermore, in secondary school, the co-ordination of PLPs in all curricular areas into the APP would be impossible.*
- *Since most parents don't have time to come to school regularly, PLPs might need to be accessible to parents through secure web sites. This would need the investment and management of a huge IT component.*

3.1.7   To a large extent, the concerns about Personal Learning Plans hijacked the discussion of what form the annual report should take. Given these concerns and the declared interrelationship between PLPs and APPs, one conclusion is that perhaps the proposal was presented prematurely. However, in trying to remain on task the discussants distinguished, perhaps inadvertently, between two issues in reporting which might merit further consideration. One was the general issue of communicating with parents about learner progress, in which the point was made that for there to be effective communication, parents and teachers need to talk face-to-face on a regular basis. This might take place informally, as when parents come into school to collect their children, or it could take place through there being increased consultation time. The advantages of this sort of continual reporting/communication were seen to be superior to the time-consuming documentation of all achievements. The other issue, which is related, was that the production of an annual report (and preference was expressed for the term 'report' to be retained) should not be a catalogue of achievements (and failures) but, rather, a summary to point up the various ways

in which the pupil is indeed making progress. Notwithstanding the disadvantages of the current arrangements for annual reporting, there was, overall, no real benefit perceived in the proposed Annual Progress Plan.

3.2   Replace the current provision of National Tests with a National Assessment Bank

3.2.1   Discussion on this proposal was dominated by the question of the purpose of providing materials, either in the form of National Tests or in the form of a National Assessment Bank. If the purpose of providing assessment materials was to offer a pool of resources in a wide range of content and format, without any requirement either to use the materials or to account for how pupils had performed, then the overall view was that such resource material could be useful in informing teachers *of where their pupils are in relation to the 'national standard'.*

3.2.2   However, with hindsight, the discussants saw the prime purpose of using National Test materials as that of accountability:

- *National Tests are used as a management tool, to pace progress.*
- *There is too much pressure from HMIE to raise levels of attainment through performance on National Tests.*
- *We should use the information to develop pupil learning, not to get them through the levels.*

3.2.3   Furthermore, there was recognition that the original purpose of providing materials had been attenuated:

- *The use of National Tests has changed. They are no longer used for their original purpose of confirming teachers' judgements but to drive the target-setting agenda, with the attendant misuse of tests.*
- *There has been an increasing lack of confidence at the primary/secondary interface on the reliability of levels because tests are **not** (original emphasis) being used as confirmation of teachers' judgements but as a judgement on the success of the teacher/school with attendant pressures to be seen to 'perform'.*

3.2.4   This change in the use of the materials was attributed to the dominance of accountability rather than quality of learning being the 'driver' for what happens in schools:

- *SEED must have the courage to desist from the practice of comparisons – between schools and between local authorities. This engenders an educational culture of fear which is not conducive to the exercising of professional judgements.*
- *The evidence from such assessments should not be use for setting targets.*

3.2.5   Because of what the discussants saw as the depressing dominance of the accountability agenda, they were fearful that any attempts to ameliorate present problems would fail.   Rather than suggest how assessment materials might be bettered, the discussants were of the view that proposed changes would inevitably lead to further pressures on pupils (to be increasingly tested) and on teachers (to be exclusively focused on test performance).

3.2.6   The proposal to introduce a National Assessment Bank was seen to have some advantages:

- *The materials can serve as a benchmarking aid and therefore help local moderation.*
- *Pupil performance in using the materials can be used as evidence for parents.*
- *The on-line delivery will encourage testing when ready.*

3.2.7   However, the proposal was not received uncritically:

- *While there are acknowledged problems with National Tests, it is not clear how the proposal overcomes problems such as holding children back and teaching-to-the-test.*
- *Materials could be useful provided that teachers can use their professional judgement in selecting from the bank.*
- *It is unclear how the existence of a National Bank will promote formative assessment (as described in the consultation papers).*

3.2.8   Local moderation was perceived to be a logical corollary of trusting teachers' professional judgements in their assessment of pupils. However, it was also recognised that:

- *The idea may be interpreted differently in primary and secondary school.*
- *Even if the idea was variously understood, there had to be some degree of consistency across different authorities.*
- *There will be a need for moderation both between and within schools.*

and so for local moderation to be fully realised, the resource implications of its introduction and implementation would have to be fully confronted.

3.2.9   At a superficial level, the question of whether to replace the current provision of National Tests with a National Assessment Bank did not really matter to the discussants. They saw the question as touching on a much more significant and fundamental issue: that of assessment purpose. Overwhelmingly, concerns were expressed about the dominance of the accountability purpose and what the discussants saw as its detrimental ramifications. However, the expressed willingness of the discussants to avail themselves of national assessment materials (as benchmarks for themselves) seems to be based on the (faulty) premise that standards can be established immutably and in a vacuum without reference to the passage of time, population changes or technical considerations.

3.3   Measure improvement in overall attainment through a Scottish Survey of Achievement rather than relying on the Annual 5-14 Survey

3.3.1   The choice to be considered here was between continuing with the annual statistical exercise (referred to as the Annual 5-14 Survey) of collating and publishing the numbers of pupils in P2-P7 and S1 and S2 who can perform at each of attainment levels A to E in English reading, English writing, Mathematics, Gaelic reading and Gaelic writing or with a refined version of the Assessment of Achievement Programme (AAP) which would be called the Scottish Survey of Achievement (SSA).

3.3.2   AAPs were established by the Scottish Office Education and Industry Department in 1981 to monitor regularly the performance of pupils in Scottish schools in English Language, Mathematics and Science, and, more recently, in Modern Languages and Social Studies.

3.3.3   While the Annual 5-14 Survey is the statistical collation of all pupil performance in the ways described above, from returns made by local authorities, the AAP:

- samples pupils from S2 and from different stages in the primary school on a triennial, or more recently, a quadrennial cycle;
- uses a range of instruments, referenced to the content and levels of the 5-14 National Guidelines;
- provides comparative and progress evidence of performance between the different stages, between girls and boys, and over time.

