Assessment and its outcomes: the influence of disciplines and institutions

Adrian Simpson

School of Education, Durham University, Durham, DH1 1TA, UK

Email: adrian.simpson@durham.ac.uk

Telephone: +44 191 3347261

Existing research provides evidence at the module level of systematic differences in patterns of assessment, marks achieved and distributions of marks between different disciplines. This paper examines those issues at the degree course level and suggests reasons for the presence or absence of those module-level relationships at this higher level. The analysis finds that both discipline and institution have large and roughly equal impact on the balance between assessment types. However, contrary to the suggestions in the literature, that balance has virtually no independent impact on degree outcomes. The analysis also discovers that while there is only a small independent impact of disciplines on average degree marks compared to the institutional impact, disciplines do have a larger relative impact on the distribution of those marks.

Keywords: assessment, disciplines, examination performance, quantitative research, degree performance

Introduction

A number of papers have raised the issue of the relationship between disciplines and teaching and learning. For example, Entwistle and Ramsden (1983) argued that students studying soft disciplines tend to score higher for deep learning than those studying hard disciplines and Smeby (1996) suggested that there are systematic differences in preparation time and instruction type between different disciplines.

Other research has looked directly at the outcomes of degree courses, often as a proxy measure for university performance. In some cases these analyses sideline the role of disciplines (e.g. Bee and Dolton 1985; Johnes and Taylor 1987). However, a number of papers have looked directly at some of the inter-relationships between assessment, marks and disciplines. For example, Neumann, Parry and Becher (2002) provided a general framework for considering the impact of disciplines on a wide range of different facets of teaching and learning, including assessment, noting that 'both modes of assessment and the determination of grades within them will reflect disciplinary characteristics' (p. 408). However, their evidence was taken from a range of previous papers which included few wider empirical studies.

In contrast, a range of papers from Bridges and colleagues (Bridges et al. 1999; Bridges et al. 2002; Yorke, Bridges and Woolf 2000) looked in detail at individual student performance across modules; sometimes examining up to ten different disciplines across up to seven different universities in the UK. They have provided intriguing evidence of systematic differences in the relationships between assessment patterns, marks and mark distributions at the module level. Until recently, such studies have had to rely on the laborious collection of data from individual departments or on the opportunities afforded to researchers to gather data from their own institutions. However, the availability in the last few years of large, reliable data sets which cover all degree courses at all institutions allows the suggestions made from examining modules to be robustly tested at the degree course level.

This paper looks at the issues of the patterns of assessments, marks and mark distributions across degree courses in the UK. In particular, the analysis examines the impact of disciplines and the relationships between patterns of assessment with marks and their distributions across a full range of disciplines and institutions.

Disciplines, Degree Courses and Data

Defining what is meant by a discipline, let alone agreeing a classification system, is difficult. Trowler (2012) provides a detailed discussion of the difficulties in developing a clear definition and, indeed, the problems associated with the word 'discipline' itself (against 'subject', 'field', 'area' or any other nomenclature which attempts to distinguish groupings consistently across the research, teaching, knowledge generation, social and power dimensions inherent in academic organisations).

Krishnan (2009) makes an attempt at a definition by providing six general characteristics of a discipline: objects of research, a body of specialist knowledge, theories, specialist language, research methods and, most importantly, institutional manifestations. Trowler's attempted definitions contain one common feature: that disciplines 'take organisational form'. The most obvious way in which disciplines have institutional manifestation and take organisational form is in the degree courses taught at universities.

There have been numerous ways in which disciplines have been analysed and organised into categories. Biglan (1973) famously characterised disciplines on the basis of faculty members' judgements about the similarity of subject areas. The analysis resulted in the identification of three dimensions: pure/applied, hard/soft and life/non-life.

Subsequent empirical research has verified Biglan's classification scheme (e.g. Muffo and Langston 1981; Smart and Elton 1982; Stoecker 1993), though when it is used, researchers generally reduce the original system of three continuous dimensions to a pair of dichotomies (pure/applied and hard/soft).

Becher (1989) described Biglan's dimensions as cognitive: based on the nature of the knowledge in the area. Others have noted other types of dimension: for example,

Becher and Trowler (2001) introduced the notions of urban/rural (depending on how closely the production and communication of knowledge is interwoven) and converging/diverging (depending on the extent to which standards are able to shift). Biglan himself suggested that data from one of the institutions in his sample indicated a possible further dimension: creative/empirical. However, given the weight of research evidence supporting the distinctions and their widespread adoption, the pure/applied, hard/soft classification is used as the highest level of categorisation in this paper. Rocconi's (2013) classification of disciplines is adapted to place courses within these categories, while a finer grained categorization adopted from the Higher Education Statistics Agency is used for examining particular disciplines.

The 2009 white paper from the UK government, *Higher Ambitions: the Future of Universities in a Knowledge Economy* (Department for Business, Innovation and Skills 2009), led to the requirement for universities and other higher education institutions to provide information to the Higher Education Statistics Agency in the form of 'Key Information Sets', one for each degree course. The justification was to provide information for prospective students about a range of features of each course and institution, though it has been argued that it marks an escalation of the market influence on higher education (Brown 2013). The information provided ranges from the cost of accommodation and the number of beds managed by the institution, through course fees, to the responses to each question on the National Student Survey (an annual survey of final year undergraduates on their views of different aspects of their degree experiences). This data is freely available from the Higher Education Statistics Agency.

Within the key information set data, there are three nested classification systems for individual degree courses, based on the 'joint academic coding system'. For the data set analysed in this paper, the coarsest level consists of 21 different areas (such as

'biological sciences' and 'historical and philosophical sciences'), the second level has 42 areas (for example, 'biological sciences' is split into areas such as 'biology and related sciences', 'psychology' and 'sports science') and the finest level has 108 areas (with, for example, 'biology and related sciences' split into areas such as 'biology' and 'zoology'). For the purposes of this paper, this last, finest level of classification is used as proxy for the disciplines, within which individual degree courses are nested.

While some have argued that the notion of discipline may be decreasingly relevant with increasing interdisciplinarity in academia (e.g. Trowler 2012), this may not be reflected in the structure or classification of degree courses. Higher Education Statistics Agency allows courses to be coded against up to three different discipline areas in their data, but 70% of degree courses are coded against a single discipline, 28% against two and only 2% use all three codes. Gibbs (2012) even argues that there is a movement away from mixed disciplinary courses in the UK in part because of the way data is attributed to courses and not to modules.

However, for this study, in many cases multiple coding does not affect the classification within broader categorisations: for example, a history and philosophy degree would be classified as 'soft/pure' within Biglan's (1973) system whether it was treated as within the discipline 'history' or within the discipline 'philosophy'. For the purposes of the analysis here, each degree course with a multiple code was assigned randomly to one of its constituent disciplines.

It is important to note that not all of the key information set data are of equal reliability. For example, data on earnings and employment destinations are based on the 'Destinations of Leavers from Higher Education' survey completed by graduates who have to agree both to fill in the survey and reveal information about salaries and job types, which is likely to lead the data set to be incomplete, conflated and biased by

many different factors. However, for the purposes of this paper, the key information set data provides information on the proportion of different forms of assessment and proportions of degrees in the honours classification system and this data comes directly from universities themselves who have an obligation to be both accurate and complete wherever they can.

The full data set was obtained from the Higher Education Statistics Agency website, which is open access. the Higher Education Statistics Agency collect data on all UK undergraduate courses of at least one year's duration and consisting of at least 120 credits. This includes courses which do not lead to an undergraduate degree, such as foundation courses, as well as courses leading to awards other than degrees (such as certificates and diplomas of higher education). This paper is restricted to the analysis of courses leading directly to the award of a degree. In addition, institutions offering a very small number of courses, which include some highly specialised institutions and colleges with a small higher education provision, could have a disproportionate effect on the results, as might any disciplines so specialised that they are only taught at a small number of institutions. So any institution with less than 10 different degree courses was omitted from the data as was any discipline taught at less than 10 institutions. Because of the difference with the Scottish school and higher education systems, degrees taught at Scottish universities were omitted. The analysis is thus based on 15709 courses at 129 institutions covering 94 disciplines in the data available from the Higher Education Statistics Agency for courses with start dates between 1st August, 2013 and 31st July, 2014.

