

THE GRADING OF PLACEMENT IN INITIAL TEACHER EDUCATION IN SCOTLAND

STEPHEN SHARP

ABSTRACT

Assessment of placement in Initial Teacher Education programmes is arguably more problematic than more traditional academic assessment. Students on placement are faced with different challenges in a way that students writing the same essay are not. Some writers have argued that the grading of placement is impossible. This paper analyses essay-based percentage marks and placement-based A to E grades obtained by students following two ITE programmes. The data show separate dimensions of achievement on placement and in essays. However the data also show that the assessment scale is used very differently for essays and for placement with a much larger percentage of high grades given for the latter. The paper concludes that the grades which have in the past been given for placement are not wholly unreliable or invalid but that placement is sufficiently distinct from academic work to be reported separately, using a scale of distinction – satisfactory – unsatisfactory.

INTRODUCTION

Most programmes of initial teacher education (ITE) seek to enable students to understand education both as an academic discipline to be studied and as a professional skill to be practised. Courses cover the social and political contexts of education and relevant psychology such as theories of child development as well as equipping students with a range of skills and competences to enable them to function effectively as teachers. This diversity of aims has implications for curriculum content. ITE provision is generally located in faculties and institutes in universities and they must pass internal review by colleagues from other social sciences whose traditions lean more strongly towards the academic aspects of the subject. At the same time however, courses, if they are to be accepted as entrances to the profession, must be externally accredited by bodies like the General Teaching Council for Scotland, whose concerns are inevitably more orientated towards the professional aspects. Course designers may not always find it easy to be the servant of two masters as it were but with careful wording of course documentation and a willingness to view the professional and academic perspectives as having a symbiotic rather than a competitive relationship, the requirements of both camps can be satisfied.

There is however one aspect of ITE courses which has been the subject of particular debate and tension: assessment. How teaching and learning in higher education should be assessed (and specifically how it should be reported) has been the cause of ongoing debate. One of the most recent contributions was the Burgess Report (Universities UK, 2004) which advocated the phasing out of the traditional honours classification for undergraduate degrees and investigating the use of Higher Education Transcripts, Progress Files, Personal Development Planning or other reporting mechanisms, which better meet the needs of different audiences. At least one academic writer (Elton from 1969 to 2004) has devoted much of his career to arguing against the first/upper second/lower second etc reporting system. Even if it is accepted that the traditional reporting system continues to be appropriate for academic assessment, it does not follow that it should be so for professional assessment.

There is a particular difficulty concerning the assessment of professional performance. In practice this means the assessment of competence shown by students while on school experience or placement (in this paper the term 'placement' will be

used for convenience). Assessment of placement is arguably more problematic than more traditional academic assessment. If two students sit an examination or write an essay or compile a literature review, the results of their labours are comparable in the sense that they have been given the same task to do and the same resources (such as libraries and opportunities for learning) to equip them to do it. Two students on placement however are not faced with the same problems. Schools differ widely in terms of the levels of challenge they pose to students and the levels of support they offer. There is also the interaction between the student's strengths and weaknesses and the school context – the same school and the same pupils might be a fulfilling and rewarding experience for one student and a professional nightmare for another. Establishing a basis for ensuring the comparability of grades so that a B in one place and time is equivalent to a B in another place at another time is that much harder, and some have questioned whether it is possible at all.

In an attempt to develop such a framework for comparability, the Scottish Executive Education Department has set out the competences which should be demonstrated by student teachers while on placement. Or at least it is possible that this was the intention behind the drafting of competences. Some writers (eg. Jones and Moore, 1993; Stronach, *et al.* 1994; Humes, 2001) suspect that the real motive is to de-professionalise teaching or to replace professional autonomy with state control. Whatever the motivation, the move to express the qualities and attributes of a teaching professional as a set of competences has been accused of being philosophically incoherent (Carr, 1993). It has also been argued that it represents a move towards the 'scientific' model of assessment based on external control, quantification, objective tests and maximum discrimination between students and away from the 'judgemental' model based on contextualised knowledge, multiple sources of evidence and professional judgement (Hager and Butler, 1996; Martin, 1997, Martin and Cloke, 2000).