3.3.4   Building on previously expressed concerns about the accountability agenda, the discussants were unreservedly of the view that the Annual 5-14 Survey was a sterile exercise. This was not the same as saying that monitoring was inappropriate, however. It was recognised that monitoring of attainment was useful to schools and to local authorities, though in various guises the point was made that *there was no need for this information to be collected centrally by the Executive.*

3.3.5   The acknowledged need to monitor achievement was seen to be well served by the introduction of the SSA. The discussants believed that if the SSA retained the existing AAP practices, then the samples would give good quality evidence of national performance. Further, the practice of disseminating AAP findings to schools in a way that enabled teachers to discuss the implications of these findings was seen as useful. While some did not agree with AAP sampling principles and considered the recent inclusion of P3 to have been *too stressful*, the overall feeling from the seminars seems to be that AAP was essentially worthwhile and therefore worthy of continuation.

3.3.6   However, there were reservations about some of the proposed extensions to be incorporated in the new SSA. The proposal to extend the sample to include S4 was questioned since there was information available from SQA on Standard Grade performance. The need to include pupils with special characteristics was seen as either contravening data protection requirements and/or an unworkable suggestion because of perceived difficulties in defining 'special characteristics'. While it was also acknowledged that information on the performance of special groups might be of use to authorities in planning provision, there was a slight concern in the perceived possibility that such information might be used to differentiate exclusively between groups.

3.3.7   Opinion was divided on the proposals to explicate the connections between overall survey sample performances and individual pupils. Some discussants made no comment, possibly because these particular proposals were declared to be *unclear and confusing*.

Some questioned the need for/value of making the details of individual pupils known while others viewed the proposals as retrograde suggestions which could once more make 'league tables' a reality. Support for the proposals was conditional on there being clear, and justifiable, purpose for the extraction of such information.

3.3.8   Views on what subjects to include and how to cycle the sampling of these were very mixed. Some discussants abstained altogether, either because they did not have enough information with which to make any judgements/comments or because they believed that this proposal was premature when many other aspects of the curriculum and of assessment had yet to be more fully developed. The inclusion of other subject areas was seen both to raise the status of such subjects and add further stress. Finally, the inclusion of embedded core skills was seen by some to be desirable, but for others it was a poorly defined idea which could not yet be contemplated.

3.3.9   As has been pointed out by many of the respondents and at different points in this consultation exercise, the issue of monitoring overall attainment on a nationwide basis is not of immediate concern to most teachers. Justifiably, the prior concern of teachers is how to enable learning and, therefore, how to use assessment to that end. However, teachers are sorely exercised when such monitoring is used for high-stakes purposes: that is, to make significant policy or evaluative decisions (about pupils, teachers, schools, authorities) on the basis of assessment instruments that are, necessarily, imperfect and/or are inappropriate for the purpose(s) to which they have been put. In considering whether to measure improvement in overall attainment through a Scottish Survey of Achievement rather than relying on the Annual 5-14 Survey, the discussants' wish to end the Annual 5-14 Survey was clear. Nevertheless, discussants accepted that monitoring was important and that, furthermore, they themselves would be happy to have knowledge of the findings of monitoring exercises to enable them to reflect on their practice and effect improvement to their teaching. How well the proposed Scottish Survey of Achievement will be able to 'deliver' on the preferences of professional teachers is not yet clear, however.

4   Key Messages from the Consultation Responses
(A fuller, more reflective version of this section is contained in the appendix)

4.1   In recent years advances in our understandings of both learning and assessment suggest that assessment can both improve and measure pupil learning. The demand by policy makers and the public for assessment that gauges learning **and** monitors achievement is particularly challenging and, as the results of this consultation exercise show, fraught with concerns about skewed classroom practice, pressure on teachers to improve pupil performance and potentially deleterious effects on teacher and pupil motivation and morale. Underpinning these very genuine concerns, however, are a number of differing interpretations of some fairly fundamental concepts in assessment. Such differences in understanding can impede progress in developing valid and reliable assessment practices.

4.2   Large-scale assessment in Scotland has developed unwittingly in the wake of the policy document, *Curriculum and Assessment in Scotland: a policy for the 90s* (1987), which listed the need for clearer definition of the curriculum; the establishment of satisfactory assessment policies; and better communication between schools and parents. While large-scale assessment can appear to be deceptively straightforward, its practices become increasingly suspect if/when the technical and theoretical limitations of assessment instruments are ignored. A lack of proper regard for principles can create considerable confusion in the planning and implementation of assessment, as was evidenced in the results of this consultation exercise. In particular, these show that there is very wide variation in meaning about:

- the basis on which any assessment judgement is made
- the notion of standards in education.

4.3   How should assessment judgements be made?

4.3.1   Assessment is a matter of judging the performance of self or other. Historically the basis for this judgement has been the performance of a defined reference group against which individual performance is compared (norm-referenced assessment). In recent years, norm-referenced assessment has been viewed as deficient because it does not describe actual achievements in education. Rather, predetermined levels of performance have become the basis for comparison (criterion-referenced assessment) in order to be able to provide explicit information as to what pupils can and cannot do. There are, however, some difficulties with referencing interpretations of performance in terms of criteria.

- One difficulty is in determining the criteria. Because of the need to provide explicit information as to what pupils can do, the specification of what constitutes excellent or adequate performance has to be precise and elaborate. Such detailed specification may reduce the assessment task to a set of routine, algorithmic subtasks making no authentic demands of the pupil.
- A second difficulty is in the type of criteria. Broadly, criteria can be task-centred, when all that is of interest is the constituent skills of a particular performance. But task-centred criteria may not always be appropriate in formal education, which is rarely concerned with one particular performance. The whole point of formal education is to enable people to think, reason, plan and make good decisions so that they can apply and extend their learning to a range of situations. Assessing a person's understanding requires construct-centred criteria that will distinguish between levels of proficiency in higher order thinking as well as the depth and breadth of subject matter knowledge.
- A third difficulty is in the use of determined criteria. The specification of clear criteria does not mean that they can be procedurally 'applied' in some reliable and consistent

fashion that is devoid of the assessor's own thought processes. Criteria themselves are the subject of interpretation and, when used to judge complex cognitive functioning, the issue of human interpretation becomes very significant. When judgements about the same event differ, whose judgement should be the benchmark?

