Psychology Teaching Review (2002) **10**, 1

Pedagogic research and evidence-based management

Roger Lindsay, Rosanna Breen & **Reneé Paton-Saltzberg**,
Psychology Department, Oxford Brookes University.

Abstract

The introduction of modular schemes and a semesterised academic year are amongst the most fundamental changes ever to occur in UK higher education. There is, however, a notable lack of pedagogic research evidence on the effects upon student learning of course frameworks and the temporal structure of large-scale learning units. In the absence of other evidence, decisions to adopt semesterisation and modular schemes seem to have been made to save money, to increase university income, or simply to follow a management trend. A small study of attitudes to semesterisation across psychology departments is reported that raises questions about the pedagogic benefits of shifts to semesterisation. A larger-scale study carried out within a well-established modular course suggests that some aspects of freedom of choice, such as choice over the number of modules studied concurrently, can have negative effects upon student achievement. It is suggested that these effects of high and fluctuating workload should be controlled through course design and through advice given to students when they are constructing programmes. Attention is drawn to the need for psychologists to contribute to the pedagogic research knowledge base in this domain.

Policy and evidence: Some examples

Elsewhere in the present issue of *Psychology Teaching Review,* Lindsay, Breen and Jenkins (2002) have presented a rationale for carrying out policy-oriented research into issues and processes that the investigators are powerless to affect directly. The argument is that generating evidence that rules out some actions and makes others more probable (or at least more rational), is itself a type of intervention to the extent that it changes the actions of agents who do have the power to act effectively. In addition it is argued that management decision-making in higher education is not sufficiently grounded in the use of evidence, and that a move to *evidence-based management* is urgently needed. In part, the rarity of evidence-based decision-making seems to result from the fact that high quality professional practice has yet to develop in this domain, but it is also fostered by the paucity of useable evidence. So long as pedagogic evidence is lacking, managers can only

make decisions on the basis of evidence to hand – often this will be evidence about costs and efficiency. Academics, including researchers, are often critical of managers for pre-occupation with the 'bottom line', but pedagogic researchers must bear some responsibility for failing to make other kinds of evidence available. More positively, however, the evidence vacuum also presents various kinds of intervention opportunity. Criticising managerial decisions that are based on financial considerations without alternative evidence of any kind is an exercise in futility. But generating evidence that favours some definite line of action, where presently there is no evidence at all, can easily create a presumption in favour of the advocated policy.

Modularisation of courses and *semesterisation* of the academic year have been amongst the most momentous changes that have affected the UK university sector during the last 20 years. The two are not entirely independent: modularisation seeks to promote flexibility and choice, and achieves efficient course delivery through centralised administration and economies of scale. However, modules must be separately assessed and a three-term year also creates dis-economies because of the length of the year and consequent assessment burden. Whilst assessing only twice-yearly within a semester system may seem seductively efficient, a non-modular course only needs to assess at the end of years 1 and 3.

The sources of evidence that underpin the decision to modularise/semesterise are such things as market analysis (attractiveness to applicants), comparative delivery costs and collateral benefits such as residential conference income during the weeks by which the semestered year is shortened. Marketing data are often fragmentary and impressionistic because they are so difficult to collect and analyse. As semesters permit a shorter academic year, with less frequent assessment, and more scope for alternative income generation, there are clear reasons why semesters might be seen as preferable to terms. During the UK debate, little or no account appears to have been taken of the impact of the course delivery framework upon such things as the effectiveness of teaching, the quality of learning, or student achievement. Partly this is a failure on the part of universities, partly a failure of pedagogic researchers to collect relevant evidence. Most surprisingly, there appears to be little or no cross-talk between universities that have already opted for modularisation or semesters, and those that have yet to decide. One small-scale study (Fearnley, 2001) from Oxford Brookes suggests that collecting evidence of this kind can be worthwhile.