In a particularly relevant recent paper, Cope, *et al.* (2003) attack the practice of grading placement (ie. the use of A to E or other reporting scales such as are often used for assessing academic work). They refer to this practice as 'an unholy union of incompatibles', a phrase which originated with a medical academic. It is not a coincidence that the grading of practice should also be particularly problematic in medicine. There is little tradition in this discipline of the use of grades in the public reporting of assessment. Medical degrees are usually awarded on a pass/fail or satisfactory/unsatisfactory basis since while most patients would be pleased to be treated by a first or possibly upper second class doctor, the presence of lower second and third class doctors in the NHS would do little to promote public confidence in the service. Doctors are either qualified to practice or they are not. Could the same argument not be forwarded for teachers and other professionals?

For many ITE courses this is already the case. One year PGCE or PGDE programmes taken after completing a degree in another subject are often awarded either on a pass/fail (or satisfactory/unsatisfactory) basis. Multiple point (usually A to E) grading is more common for four year specialist BEd degrees and, whatever its drawbacks, there are sound educational arguments in its favour. If students are following a programme where academic work is graded while teaching practice is merely certified as satisfactory, it will be very difficult to maintain the two types of work as being of equal status in the eyes of the students. Moreover, the absence of any further recognition of the quality of placement performance above the single benchmark runs the risk of encouraging minimalism. Students will do as much as necessary to ensure that they reach the necessary minimum standard (which will one assumes be placed at a point where the vast majority of students can surpass it), but as there is no summative incentive for them to do more than this, it could be strategic of some of them not to do more, or at least not as much more as they would have done had there been higher hurdles to clear.

The objections which have been raised to the grading of placement using competences are based on the inherent impossibility of comparing the levels of performance of different students with different strengths and weakness working in different schools which present different challenges and different types and degrees of support to overcome them. Referring to competences as the basis for comparability does little to reassure Cope, *et al.* who endorse the criticisms of this approach which have been a feature of the literature since it was first introduced into ITE in the early 1990s. And undeniably there are formidable problems confronting any attempt to justify the grading of comparability on theoretical grounds (empirical evidence regarding the grading of placement is presented later in this paper).

In fact the theoretical objections to comparability are perhaps stronger than the critics realise. If the objections apply to the A/B, B/C and C/D boundaries in a conventional five point grading scheme, why do they not also apply to the D/E boundary, which is usually the satisfactory/unsatisfactory cut-off point? In a way, this brings us back to the medical analogy. The use of unclassified degrees, and the fact that assessment procedures were unique to each university have until recently meant that medical schools assumed (or perhaps hoped) that their standards were comparable at the one assessment point which existed. Recent research made possible by moves towards standardising medical assessment, has enabled this to be investigated, with somewhat disquieting results (Roberts, 2004). Using only one point may reduce the size of the comparability problem but it does nothing to solve it.

If the arguments against the possibility of reliable and valid assessment of placement apply without qualification, the logical consequence is the impossibility of any summative assessment of placement at all. But there has been no attempt in the literature to argue that the assessment of placement should be abandoned entirely. In this sense, there is an incompleteness in the argument of much of what has been written about competences and their use in assessment. While writers in this area have been vociferous in articulating the deficiencies of official proposals for assessing ITE, they have not been quite so forthcoming in proposing alternatives which they regard as being more defensible. It is useful therefore to look at the current official position about the assessment of placement with a view to finding a way forward.

In Scotland, the guidelines currently in force are those published in 1998 by the then Scottish Office Education and Industry Department. Paragraph 1.5 simply states that “each block of school experience must be assessed” though paragraph 1.11 amplifies this:

The quality of students’ teaching and classroom management must be assessed and the award of a teaching qualification will be conditional on such quality being satisfactory. The scheme of assessment and award of a teaching qualification should make provision for the recognition of particular merit.

The implication here seems to be that a three point scale of unsatisfactory – satisfactory – distinction is being recommended. More recently however, the report of the ‘First Stage’ review of ITE (Scottish Executive Education Department (or SEED), 2001) recommended that

assessment processes for students on placement be considered in the light of the national profile of professional experience which is currently being introduced throughout Scotland and will take effect from June 2002... in order to assist schools in achieving consistency of assessment of student teachers across universities, a matrix be developed to show how the various universities’ assessment scales — benchmarked on the pass/fail point in each case — map on to each other.

There is some comfort here for the critics of grading in that there is no mention of a distinction or any grade other than that assessing of minimal competence. The

comparability of this across ITE programmes is the most important goal. However the national profile (essentially a set of competences devised by the SEED and the Quality Assurance Agency for Scotland) are still there as the basis for the assessment.