4.3.2   The determination, type and use of criteria are extremely complex issues in assessment. Further, the views and ideas that each of us has on these issues affect our professional practices. Where there are gaps in our understanding about, our application of, and our beliefs in assessment criteria, the implications for professional development are clear. To fail to question/check that our shared values, meanings and understandings lead to equitable and moral practices is to perpetuate the myth that how we make assessment judgements is a simple, procedural matter.

4.4   What are standards in education?

4.4.1   In both lay and professional usage, the terms 'standards' and 'norms' are used synonymously. However, in the context of assessment this conflation is possibly confusing. Norms are derived from the actual performance of a group of persons who have been assessed on some very clearly defined variable according to a set of highly specified (that is, standardised) conditions. Norms, therefore, are not static, but vary according to whether the cohort on whom the norm is based is deemed 'good' or 'bad'. Standards, on the other hand, are the defining points that distinguish between different levels on a scale of competence. They represent what is desirable or what ought to be, and so are expressions of value rather than absolute measures. When using the term 'standards', it is important to be clear about which meaning is intended because the current tendency to conflate the terms can result in further confusion.

- What has passed for national testing in Scotland does not meet the rigorous technical requirements of 'normed' instruments. Rather, national test materials in Scotland have essentially been teacher-devised tests made publicly available and, while doubtless very useful in evidencing educational achievement, they should not be used for making comparisons between and among pupils, for evaluating nationwide achievement or for predicting potential. Norms - a distinctive feature of technically standardised tests - do allow comparisons but they are both expensive and time-consuming to create. While 'standards' within the Scottish context are not norms for the school population, it is also recognised that standards may well be established on the basis of what people generally consider to be normatively appropriate behaviour. This is not inappropriate, but it is unhelpful to assume that standards necessarily contain technically accurate normative information, the interpretation of which requires a familiarity with statistical concepts.
- It is often claimed that by raising standards the quality of educational instruction will improve and thereby create better schools. There is no evidence that elevating the threshold for defining levels of competence automatically results in better educational outcomes. However, this is not to say that there is no relationship between standards and quality. Large-scale assessment can draw limited attention to the educational success or failure of pupils and systems and can imply how professional practice might change or develop. Findings from such monitoring can shape policy and professional decisions about how to manage the assessment system. But they cannot account for the many non-school factors that are enmeshed in the effects of teacher behaviour and school influences.
- Confusion about standards often focuses on their arbitrary nature. Although educational standards are arbitrary in the sense that they can never be absolute (in the way that temperature, for example, can be measured), they should not be

thought of as impulsive or whimsical. Within the Scottish context, however, the issue of standards has been further confused by the repeated emphasis (particularly from HMIE) that the levels to be achieved were *minimum* levels, thereby implying that unless certain specifications were met, pupil achievements might be deemed to be not good enough and, by extension, that the teachers/schools in whose charge such pupils lay were 'failing the standard'. While they may not be the best possible constellation of potential educational outcomes, the current Scottish standards, as expressed in 5-14 National Guidelines, are unlikely to be abandoned in the immediate future since it is general social acceptance, rather than the inherent superiority of particular types of assessment, that creates workable standards.

4.4.2   The various nuances of meaning attached to 'standards' draw attention to the complexity of the construct. Rather than being seduced into simplistic understandings that standards can be operationally defined and characterised as a set of competences or attainment targets, we should recognise that there are no definitive standards in education. Such standards as do exist are a function of the dominant value position of the persons who determined them and, as such, can change.

4.5   The contradictions in responses to the consultation document are not surprising, given that they seem to rest on two very complex, and related, ideas: how assessment judgements should be made and the notion of standards in education. Until there is clarity on both of these ideas, there will continue to be a lack of clarity about the assessment system that is to be used. While the current system is the cumulative product of various prior conceptions of learning and measurement, assessment systems do really need to keep pace with what we know about how people develop understanding, how they reason, how their knowledge is shaped by social context and the nature of the thinking processes associated with competent performance. The professional development implied by the emergence of new information on learning and assessment cannot be 'delivered' by some magical wand-waving. Rather, it has to be incorporated in the very process of assessment improvement and reform. To the extent that respondents in the consultation exercise were unanimous in their concern that assessment should reflect, and should be properly seen to reflect, the learning that is actually taking place in schools and classrooms, the desire to develop understanding of assessment theory and practice is evident.

5   Summary and conclusion

5.1   In a consultation exercise on partnership commitments to:

- provide more time for learning by simplifying and reducing assessment, ending the current system of National Tests for 5-14 year olds
- promote assessment methods that support learning and teaching
- measure improvement in overall attainment through broad surveys rather than relying on the National Tests
- improve the transitions between nursery and primary and primary and secondary education so that the system fits the needs of the children
- promote improved assessment of individual schools' progress as a better measure than national 'league tables'
- strengthen the link between parents and schools through improving the quality of information that parents receive about their children's progress, and replacing reports with Annual Progress Plans

the Scottish Executive sought views on the issues of whether to:

- replace reports with Annual Progress Plans
- replace the current provision of National Tests with a National Assessment Bank
- measure improvement in overall attainment through a Scottish Survey of Achievement rather than relying on the Annual 5-14 Survey.

5.2   While the consultation questionnaire itself was variously viewed as:

- having been issued prematurely when a review of the curriculum was not yet complete
- having been issued at a time that did not articulate well with the developing AifL programme
- having been issued too late when it was understood that most of the proposals were already being developed
- not giving sufficient attention to pre-school when the declared age range of interest was 3 to 14 years
- not being as user-friendly as it might have been,

there was, nevertheless, a substantial response to the exercise. Consultation questionnaires were returned by 1071 respondents, about three quarters of whom were professional school staff. The remainder comprised parents, local authority officers and persons from other educational organisations. In addition about 250 persons (again mostly professional school staff) engaged in seminar discussion on the issues.