From this data, one can seek evidence to address the questions about the relationships between assessment methods, degree outcomes and disciplines and thus directly address assertions made in the literature.

Assessment methods and the disciplines

Neumann et al. (2002) make a number of assertions about the relationship between methods of assessment and Biglan's classification scheme:

> 'In soft pure settings, continuous assessment, where students are assessed by topics throughout a semester, is considered preferable to an emphasis on examinations…In soft applied fields, the aims of enhancing professional practice and yielding protocols and procedures are reflected in the typical assessment practices. Not only do essay and project-based assessments predominate, but peer and self-assessment tasks are more common, the intention being to improve self-reflection and practical skills …The preference for examinations—including questions based on multiple choice—covering a large corpus of factual knowledge is commonplace in hard applied as in hard pure fields' (p. 409)

The Higher Education Statistics Agency attempts to split assessment type into three components: written examinations, coursework and practical work. While the definition of written examination appears relatively clear, the distinction between the last two is far from evident from their advice to institutions. For example, the Higher Education Statistics Agency classifies 'practical skills assessment' (which includes, for example 'laboratory techniques') as 'practical', but 'reports' (including 'a report after participating in a practical activity such as fieldwork, laboratory work') is classified as 'coursework' (Quality Assurance Agency for Higher Education 2011, pp. 18-19) and it is easy to see how these may become conflated or misclassified.

With that in mind, Neumann et al.'s assertions would suggest the hypothesis that hard disciplines use examinations more than soft disciplines and that applied disciplines use practical work more than pure. If this were the case, it should manifest itself in the data as a decreased use in the proportion of examinations for soft over hard courses and applied over pure courses, possibly with a smaller effect in the latter comparison because of the coursework/practical conflation within the classification system.

Figure 1 shows the interaction plot for these comparisons across the data set. As with all results and graphs in this paper, the analysis was undertake with the R statistical package with standard libraries (including the linear mixed effect package 'lme4'). It suggests differences both between hard and soft, and between pure and applied courses (as well as a possible interaction). A two-way, between subjects anova confirms both main effects ($F_{hard/soft}(1,15626) = 1566$, $p < 0.001$), $F_{pure/applied}(1,15626) = 164$, $p < 0.001$) and an interaction effect ($F_{interaction}(1,15626) = 70.1$, $p < 0.001$). A Tukey's HSD post-hoc test suggests that all group differences are significant (ps < 0.001). In this case, the effect for hard vs soft courses ($\eta^2 = 0.09$) is a medium effect, for pure vs applied ($\eta^2 = 0.01$) is a small effect and the interaction ($\eta^2 = 0.005$) is a very small effect, though it does seem that the difference between pure and applied is larger for hard courses than for soft ones.

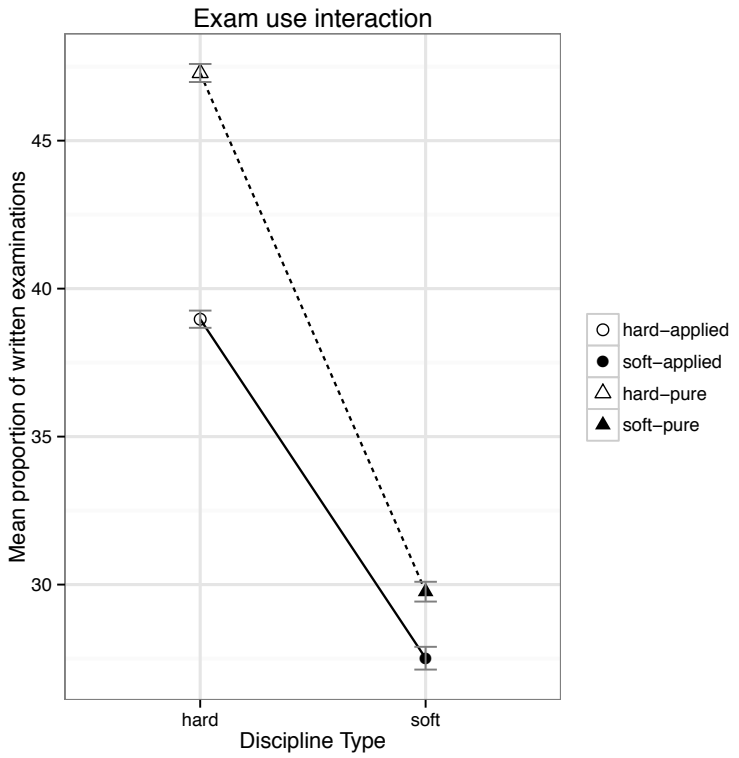

Figure 1. Interaction plot of proportion of written examinations by discipline type.

That is, the analysis confirms Neumann et al.'s assertion: hard courses use written examinations more than soft courses and pure courses use written examinations more than applied courses. However, Neumann et al.'s analysis was at the broadest level of discipline classification: the Biglan categories. Finer analysis can look for the source of these differences.

Courses can be considered as embedded within disciplines, but it is also clear that different institutions deliver different patterns of disciplines. Given that both disciplines and institutions are likely to have a direct effect on patterns of assessment, the analysis needs to account for both and for the interaction between them (effects which have not always been examined, e.g. Smith and Naylor [2001]; Yorke [2009]). While highly specialised institutions and others delivering a very small range of disciplines have been removed from the data, there is still likely to be an effect caused by different institutional coverage of disciplines. So a crossed linear mixed effects model was constructed to account for courses considered as nested both within disciplines and within institutions, allowing us to examine the impact of each while controlling for the other.

Figure 2 shows the effects of discipline on the proportion of examinations used. For example, controlling for institution, courses in mathematics and statistics use of written examinations is around 28 percentage points higher than the average across disciplines (29.8%) and fine art around 23 percentage points lower. 95% of disciplines use examinations for between 8.4% and 54.1% of their assessment.

Figure 3 shows the same analysis across institutions, controlling for discipline. 95% of institutions use examinations for between 13.8% and 68.2% of their assessment. Note that figure 3 highlights the institution 'mission groups' as at August 2013 (when

the data was collected by the Higher Education Statistics Agency). These mission groups are loose affiliations of UK universities which share similar institutional aims and figure 3 suggests that there is a strong relationship between these groupings and patterns of assessment. Figure 3 also shows that a small number of institutions very heavily skew their assessment towards examinations (even controlling for their range of disciplines).

Both discipline and institution account for a large proportion of variance in the use of examination (VPCdiscipline = 0.29, VPCinstitution = 0.39); that is, discipline accounts for around 29% of the variance in the use of written examinations, with institution accounting for around 39% of the variance.

This seems to confirm Neumann et al.'s (2002) assertion that hard courses rely on examinations more than soft courses and, while the conflation between coursework and practicals in the Higher Education Statistics Agency's data makes it hard to directly test the pure/applied distinction, the reduced difference and the interaction, suggests that there is a different pattern between all four categories of degree courses. Figure 2 shows this for individual courses and the variance partitioning suggests that there is a strong influence on the pattern of assessment coming independently from disciplines and from institutions.