In the flood of words which have been devoted to the topic of assessing placement there has been a curious lack of appeal to empirical evidence. What evidence is there that ability on placement is a skill discernibly different from the ability to write essays? In other words, can academic and professional attainment be distinguished and if so, is the difference great enough to warrant separate reporting? What would be the implications of changing from a method of calculating degree classes based on both academic and professional grades to one based only on the former (as would of course be the consequence of not grading placement)?

The empirical work which has been done in this area has mainly concentrated on looking at the consistency between different aspects of placement performance. Preece (1993), working in England, used a system based on 12 categories of performance and found that 10 of them correlated well enough with each other to form the basis of a uniform scale on which to assess placement. His scale had a Cronbach alpha coefficient of 0.94, suggesting that a very reliable scale can be constructed. But it is difficult to interpret this figure. All the ratings for each student were given by one university tutor after a placement visit and it is possible that all the grades could be affected by the overall impression which the tutor had formed. The ratings were not therefore independent in the way that Cronbach's alpha assumes. Nevertheless, scores of Preece's scale did tally fairly well with the schools' ratings of the same students, which is some evidence of the scale's predictive validity.

Brooker, *et al.* (1998) also proposed a system for assessing placement but although their article describes the system in detail, the empirical evaluation is limited to investigating whether those using it thought that it was effective or not. The responses they received were largely positive but the evidence is indirect - that users should think that a method is effective is obviously a good thing but it does not show directly that it actually fulfils its objectives. After all, many academics feel that the marking of essays and assignments is a valid and reliable way of measuring academic performance despite several decades of evidence showing the serious limitations inherent in any impressionistic assessment.

Sharp (1997) conducted a factor analysis of placement grades and essay grades for two undergraduate ITE programmes at a Scottish institute. The results showed that around 30% of the variance was explained by a general factor on which both types of assessments loaded quite strongly. In addition however there was a second factor accounting for between 10% and 15% of the variance which distinguished clearly between essay grades and placement, thus suggesting the presence of specifically placement-oriented skill.

DATA

The data to be reported in this paper consist of the third and fourth year grades obtained by students who graduated in 2005 and 2004 having completed a four-year Bachelor of Education degree at an ITE provider in Scotland. Two degree programmes were studied, in Primary Education and in Physical Education. Those graduating in Physical Education in 2003 were also included as cohort sizes are smaller in this programme and the inclusion of a third cohort enabled comparable total sample sizes. The work is being done now since, as suggested above, ITE providers may be moving in the near future to 2 or 3 point grade scales for all programmes. The cohorts graduating now may be amongst the last for which the following research questions can be pursued. These are:

- To what extent is it possible to identify 'professional' and 'academic' dimensions of student attainment on ITE programmes?

- Do essay grades and placement grades show that professional and academic courses are being assessed in the same way?

Students completing the BEd (Primary Education) programme received essay-based grades in year 3 for four courses – Education 3, Expressive Arts, Language and Mathematics. In year 4 they again had four essay-based grades – for Education 4A, Education 4B and two electives. The presence of the electives complicates things a little since by definition the courses will vary from one student to another. For present purposes, the two elective grades achieved by each student were averaged to form a composite grade which does not relate to any one course. Each student also had two placement grades in each year, one given by the school in which the placement occurred and the other given by the visiting university tutor. These overall grades are derived by combining grades given for individual aspects of placement which are in turn determined using the SEED competences for ITE (or ‘benchmarks’ as they are referred to in more recent documentation).

Students following the BEd (Physical Education) programme have a similar assessment pattern. In year 3, they received essay-based grades for three courses – Education 3, PE Curriculum 3 and PE Perspectives 3. In year 4 they had four essay-based grades – for PE Curriculum 4 (Higher Still), PE Curriculum 4 (Investigation), PE Perspectives 4 and Education 4B. Each student also had two placement grades in each year, one given by the school in which the placement occurred and the other given by the visiting university tutor. Although the nature of the placement is quite different on the two programmes, the holistic method of assessment based on the competences is similar.

RESULTS

To throw light on the first of the questions posed above, a principal components analysis was run on the grades on each programme. Only the results of students who completed the programme were included so there was no missing data. Mahalanobis’ M distances (explanations of some technical terms are given at the end of this paper) were calculated and cases where the M value was significant at the 0.1% level were deemed outliers and discarded (Tabachnik and Fidell, 2001: 68). This meant deleting 4 out of 211 cases on the Primary programme and 2 out of 250 on the Physical Education programme.