5.3   Views on the first issue, to replace reports with Annual Progress Plans, were almost evenly divided. While the current reports were seen to have disadvantages, they already had some of the alleged advantages of the proposed Annual Progress Plan. Furthermore, the proposal to change to an Annual Progress Plan was not well understood, largely because the implementation of the Annual Progress Plan was considered to be inter-related with implementation of Personal Learning Plans, the execution of which was still being piloted. In this context, 54% agreed with the proposal to develop Annual Progress Plans to a common framework which had scope for local adaptation and 46% disagreed. The alternative proposal, to produce a single national Annual Progress Plan/redesigned reports format that would be agreed and used by all schools, found support from 48% and resistance from 52% of the respondents.

5.4   The second issue of whether to replace the current provision of National Tests with a National Assessment Bank did not actually turn on a straightforward dichotomy. The overall conclusion was that respondents wanted assessment material to be available. This was confirmed in the statistic of 80% wanting the provision of National Test materials to continue. The issue for them was not whether the materials should be in the current or new form but whether the materials were available to teachers/schools as a resource or whether the availability of materials necessarily implied the constraints and stringencies of National Testing per se.   Opposition to National Testing regimes was explicit.   This was further confirmed by 82% of the respondents who supported the introduction of a National Assessment Bank in the expressed assurance of the wording of the option that the materials in the bank could be used to confirm teachers' judgements. Views varied as to what should be included in/excluded from the new assessment bank but almost all were premised on the belief that the professional decision-making of teachers should determine when and how assessment took place.

5.5   The third issue of whether to measure improvement in overall attainment through a Scottish Survey of Achievement rather than relying on the Annual 5-14 Survey triggered less response than did the other issues, partly because some of the questions were declared to be unclear, partly because the questions were not seen as pertinent and partly because the substance of respondents' views was declared to have been overtaken in the previous issue. Within this context, however, 68% of the respondents wished the Annual 5-14 Survey to end, an equivalent proportion favoured the introduction of the Scottish Survey of Achievement but 79% were resistant to new subjects being included in the survey.

5.6   The views expressed on the three issues are global ones and mask differences between class/subject teachers and groups of teachers on the one hand and senior management teams, parents and other organisations on the other hand. The class or subject teachers and the groups of teachers – in other words those whose professional practice is exclusively taken up with minute-to-minute interactions with, and decision-making on the part of, learners – were at odds with the senior management team (and others) who make overall policy and strategic decisions but who are not necessarily involved in the day-to-day minutia of curriculum delivery. They had significantly different ($p<0.001$) views on:

- the introduction of a common framework for reporting achievement
- the use of assessment banks, the inclusion of additional materials in the assessment bank and the issue of local moderation
- the introduction of a Scottish Survey of Achievement, the inclusion of new subjects in the survey and the uses to which data from such a survey might be put.

5.7   Specifically, and most notably in Aberdeenshire, Edinburgh City, North Ayrshire and South Lanarkshire:

- senior management were supportive of Annual Progress Plans which had scope for local adaptation, whilst teachers anticipated unnecessary difficulties in the proposal and would prefer to retain the existing annual report
- although for both senior management and for teachers the preference was for the introduction of the National Assessment Bank, significantly fewer of the senior management had reservations about the proposal, whilst the teachers more readily pointed up its human and financial resource implications
- most senior management were in favour of local moderation while most teachers were not, on the grounds either that moderation was difficult to effect or that it negated the notion of a national standard
- to each of the suggestions for extending the new assessment bank the teachers were unequivocally resistant, for fear that this would intensify what they saw to be an

already bureaucratic assessment culture, whilst senior managers were more relaxed on the issue

- the proposal to introduce the new Scottish Survey of Achievement was positively endorsed by majorities of both teachers and senior management but, just as for the introduction of the National Assessment Bank, greater proportions of teachers were wary of the proposal, either because of resource and workload implications or because it was perceived to exacerbate what were currently seen as unfortunate and deleterious consequences of current monitoring practices.

5.8   Even if teachers' concerns are unfounded (and there is nothing in the data to suggest that they responded other than completely truthfully), the difference between their views and those of senior managers and others suggests that full development and execution of the partnership commitments set out by the Scottish Executive are going to be considerably hampered unless there is some genuine attempt for all involved not only to understand what is meant by an effective system of assessment but also to understand what the costs and benefits are for each stakeholder of such a system and how the points of difference and tension can be reconciled. It is suggested that an effective assessment system must build on shared understandings of fundamental constructs (such as the basis for making comparative judgements and the notion of standards in education) which have to be agreed through argument and discussion and which take account of the up-to-date literature on learning and measurement.

5.9   Overall, the number of responses, together with the detail in some of the responses, points to considerable interest in, and concern about educational assessment. The desire to make educational assessment as good as it can be was evident in all responses. This is a most heartening message to emanate from the consultation exercise and one which provides a sound platform on which to work towards resolving the tensions that currently exist.

6   Appendix 1:   Reflecting on the Consultation Responses
(This is a more developed account of Section 4)

6.1   In recent years advances in our understandings of both learning and assessment challenge us all to try to find ways of improving assessment so that it can both improve and measure pupil learning. The demand by policy makers, educators and the public that large-scale assessment should serve a variety of purposes is neither new, nor unique to Scotland. To date, however, progress in developing assessment that gauges learning/provides useful feedback for pedagogical decision-making, **and** monitors achievement at pupil, school and local authority level has been slow and, as the results of this consultation exercise evidence, fraught with concerns about skewed classroom practice, pressure on teachers to improve pupil performance and potentially deleterious effects on teacher and pupil motivation and morale. Perhaps it is now time for us all to rethink some of the assumptions, values and beliefs that currently inform our view of large-scale assessment.