Survey of semesters via psychology departments

An e-mail questionnaire was distributed to 103 UK Psychology Departments (via the Association of Heads of Psychology Departments) to solicit views on semesters. Thirty responses (29 per cent) were received. It appeared that responding departments generally preferred terms to semesters. The main reasons for preferring terms to semesters were: pedagogic benefits of short learning/assessment units; terms are less likely to interfere with conferences; terms follow 'natural' breaks in the calendar year and so vacations are more convenient for staff and students with children.

Departments were asked whether they *now* taught using semesters or terms, which they preferred, and what the main reasons were for their preference.

All staff teaching in a semester system had apparently also taught in a term-based system at some point. Testing against the null hypothesis that there is a uniform distribution of responses between the two preferences, application of a chi-square test to the data show a statistically significant preference for Terms ($\chi^2 = 0.98$; df = 1, $p = 0.007$) and no evidence of an association between current practice and preference.

Table 1. Preferences for semesters or terms reported by 30 Departments of Psychology in response to an e-mail questionnaire.

	Now teaching:			**Preference**
Prefer:	Semester	Semester + Term	Terms	**Totals:**
Semester	4	1	0	5
No Preference	5	2	0	7
Terms	10	4	4	18
Totals:	**19**	**7**	**4**	**30**

Whilst only one respondent expressing a preference for semesters indicated that they felt strongly about it, five of the 18 respondents who preferred terms, made it clear that their preference was a very strong one. One department reported that their university had already reverted to terms from semesters and another department reported that a similar reversion was currently under debate within the university. Admittedly, these data come from a single discipline, and a small sample. It does however suffice to show that relevant evidence is easily collected, and perhaps to raise a doubt as to whether the current fashion for semesterisation is rationally grounded.

Modular Ecology

It is a common management assumption that course frameworks are neutral vehicles via which knowledge and skills are just more or less efficiently delivered. Lindsay (1998) has argued that contrary to this view, course frameworks resemble 'ecological niches' shaping the behaviour of the students operating within them. A commonly noted example of this is the instrumental culture that tends to develop amongst students once coursework contributes to assessment (an arrangement not unique to modular courses, but strongly associated with them). The resulting learning environment resembles a *token economy* (Ayllon & Azrin, 1968) within which desirable behaviours are rewarded with marks. Unfortunately many students learn, not only to produce the behaviours that attract marks such as essay and report writing, they also learn *not* to produce those behaviours for which marks are not awarded, such as general reading. This is not a fatal flaw in course design, but it does mean that course providers have to take much responsibility for assuring that all essential knowledge, every key skill and each important book is explicitly related to some assessment element. This creates a powerful impulsion towards a prescriptive academic bureaucracy. Even small differences in academic regulations can have dramatic effects on performance. As an illustration, Lindsay and Paton-Saltzberg's data (Lindsay, 1998) shows a failure rate in the third term of a modular course more than twice as high as the failure rate in term 2, and more than three times as high as the failure rate in term 1, even though the average mark for modules passed was roughly constant across terms. The explanation in this case appeared to be a regulation that required students to take 12 modules, but to pass only ten. By term 3, students were able to identify some modules as surplus to requirements and allocate effort accordingly.

A systematic comparison of modular with non-modular programmes, or terms with semesters is impossible whilst

respecting student freedom of choice. It is possible however to begin to understand the effects of some variables by examining variations in performance within a particular course scheme. Modular schemes are designed to offer flexibility and choice, including choice over the number of modules taken simultaneously, and choice over when a particular module is taken. It might, therefore, seem pertinent to ask how marks are affected by the number of modules concurrently studied, and how performance on a particular module is affected by the point in a student's programme at which it occurs. These issues are important to student advisors because in course credit systems a light term or a module failed now, means additional modules in some future term; choices must also be made about whether to take most- or least-favoured modules in year two, or year three. At a more strategic level, a reduction in the number of assessment periods, as through semesterisation, may also mean a greater number of assessments in any one period. Students on Oxford Brookes' Modular Programme are free to vary the number of modules they take in a particular term, and as a consequence it is possible to compare the average mark they get for each module, according to the number of concurrent assessments. The study described below examined assessment data from 1074 students graduating in the same year from Oxford Brookes' modular programme. The data set was quite complex, as assessment data maintained within the modular management database had to be first transformed by special purpose programs, then further transformed within the EXCEL spreadsheet package. The resulting flat rectangular database consisting of 330 variables was then transferred to SPSS for full analysis. The data were used to evaluate the effects of (i) *module load*, (ii) *effects of basing degree class on a subset of modules taken*, and (iii) *progression*.