For the Primary data, 207 cases were included in the analysis. Two diagnostic statistics were calculated. The Kaiser-Meyer-Olkin measure of sampling adequacy had a value of 0.73, well above the minimum of 0.60 recommended by Tabachnik and Fidell (2001: 633). Bartlett’s test of sphericity was highly significant (chi-square = 837 with 55 degrees of freedom). These tests indicate that these data are suitable for principal components.

Three components with eigenvalues greater than one were identified. The eigenvalues of the first four components are given in table 1 and the loadings on each components (without rotation) of each of the courses is given in table 2. To ease interpretation, all loadings between +0.25 and -0.25 have been omitted from table 2.

The first component explains about one third of the variance. All courses have substantial loadings (at least +0.45) on it, indicating that it is the ‘general’ component which is a very common feature of multivariate assessment data. On average, students who do well in one course of the programme tend to do well in others. Whether this is due to overall ability, application, conscientiousness or some other quality is difficult to say without further exploration but for present purposes, we need only note that its presence is not surprising. Neither it is very informative.

Rather more informative is the second component which explains about half as much variance as the first - about one-sixth of the total. This distinguishes quite

Table 1: The first four components for the Primary Education data set

Component	Initial Eigenvalues	Percentage of Variance Explained
1	3.70	33.6
2	1.84	16.8
3	1.35	12.3
4	0.85	7.7

Table 2: Component loadings for the Primary Education data set

Component	1	2	3	4
Education 3	.59	-.26		
Expressive Arts 3	.70	-.34		
Language 3	.68	-.30		-.34
Mathematics 3	.53	-.27	.26	-.29
Education 4A	.63			.27
Education 4B	.55			.68
Year 4 electives - average	.74	-.25		
Placement (School) 3	.46	.67	.51	
Placement (University) 3	.45	.65	.54	
Placement (School) 4	.50	.48	-.61	
Placement (University) 4	.48	.54	-.58	

Table 3: The first four components for the Physical Education data set

Component	Initial Eigenvalues	Percentage of Variance Explained
1	3.75	34.0
2	2.12	19.3
3	1.09	9.9
4	0.90	8.2

clearly between essay grades, which have negative loadings, and placement grades where loadings are positive. Evidently there is an identifiable placement dimension to student performance. Placement grades correlate positively with essay grades but they correlate even more positively with other placement grades. Similarly, essay grades correlate positively with placement grades but they correlate even more positively with other essay grades.

The third component with an eigenvalue greater than one distinguishes between third and fourth year achievement, but only as regards placement grades. All essay-assessed courses have low loadings on this component, indicating that year of study is not an important consideration for these. Put simply, a student who is good at essays in year 3 will be good at essays in year 4. For placement however there is a progress dimension with some students achieving better placement grades in year 4 than would be expected on the basis of their year 3 grades while others do not do as well as one would expect given their year 3 placement grades.

The fourth component is included for completeness but with an eigenvalue of less than one it is below the criterion for inclusion suggested by Kaiser (1958). In so far as it can be interpreted, it seems to distinguish between curriculum-specific essay grades and others (Education 4A is the final year dissertation while Education 4B is a conference on current developments in education). However it is not a major feature of the data.

For the physical education data, 248 cases were included. The Kaiser-Meyer-Olkin measure of sampling adequacy had a value of 0.71, and Bartlett's test of sphericity was again highly significant (chi-square = 1427 with 55 degrees of freedom). Again the data set appears to be suitable for principal components.

As with the Primary data, three components with eigenvalues greater than one were found. Tables 3 and 4 give the results of the analysis in the same way as tables 1 and 2 did for the Primary Education data.

Table 4: Component loadings for the Physical Education data set

Component	1	2	3	4
Education 3	.49	.38	-.32	-.32
PE Curriculum 3	.53	.41		-.30
PE Perspectives 3	.54	.48		
Education 4B	.34		.49	.65
PE Curriculum 4 – Higher Still	.57	.27		
PE Curriculum 4 – Investigation	.63	.38		
PE Perspectives 4	.63	.41		
Placement (School) 3	.64	-.53	.42	
Placement (University) 3	.64	-.54	.42	
Placement (School) 4	.66	-.53	-.42	
Placement (University) 4	.67	-.53	-.42	