6.2   This appendix attempts to step back from the findings to gain an overall perspective. Through relating the most common concerns expressed in the responses to what is currently understood about educational assessment, it is hoped to offer clarification on what looks like confusion in the responses. Such clarification is intended to provide a basis on which decisions about staff development and refinement of policy might be considered. It is also worth noting that such clarification is **not** making recommendations on either: merely exposing what seems to be at issue.

6.3   The responses to the questions in the consultation document at first may suggest a conflicting and perhaps unhelpful picture. In brief: the proposal to change the reporting of pupil progress is not seen as an improvement on extant arrangements; the introduction of a National Assessment Bank is seen as desirable but so, too, is the maintenance of the current provision of national test materials; there is a clear desire for the new Scottish Survey of Achievement to replace the existing Annual 5-14 Survey but there are concerns about the scope of the survey and fears that its findings might be used inappropriately; and finally, teachers and senior management teams seem to have different perspectives on what is problematic. However, reflection on the totality of the responses suggests that many of the views expressed are entangled in educationally historical events.

6.4   Until the early 80s, when government was articulating society's concerns that falling standards, the variation in quality of provision in different schools and the decline in standards of personal and social behaviour were allegedly attributable to progressive teaching methods, Scottish Education was largely devolved to local education authorities, and curriculum development, implementation and evaluation were essentially school-based. A child-centred, humanitarian philosophy was the official creed and the values espoused in the Primary Memorandum of 1965 were those to which the teaching profession allegedly aspired. Attempts to make education more effective were first manifest in *Curriculum and Assessment in Scotland: a policy for the 90s* (1987) in which the need for clearer definition of the content and objectives of the curriculum; the establishment of satisfactory assessment policies in all schools; and better communication (including reporting on pupils' progress) between schools and parents were listed as underpinning the reforms to the curriculum in Scotland. The policy document was to be the first of many in which guidance on the curriculum and on assessment was offered nationally, thereby inexorably re-routeing what had previously been relatively local decision-making through a national filter.

6.5   While policy documents on Scottish education were/are advisory rather than subject to the legislative process, there is no evidence that they are anything other than the 'drivers' for curriculum and assessment both in the primary school and in the interface between pre-school and primary education and between primary and secondary education. The current

conception of what Scottish education is like between the ages of 5 and 14 years necessarily gives rise to a curriculum that is specified, planned, implemented and evaluated by objectives. Once objectives have been specified, their achievement can be assessed. By extension, if all pupils are working to the same objectives, then the availability of a common assessment instrument allows the possibility of assessing across the school population. Additionally, by discounting the technical and theoretical limitations of assessment instruments, it is then possible to claim to describe the performance of the pupils across the nation and, hey presto, large-scale assessment is born. Further, if the results of large-scale assessment are used to make decisions about the progress of pupils, the allocation of resources or the effectiveness of teachers, such assessment is also of the high-stakes variety. Inadvertently, the significance attached to large-scale assessment (which has developed in the wake of the 5-14 National Guidelines) has created a web of confusion and misconception that finds strong, if residual, expression in the responses to this consultation exercise. An attempt to unpack this confusion follows and is structured round two issues: the basis for making comparative judgements and the notion of standards in education.

6.6   The basis for making comparative judgements

6.6.1   Stripped to its essentials, assessment is a matter of interpreting (and judging) the performance of self or other. Historically, the dominant method of interpreting performance in educational assessment has been by comparing the results of one individual with those of a well-defined reference group. The data from the relevant reference group contextualise the extent to which the individual's performance is consistent with/deviant from average. While such norm-referencing usefully gives meaning to measures such as blood pressure or cholesterol level, it is arguably less useful in educational assessment. Yes, norm-referenced information can help parents, teachers and others to determine whether pupils are progressing at the same rate as their peers or whether they are above or below the average, but it does not describe actual achievements (Glaser, 1963; 1990). To redress this perceived deficiency, predetermined levels or standards of performance became the basis for comparison in order to be able to provide explicit information as to what pupils can and cannot do. In recent years, such criterion referencing has gained popularity amongst the different stakeholders. Because pupils can supposedly be measured on the extent to which they achieve the criterion performance, policy makers and the wider community can both seek greater accountability from, and exert greater control over, the education service. Similarly criterion referencing is attractive to teachers and pupils because it can provide markers of progress and meaningful information to both constituencies. The perceived illuminative information provided by criterion-referenced assessment, together with the more diffuse societal disapproval of comparing people, has tended to favour criterion referencing in assessment practices generally. There are, however, some difficulties with referencing interpretations of performance in terms of criteria.

6.6.2   One difficulty is in determining the criteria. Because of the need to provide explicit information as to what pupils can do, the specification of what elements of performance are desired and what the criteria of excellent and adequate performance are in each case (Resnick & Resnick, 1993) can become precise and elaborate. A potential disadvantage in such detailed specification is that the assessment task is reduced to a set of routine, algorithmic subtasks making no authentic demands of the pupil, and thereby negating the pedagogical and philosophical underpinnings of assessment that is intended to support learning (Wiliam, 1998). Of more significance, however, is the decision about the type of criteria to specify.

6.6.3   Because of the intention that assessment should support learning, the performance being assessed has to be referenced against a performance domain. In other words, performance is judged on how well the constituent skills of the performance are manifest

(Motowidlo et al, 1990; Russell & Kuhnert, 1992). This, logically, results in the criteria for evaluation being task-specific. If all that counts is the quality of the artefact or performance offered for evaluation (as in a contest, a competition, a festival or exhibition), then task-specific criteria can be perfectly adequate. So long as the assessment task elicits the skills underlying the performance in the domain of interest (as in acting, dancing, painting, participative sport and so on), it is doing what it purported to do and therefore the task can be said to be valid. That the performance per se and the assessment criteria are essentially the same thing is what Messick (1994) refers to as task-driven assessment. Messick notes the increasing interest in, and his concern(s) about, task-driven assessment.