Effect of module load on performance

Two mutually incompatible beliefs exist among teaching staff concerning the relationship between number of modules taken and expected performance. One view is that the greater the module load undertaken, the worse performance will be. This is a distributed capacity theory that assumes that a fixed quantity of cognitive resources is more thinly spread as module load increases. The alternative view is that taking more modules will improve performance because of practice, transfer effects and generalisation of knowledge or skills. Load is probably a more complex variable than is recognised by either of these beliefs. Lindsay and Gibbs (1986) reported a questionnaire study which showed that while high *within-module* load was universally perceived by students as damaging to their performance, there was also a negative effect of *underload*, attributed to sparsity of feedback, and uncertainty about goals and standards. Lindsay and Paton-Saltzberg's (1993) demonstration of a decline in student marks associated with paid employment during termtime is also a reminder that factors outside the academic environment contribute to load.

The study of load reported below was restricted to load factors associated with the number of modules studied and assessed simultaneously. Even given this simplification, there are still at least three aspects of load that might affect performance: *total load* (indexed here by the number of modules taken in the six terms prior to graduation), *within-term load* (indexed by the number of modules taken in any particular term), and *between-term variability* (operationalised here as the average difference between within-term loads over the six terms prior to graduation). To illustrate the difference, row 1 of Figure 1 overleaf shows a programme with *low total load, moderate within-term load* and *low between-term variability*. In row 2, both *total load* and *within-term-load* increase, but

Figure 1. Hypothetical modular programmes illustrating different load patterns.

Term 1	Term 2	Term 3	Term 4	Term 5	Term 6	Total
3	3	3	3	3	3	18
4	4	4	4	4	4	24
5	1	5	1	5	1	18

between-term variability remains low. In the bottom row, *total load* is low, *within-term load* is high and low in alternate terms, and *between-term variability* is high.

Effect of total load on performance

Table 2 shows the relationship between total module load and *overall average*, and module load and *degree average* (programme regulations required this to be calculated over the best 18 modules, but more modules could be taken). Table 2 and Figure 2 show that as the total module load increases, the overall average mark (*all* modules taken) decreases from 61.1 when load = 18, to 49.5 when load exceeds 24.

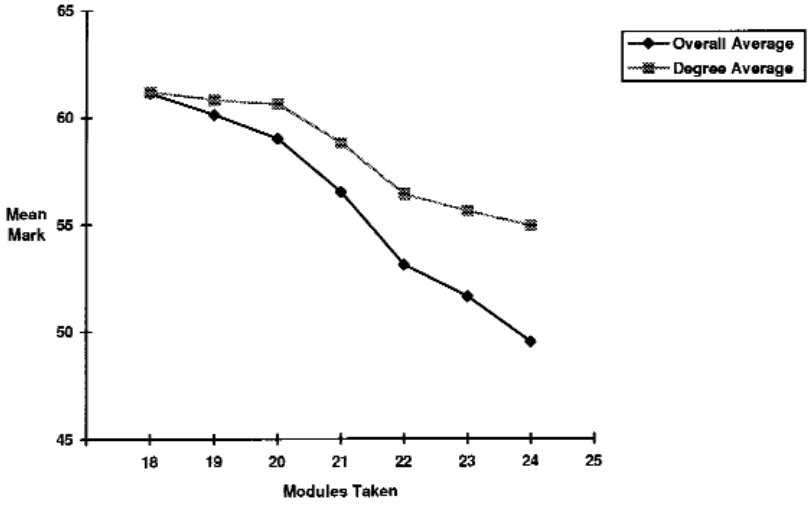
Table 2 and Figure 2 also illustrate the effect of increasing total load on degree average. Both overall average and degree average get worse as total load increases, but degree average is affected less severely. This represents the beneficial effect of being allowed to drop some modules from the calculation. Mean marks are associated with cases, not modules, (so students taking 20 modules do worse than those taking 19, etc., rather than students taking their 20th module do worse than they did on their 19th). The increasing discrepancy between degree average and overall average suggests that students taking many extra modules are doing so because they are failing or gaining low marks. Data presented in Table 2 does not necessarily imply that taking more modules impairs performance. Correlations do not specify a causal direction, and it is also possible that students with lower averages take more modules in an attempt to graduate, or to raise their average mark.