The size of the first four components, and their interpretation, are strikingly similar to those of the Primary Education data. Again the first component explains around one third of the total variance and produces positive loadings on all types of courses – it is the general component again. The second component explains slightly more of the variance (about one-fifth) than was the case with the Primary data but its nature is the

same. Loadings on one sign are associated with essay grades while those of the other correspond to placement grades. The third component again distinguishes between third and fourth year placement grades but here the interpretation is not so clear-cut. The two essay-based courses not specific to the PE curriculum (the Education 3 and Education 4B courses are common across ITE programmes) make an unexpected appearance with the signs reversed (the placement grades have negative signs for the third year and positive signs for fourth year while the two Education courses are the other way round). It is far from obvious why this should be the case. The fourth component is again included for completeness and again has an eigenvalue of less than one. As with the Primary Education data, Education 4B, the report of a conference, loads more strongly on this than on any other component but the two negative loadings are for Education 3 and PE Curriculum 3, courses which do not have anything obvious in common.

The second question posed above concerned whether essay grades and placement grades show that professional and academic courses are being assessed in the same way ie. are the grade distributions of these two types of assessments for the same students comparable? Table 5 gives the percentages of all grades which were A, A or B, and E for essay grades and placement grades for each of year 3 and year 4 for the two programmes of study.

Table 5: Percentages of grades awarded for the Primary and Physical Education data sets

	BEd Primary Grade			BEd Phys Ed Grade		
	A	A or B	E	A	A or B	E
Year 3 essays	22.1	55.2	1.9	6.0	32.4	5.5
Year 4 essays	13.2	44.7	1.4	11.7	41.9	2.6
Year 3 placement	30.5	83.2	3.2	29.0	75.9	1.2
Year 4 placement	63.5	92.1	1.0	45.8	79.2	1.6

There are substantial differences between essay grades and placement grades and between the programmes. For the essay grades, the proportions of top grades awarded falls from year 3 to year 4 in the Primary programme but increases in the PE programme. The percentage of Es (ie fails) is fairly constant for the Primary programme but falls slightly for the PE students. Similarly, failure rates for placement are around 3% or less for both years and programmes, falling slightly for the Primary programme and remaining fairly constant for the PE programme. The biggest difference however is between these and the top end of the placement rating scale. For both programmes, many more As and Bs are awarded than are awarded for the same students' essays – in fact between 75% and 90% of all placement grades are A or B. This suggests that assessors are using the A to E scale in very different ways for essays and for placement. For essays, As are awarded to a minority of students (though the size of the minority varies considerably) while A or B is awarded to between one third and one half of the cohort. For placement however the A or B assessment covers the majority of the cohort, the proportion ranging from around three quarters to over ninety per cent.

Having looked at what the data presented here have to say about whether it is possible to identify 'professional' and 'academic' dimensions of student attainment on ITE programmes, and whether professional and academic courses are being assessed in the same way, it is now possible to make some more general observations

about the assessment of placement and what the effect might be of basing degree classifications solely on essay grades.

DISCUSSION

It is necessary first to be clear about what the present data do not say. They do not say anything about the comparability of assessment between institutions since both the programmes were offered by the same institution. Neither do they throw any light on the validity of the grades. As is usual with assessment data, it is never possible to be sure quite what the grades are measuring. It may be for example that the placement grades were affected by personal characteristics of the students rather than their effectiveness on placement and that the essay grades were influenced by the students' punctuation and spelling rather than their understanding of educational ideas. Institutes offering ITE can and do implement quality assurance procedures such as assessment criteria, crossmarking, assessor training and external examiners. But beyond that it is necessary to trust the professional competence of the assessors.

Tables 1 to 4 do however say something about the consistency of the grades. In both programmes there is a 'placement attainment' dimension which appears to be at least as stable and identifiable as the 'essay attainment' dimension with which it contrasts in the second factor. This component structure is more generalisable than to just these two programmes at this time. Sharp (1997) conducted the same analysis on essay and placement grades from the same two programmes ten years earlier (students graduating in 1995 and 1996) with strikingly similar results, down to the third component distinguishing between third and fourth year grades for placement but not for essays. In the ten years between the two data sets, the institution had changed from being an independent institute to being part of a larger university. The curricular structure of both programmes had been rewritten entirely to fit the new higher education context within which they were located. As an example of this, within the Primary programme, the number of placement grades in the final two years of study had been reduced from seven to four and the number of essay grades from 14 to eight. Assessment in 2004/5 consisted of fewer, larger courses than previously but the factorial structure was effectively identical.