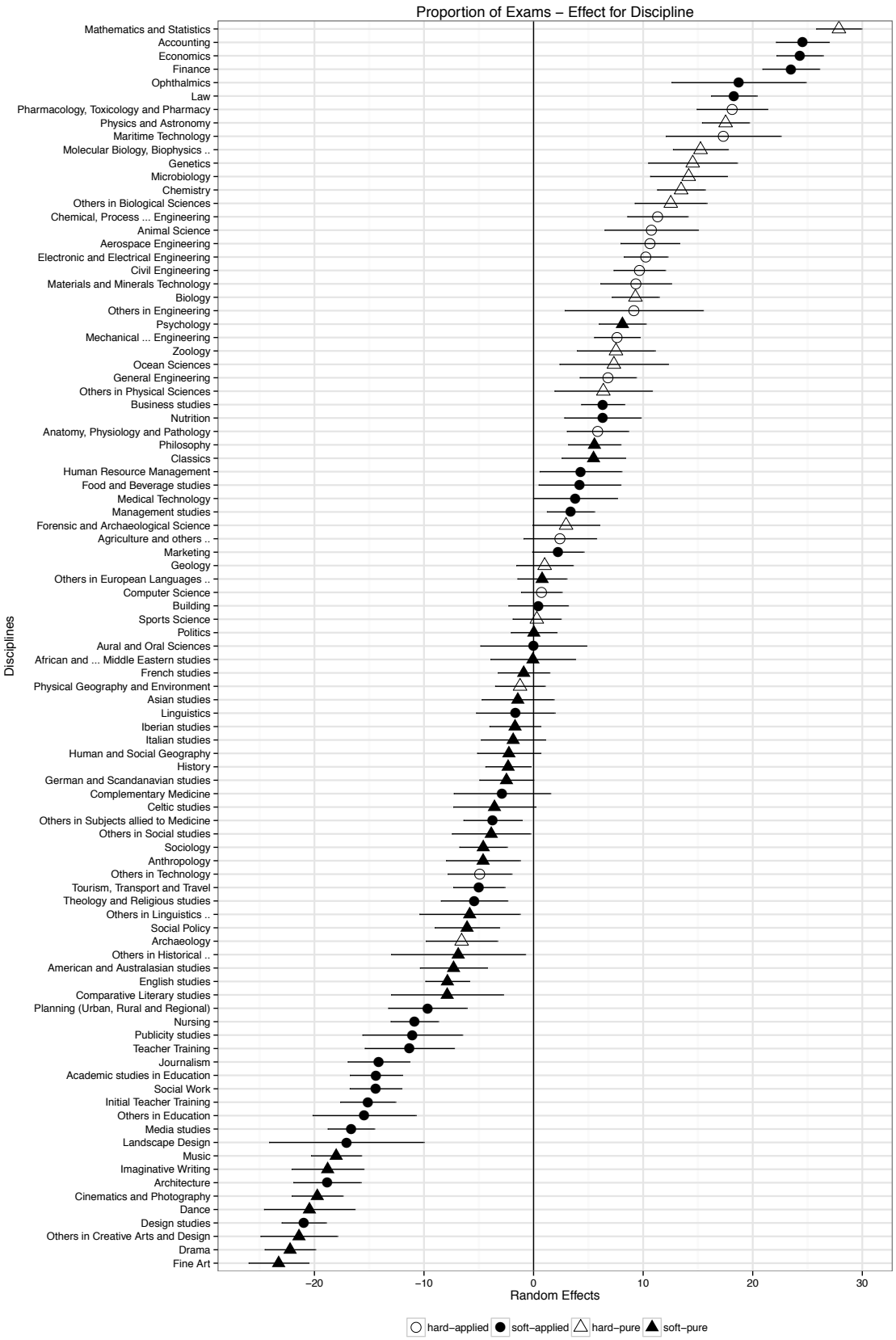

Figure 2. Effects of discipline on proportion of written examinations.

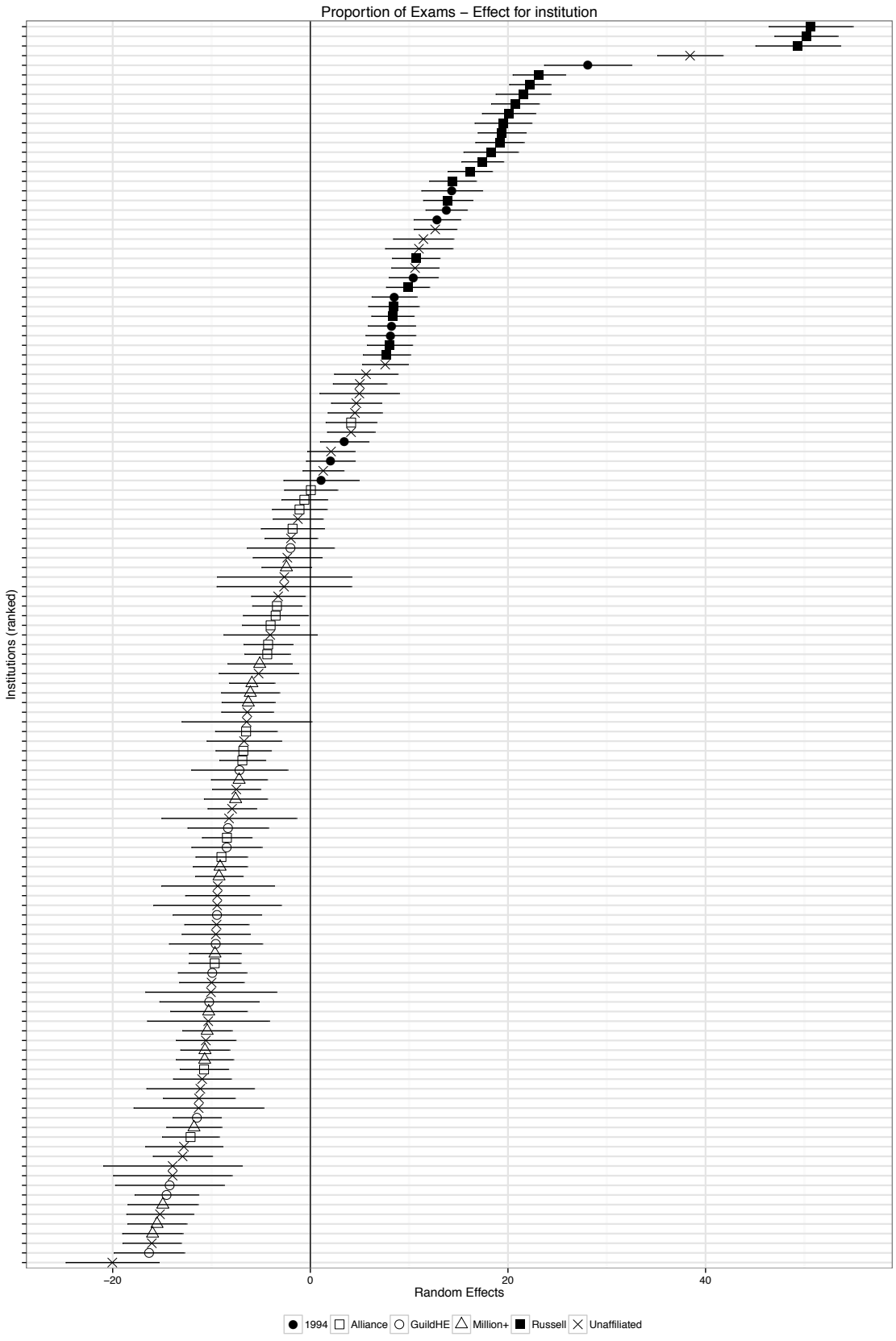

Figure 3. Effects of institution on proportion of written examinations.

Assessment outcomes and disciplines

Final Degree Marks

Bridges et al. (2002) raised concerns about the impact of differing patterns of assessment on marks achieved. They compared the performance of students on modules in six disciplines (biology, business studies, computer studies, English, history and law) at four universities in the UK with different proportions of coursework (which they took to include what the Higher Education Statistics Agency classifies as practical work) and written examinations. They noted that in the vast majority of modules, mean coursework marks were higher than mean examination marks and, to varying extents, individual students' performances tended to be higher on their coursework than on matched examinations. In some disciplines the difference between mean coursework and examination marks was as much as two thirds of a degree class (where, in the UK system, one increase in degree class would roughly correspond to 10 percentage points increase in average mark). Richardson's (2014) wide ranging literature review appears to support this: "assessment by coursework alone or by a mixture of coursework and examinations tends to lead to higher marks than assessment by examinations alone" (p. 12). Bridges et al.'s finding rightly raised concerns about how choices are made between modules with different patterns of assessment and led them to conclude that 'universities that choose to assess their students entirely through examinations may greatly disadvantage some of their students' (p. 47).