6.6.4   While for many real world applications, such as passing a driving test, confirming or disconfirming a suspected pregnancy, achieving a National Vocational Qualification (NVQ), task-driven assessment is appropriate, it may not always be appropriate in formal education, which is rarely concerned with one particular performance. If people are to learn to think, reason, plan and make good decisions (which is a significant aim of formal education), they must be able to transfer what they have learned in the past to new learning and be able to apply and extend their learning to a range of situations (Haskell, 2001). Indeed the whole point of formal education is that people should transfer learning effectively and flexibly. To assess whether or not a person understands the underlying attributes or variables that represent the crucial components of the skilled performance (and can thus draw on them at will) requires assessment that is what Messick (1994) terms construct-driven. Here the construct guides the selection of the task as well as the development of the scoring procedures. This is distinctly different from task-driven assessment, where the focus is primarily on a worthy task and then unpacking it for its constituent skills. While task-driven assessment is not necessarily devoid of constructs, reference to them in task-driven assessment is at an implicit or informal level only and is not systematically related to the scoring procedures. Because construct-driven criteria should include, according to Linn et al (1991), cognitive complexity (the processes of higher order thinking that are required to be exercised), content quality (the depth of subject matter expertise) and content coverage (the breadth of domain representation) it is not difficult to appreciate that such criteria may be complex to either conceptualise or represent.

6.6.5   The issue of whether assessment should be task-driven or construct driven or, indeed, whether it is appropriate to try to categorise assessment in this way, is far from resolved, although very much alive in the responses to the consultation exercise (as manifest in the concerns about national testing and core skills). If the interpretation of assessment performance is in terms of minutely explicated criteria, it is possible that "instructional adaptations" (Frederiksen & Collins, 1989, p28) can skew the learning to focus more on how to improve the performance score (a point that was clearly reflected in the consultation exercise) rather than on the underlying cognitive skill and knowledge which extends to a range of problems, particularly if the assessment is for accountability purposes. Indeed, Bereiter & Scardamalia (1987) argue that educational practices in schools (such as assessing only the content which has been taught on a module/unit, requiring learners to assemble knowledge on a single topic, framing assessment tasks to provoke learners' spontaneous recall of knowledge, and awarding credit to learners who appear to have learned something of the intended material even though they have not actually addressed the assessment problem as it was presented), inadvertently avoid engaging learners in intellectually generative tasks and thereby privilege the encoding and manipulation of propositional knowledge at a superficial level only, with the consequence that pupils can do no more than regurgitate such knowledge.

6.6.6   As well as difficulty in determining the criteria, use of determined criteria is also problematic. Because assessment is making judgements often about complex cognitive functioning, the issue of human judgement becomes significant. And since human

judgement about any particular event can differ, dramatically, both within persons across time and amongst persons, the reliability of judgements is a serious issue (a point that was clearly reflected in the consultation exercise). When judgements about the same event differ, whose judgement should be the benchmark? Because, according to Wiggins (1992), judges should know specifically where and what to look for in performance, the issue of reliability is often seen as being resolved in the specification of clear criteria. However, as Wiliam (1996) points out, consistency does not reside in external, pre-specified criteria and so to believe that reliable marking is a function of specifying clear criteria is naïve. That criteria themselves are the subject of interpretation is recognised in the practices of training and moderation where individuals learn to rate performances to agreed standards or otherwise acquire shared understanding of performance standards (Baker et al, 1993; Resnick & Resnick, 1993). In the process of rating, one's substantive knowledge, one's contextually derived expectations of what is appropriate and one's beliefs as to how learning occurs all subtly influence, and thereby mediate, the judgements made (Baker & O'Neill, 1994). In other words, as Angoff (1974) pointed out many years ago, "lurking behind the criterion-referenced evaluation, perhaps even responsible for it, is the norm-referenced evaluation" (p4). That criterion-referenced assessment is not as distinct from norm-referenced assessment as we might like to believe, suggests that all of us concerned with education should be prepared to question our normative assumptions to check that our shared values, meanings and understandings lead to equitable and moral practice (again a message that was reflected in the respondents' concerns for potentially discriminatory assessment and reporting practices being perpetuated or introduced). Given the importance of consistency in marking and the concerns expressed by the respondents about the lack of marking consistency in the recent history of assessment, the need for training in reliable rating is clear.

6.6.7   Although criterion referencing is the typically preferred basis for interpreting assessment information, the idea itself is fraught with practical problems and conceptual confusions. Well delineated descriptions of performance are both fundamental in criterion referencing and difficult to specify for covert cognitive functioning; criteria can be specified either by domain (in task-centred criteria) or by levels of proficiency (in construct-centred criteria); and even the specification of criteria does not guarantee agreed or uniform understanding of what is actually the target of assessment. It is within this context that teachers are trying to support learning, provide feedback to pupils, parents and other teachers, and to identify next steps in learning; and administrators are trying to manage provision and to monitor and evaluate attainment at school, education authority and national levels. Given this, it is not at all surprising that the views expressed in the consultation exercise appeared contradictory. What they point to is the need for the Scotthish Education community to determine more fully the model of assessment that is to be used.

6.7   **The notion of standards in education**

6.7.1   A further ramification of the confusion surrounding the basis for interpreting assessment information is reflected in the respondents' use of the word 'standard'. Typically they commented on the need/desire to know the national 'standard' in comparison with the performance of their own pupils. From this one can infer that the respondents were seeking to make norm-referenced interpretations However, use of the word 'standard' in this way is possibly confusing. Fairly frequently, the terms 'standards' and 'norms' are taken to be synonymous, and it is perfectly easy to understand why. Norms are used with standarised tests and the development of norms is part of the process of standardisation. But test norms are based on the actual performance of a group of persons, not on predetermined levels or standards of performance. For example, at the present time we are told that many adults are overweight, in which case the weight norm will be high. Desirable or ideal weight (in other words, the standard) is, however, explicitly documented in a range of medical literature

and inspection of it will reveal that the norm is currently higher than the standard. The word 'standard' implies a goal or objective to be reached whereas a norm is measure of the status quo and has no connotation of what is desirable or what ought to be. This point of difference is raised not merely to be pedantic. If respondents are really wanting to know the norms, as distinct from standards (in the way differentiated here), they must also appreciate that norms are not static, but will vary according to whether the cohort on whom the norm is based is deemed 'good' or 'bad'. Further, they need to appreciate that the determination of norms, in the strictest sense, involves:

- careful definition of the domain and detailed identification of the content/objectives to be assessed
- generating and trialling a large number of assessment items
- piloting the items on a representative sample of the population for whom the assessment instrument is intended
- selecting from the pool of items generated both on their appropriateness and on their facility to discriminate between testees
- piloting what is thought to be the final version of the assessment instrument (on a sample representative according to age, gender, ethnic group, socioeconomic status, geographical location) to check for clarity in the administration and marking of the assessment instrument and to develop the norms against which future individual scores will be compared.