However principal components analysis of the type reported here is based on correlations between grades. A high level of correlation does not imply that the grades constitute equivalent variables having the same distributions. In fact table 5 shows that there are substantial differences between the programmes in this regard. As regards placement (which is assessed by quite different tutor teams on the two programmes), while the third year grades were fairly comparable, in fourth year nearly 20% more Primary students were awarded grade A than PE students and over 10% more were awarded grade A or B. This may reflect differences in awarding standards, or differences between the two cohorts, or more probably some combination of these. The present data do not allow the two influences to be disentangled.

For essay grades, there were again differences between the programmes (rather larger differences than was the case for placement grades). For the PE students, the proportions of A and A/B grades increase from year 3 to year 4 while for the Primary students these proportions fall, an apparently counterintuitive trend since presumably the students' abilities were developing throughout their studies. It appears that the expectations of their tutors were also increasing, and at a faster rate. Whatever the explanation, there are certainly differences between the distributions of essay grades on the two programmes. Here however it is possible to say something about how far this was due to differences between the students and how far to differences between grading standards. Two essay-assessed courses, Education 3 and Education 4B, are taken by students on both programmes. As marking is done anonymously, the grades could not have been affected by which programme a given student was following. Comparison of the two distributions shows that on Education 3, 10% of Primary

students achieved grade A, and 45% A or B, as opposed to 5% and 33% for the PE students. On Education 4B, the proportions were 17% and 55% as opposed to 13% and 46%. On both these common courses therefore, the Primary students do slightly better. The difference however is not great (about one fifth of a grade in each case) and is less than the average of the grade differences between the programmes for essays in general. While some of this can be attributed to differences between the student cohorts, at least some appears to reflect differences in awarding standards.

It remains to ask what would be the effect on degree classifications of basing these decisions wholly on essay grades. The answer to this depends on how the grades are combined. If this is done by a simple pooling method whereby both types of grades are pooled and a single decision rule applied, the result of removing placement grades without amending the decision rule would be to make first and upper second class degrees much harder to achieve. Put another way, placement grades are relatively inflated and it could be argued that it is not appropriate to include them in the first place. In fact, for the present data set, different decision rules were applied to placement and essay grades and to be awarded a given class of degree, each student had to achieve at least that class on both rules. Given the relatively inflated nature of placement grades, most students had a higher 'placement' degree class than an 'essay' degree class. For the Primary Education cohorts for example, the 'placement' class was the higher by two classes for 16 of the 211 students, the higher by one class for 102, the same for 74, the lower by one class for 16 and the lower by two classes for just three students. So the vast majority of final degree classes were decided by essay grades. On the basis of the present data, the removal of placement grades would in most cases make no difference.

CONCLUSION

What conclusions then can be drawn from the analyses reported above as regards the assessment of placement and its proper role in the assessment of ITE? The criticisms which have been made of the use of competences as ways of formalising the ingredients of professionalism in teaching, and of using them in the assessment of placement, undoubtedly have some validity. There are indeed grounds to suppose that placement will be a particularly difficult aspect of developing professionalism to assess since it is context-dependent in a way that the assessment of more traditional academic abilities is not. It is also possible that success in placement calls upon a wider range of students' intellectual abilities, personal skills, experience and maturity than do essays and assignments.

But the fact that a job may be difficult is not a reason for not attempting it. It is however a reason for taking every precaution possible to assist the effectiveness of the assessment procedure. The critics of the competence movement have argued that the underlying motivation in introducing competences was to assert government control over the teaching profession and to reduce teachers to ciphers carrying out government directives. But twelve years after the introduction of competences, recent research on teacher satisfaction (Draper and Sharp, 2006, Sharp and Draper, 2005) indicates that professional autonomy is one of the aspects of teaching about which teachers express highest, not lowest, job satisfaction. Possibly, on this occasion at least, the official reason was the real one - to try to improve the consistency of assessment in an area where consistency is especially difficult to achieve and to highlight the importance of the mutually reinforcing link between theory and practice.

The evidence presented above suggests that complete consistency had not been achieved, even between two programmes within a single institution. But neither had total consistency of essay gradings been achieved. In fact there are more parallels between the two types of assessment than might at first be apparent. It is now accepted as good academic practice that the criteria on which essays and assignments will be assessed should be agreed between the assessors and made available to the

students before they begin their writing. It is difficult to argue that even where this good practice is adopted, the assessment of essays is beyond criticism. Wherever the professional judgement of the assessor is called upon, there will be an element of subjectivity (and hence unreliability) in the result, an element whose existence has been demonstrated in many research projects (for example Hayes, Hatch and Silk, 2000; Gentile, 2000; Read, *et al.*, 2005). True objectivity is only really achieved in the multiple choice tests so derided by the critics of the 'scientific' model of assessment (and even here it is only the marking which is objective - the choice of items and distractors is very much a skill learned by experience).