The evidence Bridges et al. used was based on module performances in a fixed number of institutions and disciplines to which they could access individual student marks. Thus their conclusions lead to the question of whether the relationship between assessment pattern and marks would be more generally true at the course level and across all disciplines and institutions.

If this effect did occur at the degree course level, one would expect to see a negative correlation between marks achieved and proportion of written examinations. Since this relationship would be affected by differences between disciplines and institutions (and the different pattern of disciplines across institutions), again a crossed linear mixed effects model for mean final degree mark against proportion of written examinations was constructed with institution and discipline as group random factors. Note that one might consider having entry tariff as a covariate in the mixed effects model, however, since institution accounts for over 80% of the variance in average entry tariff, multi-collinearity prevents disentangling tariff and institution effects.

The mean and standard deviation of the final marks for each degree course in the data set were calculated using Sheppard's correction to account for binned data. Recall that specialist and other institutions with small higher education provision were removed as were disciplines with few courses and, for this analysis, any course that reported degree classification information for fewer than 20% of their students.

First it was noted that the model involving the proportion of written examinations was a better fit than a null model (which would attempt to account for the mean final mark only by degree course and institution without accounting for balance of assessment type): $\chi^2(1) = 16.6$, $p < 0.001$. However, examination of the fixed effect parameters in table 1 suggests that, while there is an influence, it is a very weak one.

	b	se(*b*)	95% CI
Intercept	62.6	0.23	(62.1, 63.1)
Examinations	-0.007	0.0017	(-0.093, -0.0027)

Table 1. Fixed effects for mean final degree mark model.

That is, controlling for institution and discipline, a 10% increase in the contribution of coursework to the balance of the assessment pattern would increase the final degree mark by an average of 0.07%. Indeed, the impact of the discipline and institution are both much larger by comparison, though the influence of institution far outweighs that of discipline on the variance in average final degree mark ($VPC_{discipline} = 0.06$, $VPC_{institution} = 0.42$) that is 42% of the variance in average final degree mark might be apportioned to institution while 6% can be apportioned to discipline.

Figure 4 shows the caterpillar plot of the effects of discipline on average final degree mark. For example, controlling for institution and the balance between examinations and coursework, anatomy degrees get an average final mark about two percentage points higher than the average across disciplines and social policy about two percentage points lower. 95% of the disciplines' effects lie between -1.57 and 1.62 percentage points away from the overall average. Figure 5 shows the same image for (anonymised) institutions. 95% of the effect for institution lies between -3.75 and 4.65 percentage points away from the average. Again, figure 5 suggests a strong relationship with mission group.

So, while Bridges et al. (2002) and Richardson (2014) maintain that there is a relationship between marks and assessment method at the module level, at the degree level it is extremely small and is considerably outweighed by the influence of institution and discipline. Moreover, the variance partitioning suggests that the key effect on the average final degree mark is the institution, much more than the discipline.

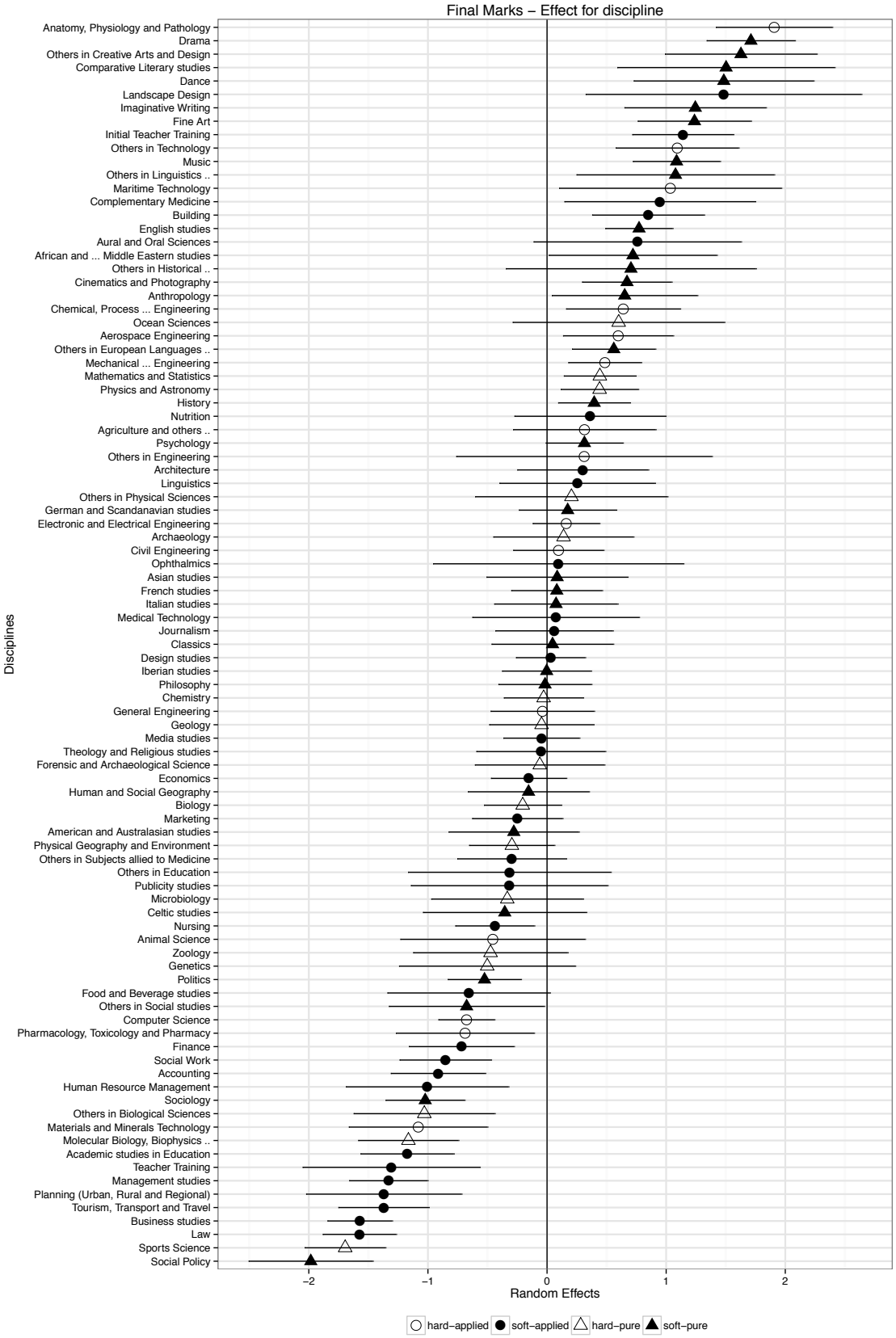

Figure 4. Effects of discipline on average final degree mark.