Because of the large variation in sample size between students taking different total

Table 2. Degree average compared with overall average for students taking 18 or more modules.

		Overall Average		Degree Average	
MODULES TAKEN	**N**	**SD**	**MARK**	**MARK**	**SD**
18	236	5.6	61.1	61.2	5.5
19	303	5.8	60.1	60.8	5.7
20	194	5.0	59.0	60.6	4.8
21	83	5.3	56.5	58.8	4.9
22	28	6.5	53.1	56.4	6.3
23	15	6.2	51.6	55.6	6.3
24–33	10	7.5	49.5	54.9	6. 7

Figure 2. Degree average compared with overall average for students taking 18 or more modules



module loads (e.g. 303 students take 19 modules but only 28 students take a load of 22) statistical comparisons cannot be made across the full range of loads. To enable such comparisons to be made, module load was collapsed into three categories as reported in Table 3.

An analysis of variance showed that there is a significant difference between the overall average mark (all modules taken) achieved by the three groups <18, 18, and >18 ($F = 22.81$; $df = 2$; $p < 0.001$). A post-hoc test confirmed that the significant contrasts ($p = 0.05$) are those between the >18 group and the other two groups. Table 3 shows that students taking more than 18 modules have a lower average mark (58.6) than students taking exactly 18 (61.1) or less than 18 (61.2). The same pattern was observed for the degree average (best 18 modules) comparisons ($F = 9.84$; $df = 2$; $p < 0.0001$). Again, post-hoc tests showed that the >18 group is associated with a lower mean mark than the <18 and =18 groups.

When analysis was restricted to students whose record contained no module marks lower than 40 per cent (the no-fail group), there was no difference between the average mark of students taking more than 18 modules, and those taking exactly 18, regardless of whether overall average or degree average was used as the dependent

Table 3. Degree average compared with overall average for students taking less than 18, exactly 18 and more than 18 modules.

NO. OF MODULES	OVERALL AVERAGE	DEGREE AVERAGE
Less than 18 modules (<18)	61.2	62.0
Exactly 18 modules (=18)	61.1	61.2
More than 18 modules (>18)	58.6	60.1

variable. A two-way analysis of variance using overall average as the dependent variable showed that while there was a significant effect of fail/no-fail ($F = 84.9$; $df = 1$; $p < 0.001$) and a significant effect of </=/>18 ($F = 4.73$; $df = 2$; $p = 0.009$) there was no interaction between these factors.

In summary, it seems that the facility to take more than 18 modules is used by weaker students to bring them closer to average performance, rather than by strong students seeking to maximise their advantage. The benefits of the 'best n' rule are apparently greatest for a very small number of students who take substantially more modules than most, but whose achievement is poorest.

Effects of within-term load on performance

Though increases in total load seem to be associated with poorer performance, it remains open whether this is because of selection effects (poorer students take more modules) or causal processes (taking more modules impairs performance). Data relating to within-term load bears more closely upon this issue for two reasons. Firstly, within-term load, rather than total load, is likely to be the variable via which any causal effects of load are mediated. A hard-pressed student really does have to reconcile the competing pressures of five modules taken in the same term, whereas total load is less directly related to the experience of studying. Secondly, as row 1 of Figure 1 illustrates, it is possible to accumulate a relatively high total load (24) without ever taking more than 4 modules in any single term. Three-hundred-and-fifty-one students in the sample took a within-term load of four on at least one occasion, and a within-term load of four was the load most frequently taken on two out of six terms surveyed. A load of four is thus a relatively standard load.