The criticisms of the competence model, and of the use of competences in the assessment of placement, may be justified if the competences are seen as a complete account of professionalism and if they are seen as a method on their own of serving as the basis of assessment. But they are clearly neither of these – the experience, intuition and judgement of the assessor are key to the validity of the process, just as they are in the grading of essays and assignments. Competences are merely aids in a larger process. To argue convincingly that the fallibility of assessing placement is a reason for abandoning it, it would be necessary to show that the assessment of placement was fallible in a way that the assessment of essays is not. The evidence presented above and in the research conducted on essay marking does not suggest that this is the case. If the assessment of placement is of such unreliability that it should be discarded, the same conclusion should be drawn for essays also. But the critics of competences have not argued for the dropping of all summative assessment on ITE courses.

Had the arguments of the critics been supported by evidence, the implication would have been that all the degrees held by students where placement grades had contributed to the degree class were of suspect validity. Fortunately, the evidence reported here does not support such a pessimistic conclusion. There is a placement dimension and placement grades measure it as well as essay grades measure the academic dimension.

Having said this, it appears likely that grading will in the future indeed be phased out as a contributor to degree classifications. The reasons for this however are very different from those evinced in the literature on this subject. One of the drivers is the transfer of responsibility for all aspects of placement from ITE providers to the schools in which they will work. This in turn is driven by the closer relationships which students have with schools as they move from being students towards being probationer teachers, one of the outcomes of the McCrone reforms (SEED, 2000). Another driver is the financial and personnel costs of having university-based tutors spend so much time and effort in travelling to schools in an era of intense pressure on the financing of higher education. But the phasing out of the grading of placement does not signify the acceptance by government of the arguments put forward by the critics of competences. Had this been the case, the competences themselves would have been phased out. In fact they will remain the basis of the assessment of placement whether it is graded or not and whether it is undertaken by schools only, or a combination of schools and ITE providers.

Will the phasing out of placement grading matter? Not much as far as degree classes are concerned, on the basis of the data reported here. The real danger lies in the potential devaluation in the eyes of the students of the school-based parts of their programmes compared to the university-based parts. But if professional attainment and academic attainment are different then they should not be reported together using a linear scale of degree classes. Perhaps the best compromise would be to report the two types of attainment separately, academic attainment using the degree classes which (rightly or wrongly) have always been used to report such attainment, and professional attainment based on a three-point scale of satisfactory-unsatisfactory-distinction. The distinction would not be equivalent to the grade

A for placement but would be awarded to a percentage of students (similar to the percentage currently who fail) whose work in schools is outstanding. Those who propose assessing placement using a scale of satisfactory-unsatisfactory could not object to this three point scale on technical grounds since the upper benchmark is the mirror image of the benchmark they already accept. The award of the degree would imply that the lower benchmark had been met and the meeting of the upper benchmark would be reported as for example 'Upper second class with distinction in teaching practice'. This would retain for students the incentive to excel on placement regardless of level of their essay grades, would be consistent with the ITE guidelines and would be consistent with research evidence in reflecting the two-dimensional nature of achievement on ITE courses.

NOTE

Mahalanobis distances are a measure of how far the values of the independent variables for a given 'case' (which means 'student' in this paper) differ from the mean values over all cases. A large Mahalanobis distance identifies a case as being an outlier and hence unsuitable for inclusion in a factor analysis. The Kaiser-Meyer-Olkin measure of sampling adequacy tests whether the partial correlations among variables are small. Small values denote 'diffuse' patterns of correlation while large values indicate 'compact' patterns. The latter are preferable for factor analysis. Bartlett's test of sphericity tests for the existence of a factor structure. If the result is not significant, no structure is detected and factor analysis is clearly not appropriate. Eigenvalues convey the amount of the variance in the original data which is associated with a particular factor. Factor loadings indicate the strength and direction of the relationships between the raw variables entered into a factor analysis and the latent variables which the analysis identifies.

ACKNOWLEDGEMENT

The author would like to acknowledge the assistance given by Pamela Munn and Dave Thomson on an earlier draft of this paper. The views expressed in it are however those of the author and do not necessarily represent those of any other individual or organisation.