6.7.2 No claim is being made here for superiority of technically standardised tests. Standardised tests report performance in some form of standard score, which is immediately and commonly understood by those with sufficient statistical knowledge. However, the stringencies of test construction within the psychometric paradigm mean that the technical matters of validity, reliability and the appropriate use of norms can result in the construct to be assessed being very narrowly or idiosyncratically defined. In other words, standardised tests can describe very accurately the performance of populations of pupils on some very specifically defined measure of educational achievement, but the range of behaviours included in the achievement, while satisfying technical specifications, may not be commonly agreed to be relevant and/or appropriate. Because the enormous power of standardised tests is that they enable the comparison of performance between one pupil and his/her contemporaries, such tests may be viewed as promising more than they deliver and so used inappropriately. While there have been standardised tests constructed for criterial purposes (such as diagnosing strengths and weaknesses), such tests are fraught with content and logistical problems, and so are not best suited to the pedagogical purposes of promoting learning. By implication, the assessment tasks devised by teachers become all the more significant. With this comes the very onerous task of ensuring that the assessment methods used before, during and after instruction are as valid and reliable as they can be, and while perfect reliability and validity is never possible, the teacher's responsibility to gather and interpret valid and reliable data for decision making are fundamental to the ethics of assessment. Given that standardised tests serve limited purposes (which have little overlap with the intentions of the 5-14 National Guidelines), it becomes clear that what are referred to as 'standards' within the Scottish context are not norms for the school population. What has passed for national testing in Scotland does not meet the rigorous technical requirements of 'normed' instruments. National test materials in Scotland have essentially been teacher-made tests made publicly available, and while doubtless very useful in evidencing educational achievement, they should not be used for generating norms, for evaluating nationwide achievement or for predicting potential. If norms are really what are wanted in order to satisfy comparative purposes, it has to be recognised that they are both expensive and time-consuming to create.

6.7.3   Standards are the defining points that distinguish between different levels on a scale of competence. The size of this scale will vary depending on the number of gradations desired to describe performance. While there are well developed (but problematic) psychometric procedures for setting standards (in the parlance, for setting cut-scores), standards represent diverse values about what we think is important in education, resulting in ambiguity in meaning if not controversy among stakeholders (Gipps, 1990; Moss, 1992; Pring, 1992). Debate around the credibility of standards takes a variety of forms. One is the perennial issue of whether standards are rising or falling, largely fuelled by hopelessly inadequate media reporting of assessment 'results'. This issue is essentially one of comparing norms across cohorts and has been dealt with in the previous discussion.

6.7.4   Another issue is the relationship between standards and improvement, often rehearsed in the spurious claim that by raising standards the quality of educational instruction will improve and thereby create better schools. There is no evidence that elevating the threshold for defining levels of competence automatically results in better educational outcomes. Indeed, Coffman (1993), Linn (2000) and others would question whether standards can be raised or whether the reporting of elevated achievement is a function of deflated norms, repeated use of the same assessment instruments, the teaching of test-taking skills and the exclusion/non-participation of particular categories of pupils in the baseline for comparison. Just as it is inaccurate to equate standards with educational quality, so, too, is it inaccurate to claim that there is no relationship.

6.7.5   While it is true that weighing a child does not cause the child to grow, periodic checks on the child's weight give an indication of whether diet and exercise or further physiological tests are implied. Similarly, large-scale assessment can shed light (albeit limited) on the educational health of pupils and systems and can imply how professional practice might change or develop. Just as the monitoring of survival rates of cancer sufferers does not detract from the oncologists' skill to diagnose and treat malignancy, the findings from such monitoring can shape policy and professional decisions about how to manage the disease. And just as oncologists should not be considered the sole explanatory mechanism for the success or otherwise of their medical interventions, so, too, should teachers not be considered the sole explanatory mechanism for the educational success or failure of pupils when many non-school factors cannot be separated from the effects of teacher behaviour and school influences (Coffman, 1993).

6.7.6   A third issue in the debate about standards is their arbitrary nature. While, of course, educational standards are arbitrary in the sense that they can never be absolute (in the way that temperature, for example, can be measured), they should not be thought of as impulsive or whimsical. Far from being selected at random and without reason, clearly articulated educational standards are meant to serve as the benchmarks that everyone understands to be the important and tangible outcomes of education. While they may not be the best possible constellation of potential educational outcomes, the current Scottish standards as expressed in levels A-F of the 5-14 documentation are unlikely to be abandoned in the immediate future since it is general social acceptance, rather than the inherent superiority of particular types of assessment, that creates workable standards. Within the Scottish context, however, the issue of standards has been further confused by the repeated emphasis (particularly from HMIE) that the levels to be achieved were *minimum* levels, thereby implying that unless certain specifications were met, pupil achievements might be deemed to be not good enough and, by extension, that the teachers/schools in whose charge such pupils lay were 'failing the standard'.