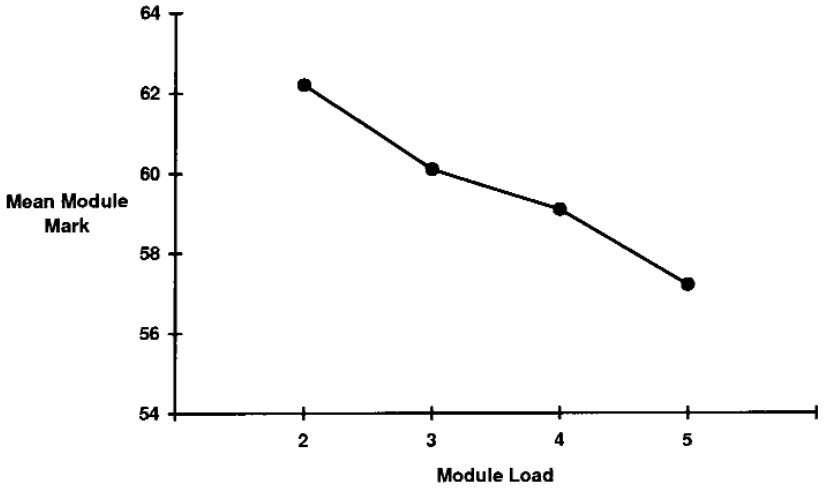
Table 4 shows the relationship between within-term module load and academic performance. The average mark for module load one means the average over all the modules taken by a student in which exactly one module was taken in a term, the average for load two means the average over terms in which two modules were taken and so on. The values for module load one presented in Table 4 should be treated with caution because approximately 70 per cent of the load one sample are part time students, and load six should be treated with caution because of the small sample size ($n = 11$).

Figure 3 illustrates the relationship between module load and mean mark when loads one and six are excluded. The marks are means per student, and are generalisations over a host of other relevant variables such as discipline, module size, etc. Examining such variables would require a much larger sample size. A repeated measures analysis of variance showed that

Table 4. Average marks gained by students taking within-term module loads of 1–6.

Module Load	1	2	3	4	5	6
Average Mark	61.5	62.2	60.1	59.1	57.2	57.5
Standard Deviation	7.9	8.5	6.7	6.6	7.3	9.6
N	99	239	900	727	124	11

Figure 3. Average marks gained by students taking within-term loads of 2–5 modules.



the average marks associated with loads two, three and four are significantly different ($F = 4.66$; $df = 2$; $p = 0.01$). A linear regression analysis showed that the mean mark is 63.9 and the slope coefficient is –1.2 [expected mean mark with load n is thus 63.9-1.2n. Crudely, 1.2 per cent is lost from 63.9 for every module taken above two modules]. The slope coefficient is significantly different from zero ($F = 54.75$; $df = 1$, 2098; $p<0001$). The regression analysis gave a slightly higher intercept value, and a slightly steeper slope when load 1 was excluded (intercept = 64.7; slope = –1.5), but remained stable when the analysis was confined to full-time students (intercept = 64.5; slope = -1.5). Though the loss of 1.2 - 1.5 per cent may seem trivial, it should be recalled that the assessment scale actually used in assessing degree-level work does not in practice, extend much beyond 70 per cent. The percentage lost should, therefore, be seen as a proportion of the scale subset within which most of the discrimination between candidates occurs, i.e. 55–70 per cent. In these terms, the mark loss incurred may be as much as 10 per cent of the effective scale.

Effects of load variability on performance

The standard deviation of term-by-term load was calculated as a measure of load variability. The range of the standard deviation was subdivided into five categories using the standard error of the load variation as an interval. Average marks were then compared across the five load variability categories and the data underlying this comparison is presented in Table 5. The analysis is made more complicated because students on Oxford Brookes' Modular Programme may be taking two *single* fields or one *double* field. A double field is roughly equivalent to a single honours degree elsewhere, and two single fields corresponds to joint honours in two subjects. An analysis of variance showed that there were no effects of load variability over the sample as a whole when overall average, degree average, or field one average were used as dependent variables. When field two average which excludes double field students was used, a significant effect was observed ($F = 3.14$; $df = 4$; $p = 0.01$). Post-hoc tests showed that low variability programmes tended to yield higher average

Table 5. Average marks for students with different levels of within-term load variability in their programmes.