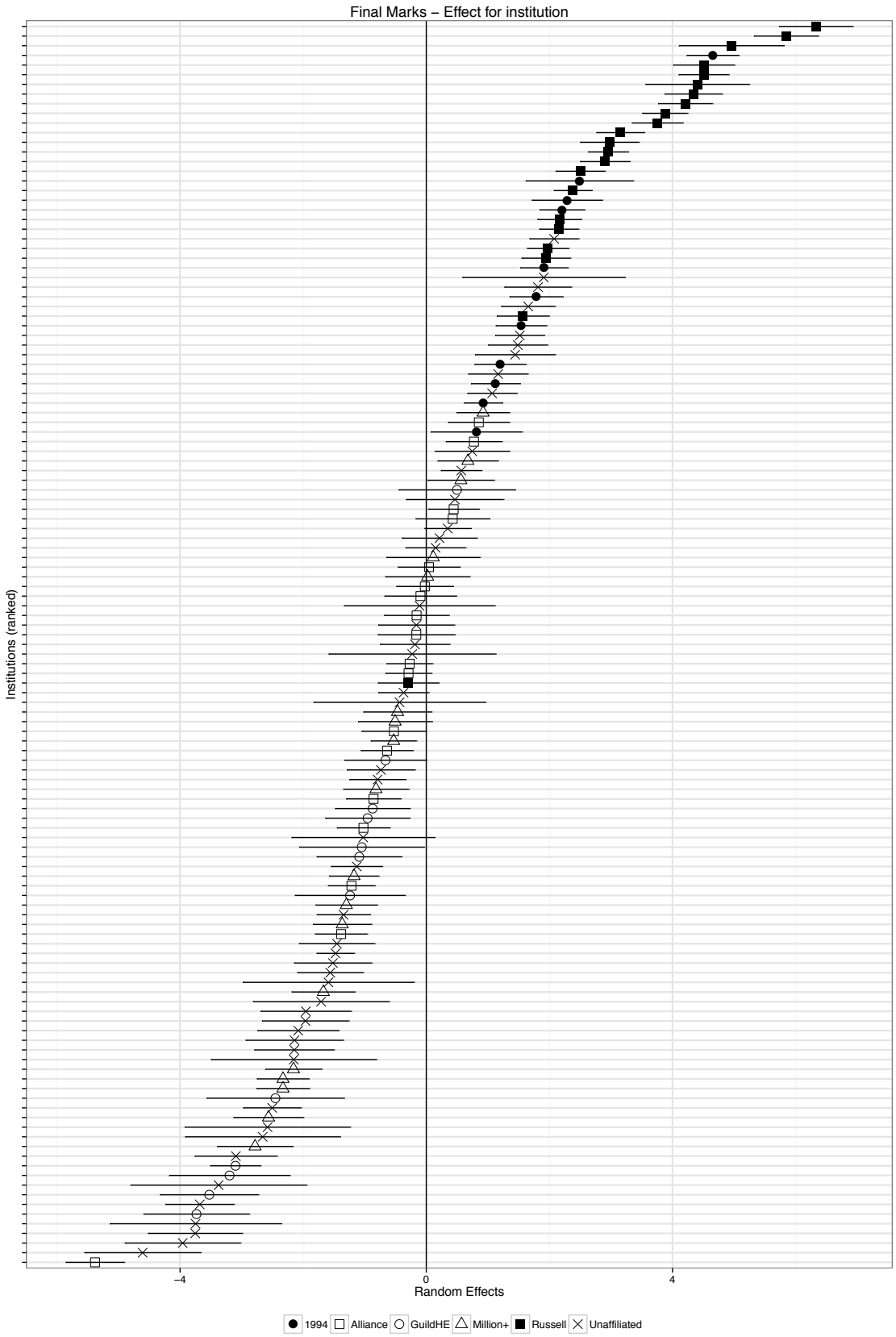

Figure 5. Effects of institution on average final degree mark.

Distribution of marks

Bridges et al. (1999) argued that the distribution of marks (at the module level) is linked to the nature of the discipline: that is, disciplines with qualitative subject matter have the narrowest spread and the disciplines with quantitative subject matter have the widest spread. They examined 10 disciplines at two institutions (although two of the disciplines were examined at only one institution) and looked in detail at a very large number of results across a wide range of modules within those disciplines. Their analysis suggested that the widest spread of marks was seen in modules within subjects classified as hard in Biglan's hard/soft categorisation. At the module level, as Yorke et al. (2000) pointed out, this has potentially substantial consequences in degree courses which allow broad module choice: they argued that the wider the spread of marks in a module, the more influence it can have on the degree classification.

As with the issue of the final degree mark, their analysis leads to the question about whether the relationship noted at the module level would hold at course level, across all disciplines and institutions. Figure 6 (a) shows the interaction plot which suggests there are indeed differences in the spread of marks between hard and soft courses and there are also differences between pure and applied courses. A two-way, between subjects anova showed a main effect for hard/soft ($F_{hard/soft}(1, 15626) = 3700$, $p < 0.001$) and for pure/applied ($F_{pure/applied}(1,15626) = 1150$, $p < 0.001$) and a comparatively small interaction effect ($F_{interaction}(1,15626) = 12.3$, $p < 0.001$). A post-hoc Tukey's HSD test showed that all group differences were significant (ps < 0.001), but the main effects for the difference between hard and soft courses ($\eta^2 = 0.18$) and between pure and applied courses ($\eta^2 = 0.06$) are large and medium respectively, while the interaction effect ($\eta^2 = 0.001$) is extremely small.

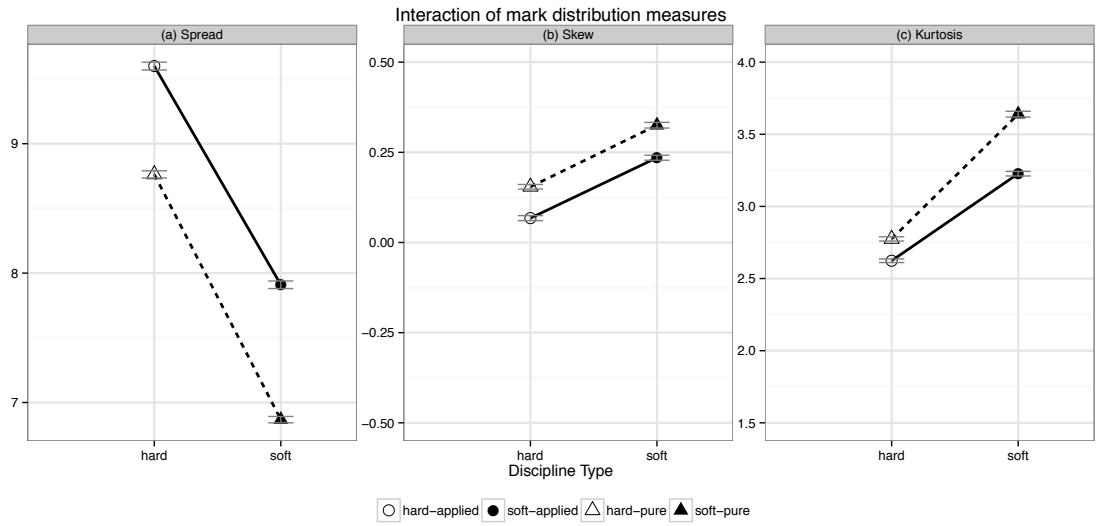

Figure 6. Interaction plots of measures of distribution of final degree marks by discipline type (scaled to reflect relative effect sizes).

Bridges et al. (1999) also noted the relationship between disciplines and third and fourth moment measures of distributions (i.e. skew and kurtosis) of marks at the module level. This is also visible at the degree course level. For skew, the two-way, between subjects anova showed main effects for both hard/soft ($F_{hard/soft}(1, 15626) = 504$, $p < 0.001$) and pure/applied($F_{pure/applied}(1,15626) = 148$, $p < 0.001$), though there was no interaction effect ($F_{interaction}(1,15626) = 0.05$, $p = 0.83$). Again, post-hoc Tukey's HSD tests showed that all group differences were significant (ps < 0.001). However, the main effect for the difference between hard and soft discipline was small ($\eta^2 = 0.03$) and between pure and applied was very small ($\eta^2 = 0.009$). This is illustrated in figure 6(b).

For kurtosis, the pattern was similar to spread: there was a medium main effect for hard/soft ($F_{hard/soft}(1, 15626) = 1721$, $p < 0.001$, $\eta^2 = 0.10$), a small main effect for pure/applied ($F_{pure/applied}(1,15626) = 349$, $p < 0.001$, $\eta^2 = 0.02$) and an extremely small interaction effect ($F_{interaction}(1,15626) = 54.6$, $p < 0.001$, $\eta^2 = 0.003$), with all group

differences being significant (post hoc Tukey's HSD test *p*s < 0.001). This is illustrated in figure 6(c).

These measures of the distribution at the course level generally confirm the phenomena seen by Bridges et al. at the module level: soft disciplines tend to have narrower and more peaked distributions than hard ones and pure courses tend to have narrower and more peaked distributions than applied ones (though this has less of an influence than the hard/soft distinction).