REFERENCES

- Brooker, R., Muller, R., Mylonas, A. and Hansford, B. (1998) Improving the assessment of practice teaching: a criteria and standards framework, *Assessment and Evaluation in Higher Education*, 23(1): 5–24.
- Carr, D. (1993) Guidelines for teacher training: the competency model, *Scottish Educational Review*, 25(1): 17–25.
- Cope, P., Bruce, A., McNally, J. and Wilson, G. (2003) Grading the practice of teaching: an unholy union of incompatibles, *Assessment and Evaluation in Higher Education*, 28(6): 673–684.
- Draper, J. and Sharp, S. (2006) Continuity or change? The initial impact of the post-McCrone agreement, *Scottish Educational Review*, 37 (Special Edition): 100–113.
- Elton, L. (1969) *Assessment of Undergraduate Performance*, London: Committee of Vice Chancellors and Principals/ Association of University Teachers.
- Elton, L. (2004) Should classification of the UK honours degree have a future? *Assessment and Evaluation in Higher Education*, 29(4): 415–422.
- Gentile, J.R. (2000) An exercise in unreliability, *Teaching of Psychology*, 27(3): 210–212.
- Hager, P. and Butler, J. (1996) Two models of educational assessment, *Assessment and Evaluation in Higher Education*, 21(4): 367–378.
- Hayes, J.R., Hatch, J.A. and Silk, C.M. (2000) Does holistic assessment predict writing performance? Estimating the consistency of student performance on holistically scored writing assignments, *Written Communication*, 17(1): 3–26.
- Humes, W. (2001) Conditions for professional development, *Scottish Educational Review*, 33 (1): 6–17.
- Jones, L. and Moore, R. (1993) Education, competence and the control of expertise, *British Journal of Sociology of Education*, 14(4): 385–398.
- Kaiser, H. (1958) The varimax criterion for analytic rotation in factor analysis, *Psychometrika*, 23, 187–200.
- Martin, S. (1997) Two models of educational assessment: a response from initial teacher education: if the cap fits..., *Assessment and Evaluation in Higher Education*, 22(3): 337–343.

- Martin, S. and Cloke, C. (2000) Standards for the award of qualified teacher status: reflections on assessment implications, *Assessment and Evaluation in Higher Education*, 25(2): 183–190.
- Preece, P. (1993) The assessment of teaching practice performance, *Research in Education*, 49, 23–27.
- Read, B., Francis, B. and Robson, J. (2005) Gender, “bias”, assessment and feedback: analyzing the written assessment of undergraduate history essays, *Assessment and Evaluation in Higher Education*, 30(3): 241–260.
- Roberts, T. (2004) Defining the concept of the minimally qualified or borderline student. Paper presented to the annual conference of the Society for Research into Higher Education, University of Bristol, December 2004.
- Scottish Executive Education Department (SEED) (2000) *A Teaching Profession for the 21st Century* (‘The McCrone Report’), Edinburgh: HMSO.
- Scottish Executive Education Department (2001) Report of the ‘First Stage’ Review of Initial Teacher Education. Available: www.scotland.gov.uk/library3/education/tefs-00.asp (accessed 13 December 2005).
- Scottish Office Education and Industry Department (1998) Guidelines for Initial Teacher Education Courses in Scotland. Available: www.scotland.gov.uk/library/documents-w3/git-01.htm#REQUIREMENTS (accessed 13 December 2005).
- Sharp, S. (1997) A factorial study of student performance in initial teacher education, *Studies in Higher Education*, 22(3): 363–378.
- Sharp, S. and Draper, J. (2005) Teachers’ Job Satisfaction: its nature and change in a longitudinal study, Paper presented to the annual conference of International Conference on the Education of Teachers, University of Pretoria, July 2005.
- Stronach, I., Cope, P., Inglis, B. and McNally, J. (1994) The SOED competence guidelines for initial teacher training: issues of control, performance and relevance, *Scottish Educational Review*, 26, 118–133.
- Tabachnik, B. and Fidell, L. (2001) *Using Multivariate Statistics* (4th edition), Boston: Allyn and Bacon.
- Universities UK (2004) Measuring and recording student achievement: Report of the Scoping Group chaired by Professor Robert Burgess. Available: bookshop.universitiesuk.ac.uk/downloads/measuringachievement.pdf (accessed 18 January 2006).