VARIABILITY LEVEL (1 = Low, 5 = High)	FIELD 2 AVERAGE
1	62.8
2	59.8
3	58.2
4	59.1
5	59.3

field two marks than high variability programmes. When the sample was restricted to students with no module failures in their programmes all effects of differences in variability were lost.

It seems that there is an association between load variability and performance among single field students, but that the association is confined to single field students who have failed modules. The performance of struggling students taking two different disciplines appears to be worse when there is more variation in their termly load than when there is less.

Progression and 'value added'

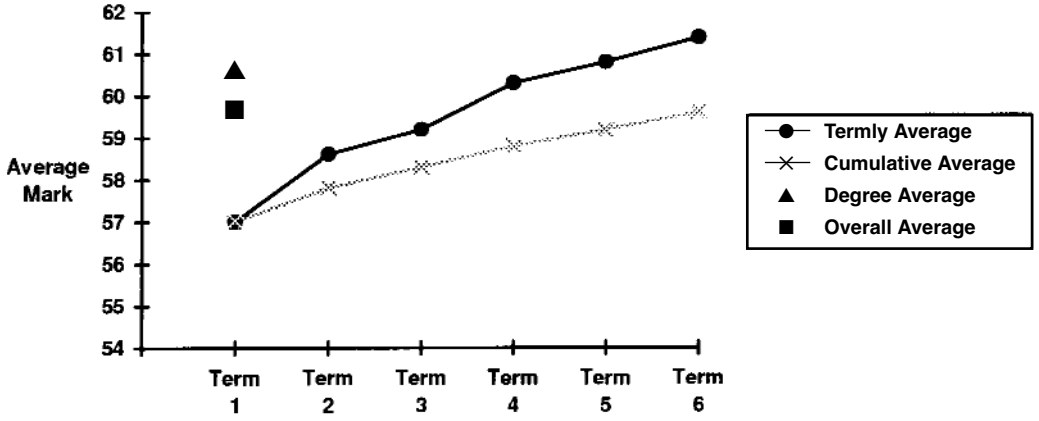
All advanced modules in the Oxford Brookes Undergraduate Modular Programme are pitched at the same nominal level of academic difficulty, and full-time students spend six terms drawing from the general pool of modules at this level. If there is learning progression or 'value added', it is to be expected that students will improve their performance during this time because of increased knowledge, enhanced transferable and discipline-specific skills, and greater understanding of implicit features of the course (the 'hidden curriculum'). With units of constant difficulty assessed on the same scale, improved performance should manifest itself in the form of higher average marks. Table 6 and Figure 4 show that improvement does occur.

A repeated measures analysis of variance showed that the improvement in average mark across the 6 terms of a full-time student's undergraduate career was significant ($F = 103.74$; $df = 5$; $p < 0.001$). A linear regression analysis showed that the intercept for mean mark by term number is 56.7 and the slope coefficient is 0.8. (expected mean mark in term n is thus (56 + 0.8n; most simply the average student gains a mark of 56.7 in term 1 and improves by 0.8 per cent in every successive term). The slope coefficient is significantly greater than zero ($F = 194.2$; $df = 1, 5996$; $p < 0001$).

Table 6. Average mark gained by Oxford Brookes Undergraduate Modular Programme students in each of their final 6 terms of study.

	TERM 1	TERM 2	TERM 3	TERM 4	TERM 5	TERM 6
Average Mark	**57.0**	**58.6**	**59.2**	**60.3**	**60.8**	**61.4**
Standard Deviation	7.7	7.7	7.4	7.4	7.1	7.6
N	933	935	934	1074	1074	1048

Figure 4. Term mark achieved by Undergraduate Modular Programme students in each of their final 6 terms of study, and the relationship between the termly average, the cumulative average at the end of the same term, the overall average at graduation, and the 'degree average' over the best 18 modules.