Again, further analysis can look to individual disciplines and institutions to see the underlying detail of these effects. A linear mixed effects model for standard deviation of final degree mark against proportion of written examinations was constructed with institution and discipline as crossed group random factors

It was noted that, while it is a better fitting model than the null model which ignores the balance of assessment ($\chi^2(1) = 21.2$, $p < 0.001$) there is a very small influence of the balance of coursework and examinations on the spread of final degree marks (table 2)

	b	se(b)	95% CI
Intercept	7.9	0.13	(7.64, 8.17)
Examinations	0.004	0.0009	(0.002, 0.006)

Table 2. Fixed effects for standard deviation of final degree marks model

That is, controlling for discipline and institution a 10% increase in the proportion of examinations in the assessment pattern would lead to a 0.04% increase in the standard deviation of the final degree marks. Again, this effect is overwhelmed by the effect for discipline and institution, though in this case the proportion of variance in

the spread of degree marks which can be accounted for by discipline is around the same as that for institution ($VPC_{discipline} = 0.26$, $VPC_{institution} = 0.22$).

Figure 7 shows the random effects for discipline on the spread of final degree marks. For example, controlling for institution and the pattern of assessment, mathematics and statistics standard deviation is 2.47 percentage points higher than the mean across disciplines and drama is 1.65 points lower. 95% of disciplines' standard deviations lie between -1.44 and +2.22 percentage points of the average. Figure 8 shows the effects of institution on the spread of final degree marks. 95% of the institutions' effects lie between -1.49 and 1.79 percentage points of the average and again indicates a relationship to mission group.

This seems to confirm that what Bridges et al. saw at the module level for 10 courses at two universities is generally true at the course level across all disciplines and institutions: hard courses appear to have a wider spread of marks than soft ones and applied have a wider spread than pure courses. There is a barely noticeable effect for the balance of assessment between coursework and examinations, as was the case with the mean final degree mark, which contradicts Richardson's (2014) assertion that coursework tends to reduce the variability in marks. Unlike the final degree mark, however, both discipline and institution have a large, and roughly equal influence on the spread of marks.

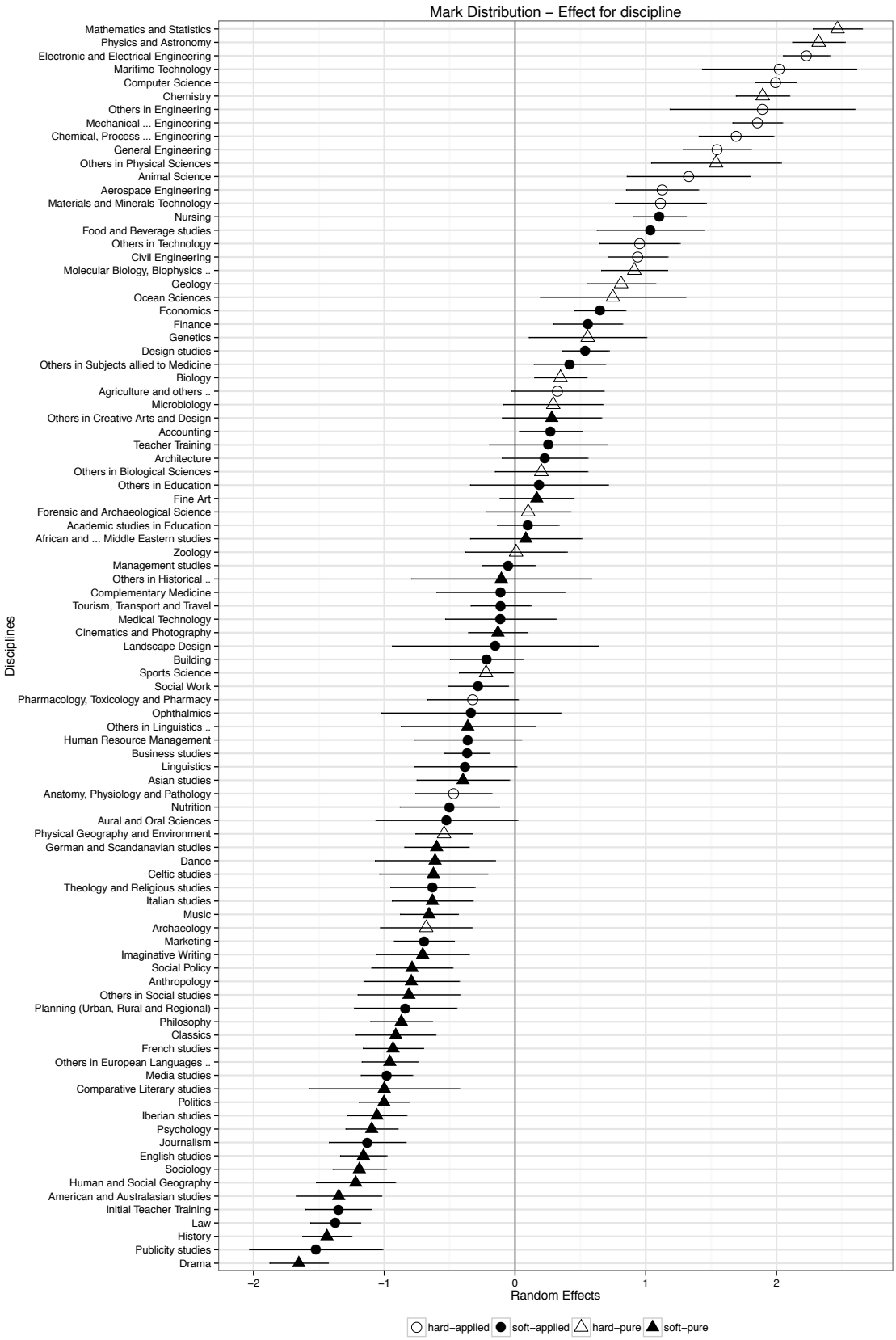

Figure 7. Effects of discipline on spread of final degree marks.