6.7.7   The various nuances of meaning attached to 'standards' draw attention to the complexity of the construct. Rather than being seduced into simplistic understandings that standards can be operationally defined and characterised as a set of competences or

attainment targets, we should perhaps recognise that there are no definitive standards in education. Such standards as do exist in the current 5-14 documentation are a function of the dominant value position of the persons who determined them and, as such, can change. While it is currently of national priority to "raise the standards of attainment for all in schools", "to help every pupil benefit from education" and "to equip pupils with the foundation skills, attitudes and expectations necessary to prosper" (*Educating for Excellence: the Executive's Response to the National Debate*, 2003), such standards cannot be achieved by schools or teachers alone. Yes, there may well be advances in pedagogy as a result of the *AiFL* programme (though that has not yet been fully worked through) which will harness the creativity and capacity of our pupils, but we must also recognise that there are dramatic differences between pupils on entry to, and throughout, school. Some will learn despite considerable obstacles and some will thrive in supportive environments. Others, however, may not be as open to learning or, indeed, may not be prepared to learn. The differential conditions that affect learning can be enormous. However, these would seem to be well recognised by the Scottish Executive who declare social and economic action in the achievement of excellence. In the meantime it might be helpful if we could all extend our understanding of the concept of standards so that they are not represented as ill-summarised versions of pupil achievement, devoid of the conditional interpretations that richly characterise achievement, which are then inadequately interpreted/reported by the media and general public.

6.7.8   It has been suggested that the contradictions in responses to the consultation document derive from assumptions, values and beliefs that have evolved over time but that have not been scrutinised in the light of the current political and educational context. These contradictions in respondents' views are not surprising, given that they seem to rest on two very complex, and related, ideas: the basis for making comparative judgements and the notion of standards in education. Until there is clarity on both of these ideas, there will continue to be a lack of clarity about the assessment system that is to be used. While the current system is the cumulative product of various prior conceptions of learning and measurement, and while some of these foundations may still have qualified use, assessment systems really do need to keep pace with what we know about how people develop understanding, how they reason and build structures of knowledge, how their knowledge is shaped by social context and the nature of the thinking processes associated with competent performance. At the same time it is important to recognise that assessment works within the constraints of the larger educational system and so the positive potential of new forms of assessment can be impeded by limited resources, by large class sizes, by too litte time for the various stakeholders in education to interact and by the lack of alignment in curriculum, instruction and assessment. These constraints cannot be ignored, but equally they cannot be resolved by some magical wand-waving. Rather they have to be incorporated in the very process of assessment improvement and reform. To the extent that respondents in the consutation exercise were unanimous in their concern that assessment should reflect, and should be properly seen to reflect, the learning that is actually taking place in schools and classrooms, the desire to develop understanding of assessment theory and practice is evident.

References

Angoff, W. (1974) Criterion referencing, norm referencing and the SAT, *College Board Review,* 92, 3-5, 21.

Baker, E., & O'Neil, H. (1994) Performance assessment and equity, *Assessment in Education* 1(1), 11-26.

Baker, E., O'Neil, H. &Linn, R. (1993) Policy and validity prospects for performance-based assessment, *American Psychologist* 48, 1210-8.

Bereiter, C. & Scardamalia, M. (1987) *The Psychology of Written Composition*. NJ: Lawrence Erlbaum Associates.

Coffman, W. (1993) A king over Egypt, which knew not Joseph, *Educational Measurement: Issues and Practice*12(2), 5-8.

Frederiksen, J. & Collins, A. (1989) A systems approach to educational testing, *Educational Researcher* 18(9), 27-32.

Gipps, C. (1990) *Asessment: a teacher's guide to the issues*. London: Hodder & Stoughton

Glaser, R. (1963) Instructional technology and the measurement of learning outcomes: some questions, *American Psychologist*, 18, 519-21.

Glaser, R. (1990) Toward new models for assessment, *International Journal of Educational Research*, 14(5), 475-83.

Haskell, R. (2001) *Transfer of Learning.* London: Academic Press.

Linn, R. (2000) Assessments and accountability, *Educational Researcher*, 29(2), 4-16.

Linn, R., Baker, E. & Dunbar, S. (1991) Complex performance-based assessment: expectations and validation criteria, *Educational Researcher* 20(8), 15-21.

Messick, S. (1994) The interplay of evidence and consequences in the validation of performance assessments, *Educational Researcher* 23(2), 13-23.

Moss, P. (1992) Shifting conceptions of validity in educational measurement: implications for performance assessment, *Review of Educational Research* 62, 229-58.

Motowidlo, S., Dunnette, M. & Carter, G. (1990) An alternative selection procedure: the low-fidelity simulation, *Journal of Applied Psychology* 75, 640-7.

Pring, R. (1992) Standards and quality in education. *British Journal of Educational Studies*, 40(1), 4-22.

Resnick, L. & Resnick, D. (1993) Assessing the thinking curriculum: new tools for educational reform <u>in</u> B. Gifford & M. O'Connor (Eds) *Changing Assessments: alternative views of aptitude, achievement and instruction.* The Netherlands: Kluwer Academic Publishers.

Russell, C. & Kuhnert, K. (1992) New frontiers in management selection systems: where measurement technologies and theories collide, *Leadership Quarterly* 3, 109-36.

Scottish Executive Education Department (1999) A Review of Assessment in Pre-school and 5-14.

Scottish Executive Education Department (2000) Improving Assessment in Scotland.

Scottish Executive Education Department (2003) A Partnership for a Better Scotland: Partnership Agreement.

Scottish Executive Education Department (2003) Educating for Excellence: the Executive's Response to the National Debate.

Scottish Office Education and Industry Department (1987) Curriculum and Assessment in Scotland: a policy for the 90s

Wiggins, G. (1992) Creating tests worth taking, *Educational Leadership* 49(8), 26-33.

Wiliam, D. (1996) Meanings and consequences in standard setting, *Assessment in Education* 3(3), 287-307.

Wiliam, D. (1998) Construct-referenced assessment of authentic tasks: alternatives to norms and criteria. Paper presented at the 24th Annual conference of the International Association for Educational Assessment – Testing and Evaluation: Confronting the Challenges of Rapid Social Change, Barbados, May 1998.