The importance of this evidence of progression-related value added should not be underestimated. It is extremely difficult to demonstrate progression within a traditional course. Sometimes this is because only endpoint assessment data from final examinations is available. Sometimes it is because within-course assessments are not comparable (e.g. second years never take third year assessments and vice versa), and often because repeat assessments are administered by the same agents whose assessments may be affected by their expectations. The information presented in Figure 4 and Table 6, has been calculated solely for the purposes of the present study and has not been previously available to influence assessor behaviour. The Oxford Brookes modular programme does not distinguish between level 2 and level 3 modules (except via prerequisite relationships). The modules on which some students are doing comparatively well in term six of their programme, will therefore be the same modules on which other students are doing less well because they have elected to take them in term three.

An intriguing feature of the data in Table 6, is the fact that term-by-term improvement is of the same magnitude between terms 5 and 6 as between terms 2 and 3. While this indicates that modules remain challenging and produce benefits for students in their final term of study, it also implies that the academic development of undergraduates on Oxford Brookes' undergraduate modular programme remains incomplete at the end of term 6. Performance improvement is arbitrarily terminated by the end of the course, rather than reaching a plateau at a level of performance higher than that at entry. This may be a hitherto unacknowledged property of all undergraduate education, or it may be specific to the Oxford Brookes undergraduate modular programme. The most arresting implication of the finding concerns conventional interpretations of degree class as an indicator of ability. For example, graduate programmes commonly use degree class as a measure of potential for graduate study. The data reported here suggests that this view is mistaken as academic performance is likely to continue to improve for some time beyond the duration of an undergraduate course.

Conclusions

The studies reported in the present paper, along with those described by Lindsay, Breen and Jenkins (2002, this issue) are intended to illustrate how teachers and researchers in particular disciplines can use a variant of action research methodology to influence policies over which they have no direct control. A secondary theme of both papers is that policy-relevant evidence of the kind required for indirect intervention, is also badly needed to assist higher education managers in moving to evidence-based decision-making. Lindsay, Breen and Jenkins (2002) provides a range of examples of policy-oriented studies that have been successful to varying extents in effecting policies and practice outside the department that originated the research. The present paper begins by noting that radical change in the organisation of frameworks for course delivery in UK higher education has been introduced without reference to academic considerations. The consequences for student learning and performance of switching to modular courses and a semesterised academic year are almost entirely unexplored. The paper goes on to present data that has had little if any impact upon decision-making, but serves as an example of the kind of data that is needed to gain a better understanding of whether or not semesterisation is pedagogically benign, and how some of the variable parameters of modular courses impact upon student achievement. It is hoped that the studies reported, will (although they are undoubtedly replete with imperfection), serve as models that may encourage others to collect and analyse data of a similar kind so as to contribute to the cumulative knowledge base that is so sorely needed to support evidence-based management in higher education.

References

Ayllon, J. & Azrin, N. (1968). *The token economy.* New York: Appleton-Century-Crofts.

Fearnley, S. (2001). *Report on Survey of Semesters via Psychology Departments.* Oxford: Oxford Brookes University Psychology Department. Mimeo.

Lindsay, R.O. (1998). Analysing Student Performance. In C. Rust (Ed.), *Improving student learning: Improving students as learners.* Oxford: OCSLD. 32-55.

Lindsay, R.O., Breen, R. & Jenkins, A. (2002). Evidence-based management and action-at-a-distance. *Psychology Teaching Review, 10*(1), 20–30.

Lindsay, R.O. & Gibbs, G. (1986). 'A Study of Workload Variability'. *Teaching News, 13,* 20–27. Oxford Polytechnic.

Lindsay, R.O. & Paton-Saltzberg, R. (1993). *The effects of paid employment on academic performance. Report to the Academic Standards Committee of Oxford Brookes University.* Oxford: Oxford Brookes University.

Correspondence

Dr R.O. Lindsay
Psychology Department
Oxford Brookes University
Gipsy Lane
Oxford OX3 OBP.