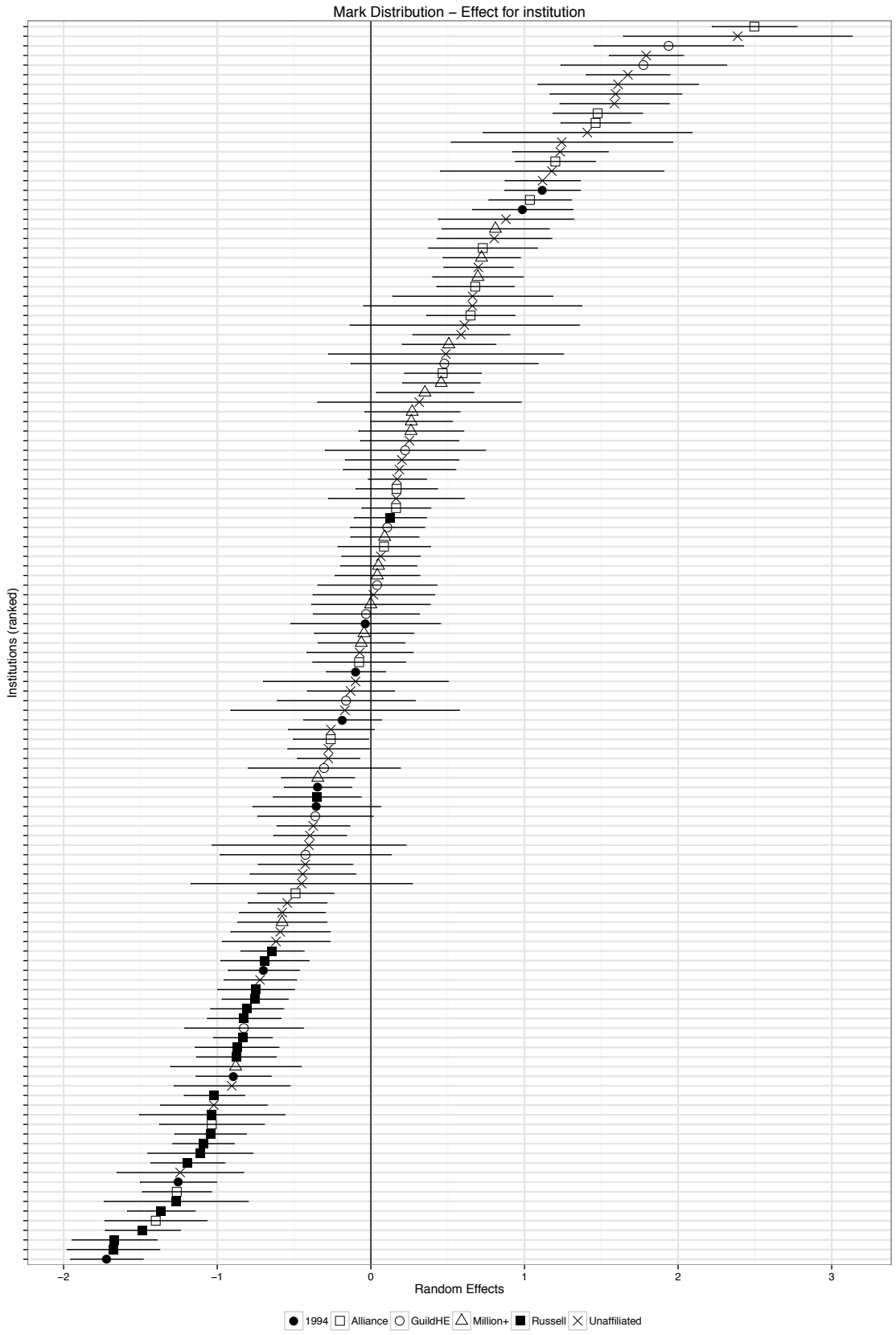

Figure 8. Effects of institution on spread of final degree marks.

Discussion

The analysis of the Higher Education Statistics Agency data related to the balance of assessment and the final degree outcomes confirms some, but not all of the suggestions in the literature concerning the influence of disciplines. It is clear that different disciplines balance the use of written examinations with coursework (interpreted widely) in very different proportions. In extreme terms, accounting for institutional differences, mathematics degrees use written examinations about 50 percentage points more than fine art degrees and 95% of disciplines range across 45 percentage points difference in their use of exams. But institutions also have an impact on the balance of assessment: accounting for discipline differences, the top ranked university uses examinations about 65 percentage points more than the bottom ranked university with 95% of institutions ranging across 35 percentage points difference.

It should be noted that this relationship between institution and proportion of written examinations has been commented on before, at least in the case of particular disciplines. For example, in mathematics, Iannone and Simpson (2011) described a strong negative correlation between league table ranking position and proportion of written examinations used (though ranking is likely to be highly correlated with other institutional factors).

The analysis in this paper confirms Neumann et al.'s (2002) suggestion about the pattern of assessment and its relationship to Biglan's (1973) classification of disciplines: hard courses tend to use written examinations much more than soft courses and pure courses use them somewhat more than applied ones. In this case, the influences from institutional and discipline factors on this balance appear to be large and roughly equal.

However, the balance of assessment appears to have very little impact on either the average or distribution of final degree marks: Bridges et al.'s (2002) finding that, at

the module level, increased coursework tends to result in increased marks appears to be a much less important influence on the final degree marks when examined across all disciplines and institutions. The very small relationship between assessment type and average mark compared to the relationship with discipline may also support the argument from Smith and Miller (2005) that discipline influences student learning, but that assessment type does not.

But in this case, there is a mismatch between the influence of institution and discipline: there is only a relatively small effect for discipline and a much larger effect for institution on the average final degree marks, even controlling for different assessment regimes. In extreme terms, the difference between the mean final degree mark for the top ranked discipline (anatomy) and the bottom ranked discipline (social policy) is around 4 marks and the averages in 95% of disciplines lie within about 3 marks of each other. However, the difference between the top ranked university and bottom ranked university is about 12 marks, more than a full degree class, with 95% of institutions ranging over more than 8 marks difference in their averages. One interpretation is that 'institution' in this case, is a proxy for 'entry grade' and that it is entry grade which has the most substantial effect on the average degree mark across different courses, though Yorke (2011) suggested it is unlikely to be a major factor in accounting for an effect of this type and size, which agrees with an earlier meta-analysis which suggested that while it was significant, the effect was small (Peers and Johnston 1994).

This mismatch in the influences between discipline and institution does not appear to be present in the analysis of mark distributions: disciplinary differences appear to have a large impact on the distribution and one which is about equal to that of institutions. The analysis in this paper supports, at the final degree mark level,

something Bridges et al. (1999) noted at the module mark level: courses classified as soft have a narrower spread and more peaked distribution of final marks than hard courses, and pure courses have a narrower spread and more peaked distribution than applied ones. However, Richardson's (2014) assertion that coursework reduces variability in marks is not borne out at the degree course level.

One might consider what balance of regulatory forces could account for the patterns of influence of disciplines and institutions on marks and their distributions. It is important to recognise that degree marks and classes are not formally standardised between universities. For the UK institutions analysed here, entry is normally based on A level grades: examinations sat in the last year of schooling which are provided by centralized examination bodies independent of schools and which are standardized. However, universities set their own examinations and the influences on grading between institutions are more subtle.

Yorke (2011) provides a detailed critique of marks and grades as measures. He notes the complexity involved in grading, combining and standardizing across modules within courses, courses within institutions and disciplines between institutions. His closely argued discussion of the flaws and inequities in differing systems makes it all the more important to see how one can account not only for the difference in the influences of institution and discipline on average final degree mark and mark distributions, but also the relative lack of difference between say, the average final mark in different disciplines once we control for institutional factors.

Decisions on marks on a university degree course are generally in the hands of internal examiners, but there are a number of competing regulatory forces which act on their decisions: some which are wholly institutional, some wholly disciplinary and some which impact on both.

For example, the analysis here suggests that different disciplines have very different patterns of assessment type and while this does not account for more than a very small amount of the variance in average final marks and their distributions, it may lead to the hypothesis that disciplinary factors such as curriculum design, structure and teaching, may act as regulatory forces; and, of course, in some cases, professional bodies play a disciplinary regulatory role. On this last point, professional bodies may accredit courses (or parts of courses) for different purposes. For example, a body may recognize the degree for partial credit to a further professional qualification which gives license to practice. Harvey (2004) suggested that up to 100 professional regulatory bodies may have some say in accreditation of degree programmes in the UK. These bodies will undoubtedly have at least an indirect influence on the content and form of assessment for any degree course which seeks accreditation.

Similarly, the UK Quality Assurance Agency (an independent body which monitors and advises on standards in higher education) has laid out subject benchmarks, which are intended to provide broad sets of expectations about knowledge and skills of graduates at different levels in different subjects. These too might act as a disciplinary regulatory force. In an analysis of them, Yorke (2002) notes that different disciplines' benchmark statements contain different amounts of detail regarding level descriptors and appears to show, for example, that hard subjects tend to include level descriptors for excellence more than soft subjects do.

Another regulatory force which may influence the discipline is the external examining system used in the UK (amongst other countries). This system involves an academic from a cognate department external to the institution monitoring the assessment tasks and checking samples of students' work for a given degree course. It is clear that the role they play has moved in the last forty years from being a powerful

influence on individual marks with a stated responsibility to maintain standards between institutions (Williams 1979) to one of quality assurance and enhancement (Biggs 2001). However, Bloxham and Price (2015) argue that while the influence they play may rest on challengeable assumptions they still play an important role in "the social construction of standards and … the inter-institutional sharing and debating of them as the basis for a community consensus" (p. 207).

There are also likely to be less visible and formal disciplinary frameworks through which marking standards are regulated. For example, internal examiners will normally have their own experience of their own discipline at other institutions, not least from the universities from which they graduated (which, as Becher [1989], notes are likely to be from a disproportionately narrow, elite band).

But at the institutional level, there will be other, perhaps more immediate regulatory forces. Moreover, institutions may differ in the ways in which they grade, determined by institutional assessment policies. For example, one institution might require criteria based grading across departments, one might require elements of normative grading and another might leave decisions to faculty or department level. Institutions decide, to differing extents, on grading systems and mechanisms by which module marks get aggregated to determine final marks and degree classifications, which have to be adhered to across all disciplines in the institution. For example, Bridges et al (1999) noted some universities give grades from a small set of discrete options (e.g., 'A', 'B+', 'B', 'C' and 'F') and others use a percentage scale and they indicated the substantial effect the choice of system could have.

Again, less visible regulatory forces may be in play at institutional levels: comparison of degree outcomes between disciplines within a single institution can draw markers' attention to cross-disciplinary differences and act explicitly or implicitly to

decrease those differences. Other mechanisms such as league tables (which often include degree outcomes as an element) may also act to regulate both institutionally and within disciplines.

Gibbs (2012) argues that the availability and use of performance indicators, such as the National Student Survey impacts on institutions in complex ways, not all of which positively align with improved teaching and learning. For example, he notes that institutions are "making very broad scale changes that affect all degree programmes and all teachers … Some centrally determined changes will limit teachers' scope to enhance teaching in contextually sensitive ways, and will make things worse" (p. 10). That is, a superficial analysis of data or an emphasis on chasing ratings can lead institutions to make changes which may not fit all disciplines and this regulatory force may go some way to explaining the disproportionate impact of institution over discipline on, say. average final marks.

However, the relatively small effect for discipline on average final marks would appear to suggest that institutional regulatory forces are particularly strong on minimizing cross-disciplinary differences within an institution, but currently have a less overwhelming effect on distribution.

This provides an intriguing final thought. Recently, in the UK at least, there has been increasing pressure to transform assessment practices and diversify methods, which may significantly alter the balance of assessments (HEA, 2012). Distributions of marks may allow us to monitor the effect of institutions' central influence against disciplines' distributed influence as assessment patterns shift. In particular, given the concern that average degree marks appear to be increasing over time for many subjects and in many institution groups (Yorke 2009), one might monitor changes in distributions of marks. Institutional factors appear to influence average mark

substantially more than discipline but discipline factors influence distributions as much as institutions, so by examining the change of average mark over time and the change in distribution of marks over time one may be able to develop a proxy measure for the relative strength of institutional and disciplinary regulatory forces. This may provide an empirical measure of shifts in where power lies within academia.

References

Becher, T. 1989. *Academic Tribes and Territories: Intellectual Enquiry and the Culture of Disciplines*. Buckingham: Open University Press.

Becher, T. and P. Trowler, 2001. *Academic Tribes and Territories: Intellectual Enquiry and the Culture of Disciplines (2nd Edition)*. Buckingham: Open University Press.

Bee, M., and P. Dolton. 1985. "Degree Class and Pass Rates: An Inter-University Comparison." *Higher Education Review* 17(2): 45-52.

Biggs, J. 2001. "The Reflective Institution: Assuring and Enhancing the Quality of Teaching and Learning." *Higher Education*. 41(3): 221-238.

Biglan, A. 1973. "The Characteristics of Subject Matter in Different Academic Areas." *Journal of Applied Psychology* 57(3): 195-203.

Bloxham, S., and M. Price. 2015. "External examining: fit for purpose?" *Studies in Higher Education*, *40*(2): 195-211.

Bridges, P., B. Bourdillon, D. Collymore, A. Cooper, W. Fox, C. Haines, D. Turner, H. Woolf, and M. Yorke. 1999. "Discipline-related Marking Behaviour using Percentages: a Potential Cause of Inequity in Assessment." *Assessment & Evaluation in Higher Education* 24(3): 285-300.

Bridges, P., A. Cooper, P. Evanson, C. Haines, D. Jenkins, D. Scurry, H. Woolf, and M. Yorke. 2002. "Coursework Marks High, Examination Marks Low: Discuss." *Assessment & Evaluation in Higher Education* 27(1): 35-48.

Brown, R. 2013. "Mutuality Meets the Market: Analysing Changes in the Control of Quality Assurance in United Kingdom Higher Education 1992–2012." *Higher Education Quarterly* 67(4): 420-437.

Department of Business, Innovation and Skills. 2009. *Higher Ambitions: The Future of Universities in a Knowledge Economy*.

Entwistle, N., and P. Ramsden. 1983. *Understanding Student Learning*. London: Croom Helm.

Gibbs, G. 2012. *Implications of 'Dimensions of quality' in a market environment*. York: Higher Education Academy.

Harvey, L. 2004. "The Power of Accreditation: Views of Academics." *Journal of L. Education Policy and Management* 26(2): 207-223.

Higher Education Academy, 2012. *A Marked Improvement: Transforming assessment in higher education.* HEA: York.

Iannone, P., and A. Simpson. 2011. "The Summative Assessment Diet: How we Assess in Mathematics Degrees." *Teaching Mathematics and its Applications* 30(4): 186-196.

Johnes, J., and J. Taylor. 1985. "Degree Quality: an Investigation into Differences between UK Universities." *Higher Education* 16(5): 581-602.

Krishnan, A. 2009. *What are Academic Disciplines: Some observations on the Disciplinarity vs. Interdisciplinarity Debate*. University of Southampton: ESRC National Centre for Research Methods.

Muffo, J., and I. Langston IV. 1981. "Biglan's Dimensions: Are the Perceptions Empirically Based?" *Research in Higher Education* 15(2): 141-159.

Neumann, R., S. Parry, and T. Becher. 2002. "Teaching and Learning in their Disciplinary Contexts: A Conceptual Analysis." *Studies in Higher Education* 27(4): 405-417.

Peers, I., and M. Johnston. 1994. "Influence of learning context on the relationship between A-level attainment and final degree performance: a meta-analytic review." *British Journal of Educational Psychology* 64(1): 1-18.

Quality Assurance Agency for Higher Education. 2011. *Explaining Contact Hours: Guidance for Institutions Providing Public Information about Higher Education in the UK*. Gloucester: QAA.

Richardson, J. 2014. "Coursework versus examinations in end-of-module assessment: a literature review". *Assessment & Evaluation in Higher Education*, 40(3): 439-455.

Rocconi, L. 2013. "Analyzing Multilevel Data: Comparing Findings from Hierarchical Linear Modeling and Ordinary Least Squares Regression." *Higher Education* 66(4): 439-461.

Smart, J., and C. Elton. 1982. “Validation of the Biglan Model." *Research in Higher Education* 17(3): 213-229.

Smeby, J. 1996. “Disciplinary Differences in University Teaching”, *Studies in Higher Education*, 21(1): 69-79.

Smith, J., and R. Naylor. 2001. “Determinants of Degree Performance in UK Universities: a Statistical Analysis of the 1993 Student Cohort.” *Oxford Bulletin of Economics and Statistics* 63(1): 29-60.

Smith, S., and R. Miller. 2005. “Learning Approaches: Examination Type, Discipline of Study, and Gender.” *Educational Psychology* 25(1): 43-53.

Stoecker, J. 1993. “The Biglan classification revisited.” *Research in Higher Education* 34(4): 451-464.

Trowler, P. 2012. “Disciplines and Interdisciplinarity: Conceptual Groundwork.” In *Tribes and territories in the 21st century* edited by Paul Trowler, Murray Saunders and Veronica Bamber, 5-29. Abingdon: Routledge.

Williams, W. 1979. “The Role of the External Examiner in First Degrees.” *Studies in Higher Education*. 4(2): 161-168.

Yorke, M. 2002. “Subject Benchmarking and the Assessment of Student Learning." *Quality Assurance in Education.* 10(3): 155-171.

Yorke, M., 2009. *Honours Degree Classifications: What we can and cannot tell from the Statistics*. Gloucester: QAA.

Yorke, M. 2011. “Summative Assessment: Dealing with the ‘Measurement Fallacy’.” *Studies in Higher Education*. 36(3): 251:273.

Yorke, M., P. Bridges, and H. Woolf. 2000. “Mark Distributions and Marking Practices in UK Higher Education: Some Challenging Issues.” *Active Learning in Higher Education.* 1(1): 7-27.