



The influence of disciplinary assessment patterns on student learning: a comparative study

Tansy Jessop & Barbara Maleckar

To cite this article: Tansy Jessop & Barbara Maleckar (2016) The influence of disciplinary assessment patterns on student learning: a comparative study, *Studies in Higher Education*, 41:4, 696-711, DOI: [10.1080/03075079.2014.943170](https://doi.org/10.1080/03075079.2014.943170)

To link to this article: <https://doi.org/10.1080/03075079.2014.943170>



© 2014 The Authors. Published by Taylor & Francis.



Published online: 27 Aug 2014.



Submit your article to this journal [↗](#)



Article views: 2641



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 8 View citing articles [↗](#)

The influence of disciplinary assessment patterns on student learning: a comparative study

Tansy Jessop^{a*} and Barbara Maleckar^{b†}

^a*Learning and Teaching Development Unit, University of Winchester, West Hill, Winchester, SO22 4NR, Hampshire, UK;* ^b*Department of Psychology, University of Winchester, West Hill, Winchester, SO22 4NR, Hampshire, UK*

This paper explores disciplinary patterns of assessment and feedback, using data from the Transforming the Experience of Students through Assessment project. Its central research question concerns the effect of disciplinary assessment patterns on student learning. Audit data from 18 degree programmes at 8 UK universities showed variations in assessment patterns across three disciplinary fields: Humanities, Professional and Science courses. There were variations in assessment demands; in the quantity of feedback and in the proportion of examinations. Statistical analysis of Assessment Experience Questionnaire data ($n = 762$) explored whether these differences influenced students' perceptions of learning across the disciplines. Findings showed that there were no significant differences in students' perceptions of learning from examinations. Humanities students evaluated the appropriateness of their assessment lower than other discipline groups; professional students were less clear about goals and standards. The researchers propose explanations for these findings and suggest avenues for further research.

Keywords: disciplines; assessment patterns; feedback; deep learning; authentic assessment

Introduction

The idea of disciplinary groups dates back to the medieval university, when the disciplines of logic, rhetoric and grammar were distinguished from arithmetic, geometry, astronomy and music, dissecting knowledge and learning into the 'inner' domains of arts and humanities, and the 'outer' domains of understanding the world using mathematical principles (Fanghanel 2009, 566). More recently, there has been a resurgence of interest in the discipline as an important, yet overlooked organising principle, knowledge domain and social construct within higher education (Becher and Trowler 2001, 37–38). Research has explored how disciplines demonstrate distinctive social, cultural, epistemological and even ideological dimensions (Biglan 1973a, 1973b; Kolb 1981; Becher 1987, 1989, 1994; Becher and Trowler 2001; Wareing 2009). The purpose of this study is to provide insights into the pedagogy of the disciplines,

*Corresponding author. Email: tansy.jessop@winchester.ac.uk

†At the Institute of Media, Knowledge and Communication, University of Augsburg, Augsburg, Germany.

by distinguishing disciplinary differences in assessment patterns on degree programmes and determining their influence on student learning. The researchers use evidence of disciplinary assessment patterns to explore plausible explanations for effects on student learning. These explanations suggest avenues for further research to enhance student learning from assessment. In the following sections, the authors provide an overview of research on disciplines and pedagogy and a rationale for the present study.

Research on disciplines

Studies by Biglan (1973a, 1973b) showed that disciplines with a single paradigm, holding a generally accepted body of theory, displayed greater consensus about content and methods, and greater social connectedness between members of the discipline than subjects with more divergent knowledge fields. His research portrayed the sciences as a paradigmatic subject field, in contrast to social sciences, humanities and arts which he described as non-paradigmatic, with loosely structured knowledge domains encouraging less collaborative research. Out of this research, Biglan (1973b) developed a typology to classify disciplines into six groups with similar characteristics. Following from this work, Becher (1989) extended the idea of paradigmatic and non-paradigmatic disciplines in wide-ranging research about how academics from different disciplines ply their trade as researchers, teachers, social actors and career professionals. Becher's classic study of 12 disciplines at Berkeley in California resulted in the following four disciplinary categories:

- 'Hard pure' – denoting pure sciences such as Physics, Chemistry, Mathematics and Biology – quantifiable, impersonal knowledge forms, containing universally accepted 'truths'.
- 'Hard applied' – application of sciences to pragmatic and functional ends, for example in Engineering, Computer Science and Pharmacy.
- 'Soft pure' – humanities, arts and social sciences characterised by complexity, by being personal, value-laden and open to interpretation, for example, the study of Philosophy, English and History.
- 'Soft applied' – using soft knowledge for functional purposes, such as in Law, Social Work and Teaching.

Becher's classification of disciplinary groups is widely regarded as an instructive and reasonably accurate depiction, clarifying and simplifying earlier categorisations developed by Biglan (1973a, 1973b) and Kolb (1981). However, the idea of disciplines and disciplinary groupings has its detractors, particularly in a changing higher education landscape, with the growth of inter-disciplinary subjects, and the fragmentation of traditional knowledge fields. Disciplinary epistemology has become more complex and fluid, particularly in the arts, social sciences and humanities, which have always been more divergent and loosely structured than scientific knowledge fields (Becher and Trowler 2001, 184). Recently, there has been a spawning of hybrid, inter-disciplinary and new subjects (for example, Sports Science; Human Development and Minority Studies). This has made disciplinary classification more complex and ambiguous. Disciplines can no longer be viewed as fixed, unchanging and monolithic entities (Jones 2009, 85), nor are they always 'internally consistent' in practice and philosophy (Wareing 2009, 922). Yet, in spite of these complexities, disciplines remain a fundamental organising feature of universities, and the primary context for social interaction

for academics, and between academics and students (Smeby 1996; Becher and Trowler 2001; Neuman and Becher 2002). Departments and faculties are generally organised in a way which recognises the influence of shared disciplinary cultures and resource needs, for example, as pure sciences with requirements for laboratories. Most academics identify themselves with their disciplinary ‘tribes’, developing day-to-day intellectual and social practices in line with a community of practice which links to their disciplinary culture (Becher and Trowler 2001; Krause 2014, 5).

Disciplines and pedagogy

In most modern university contexts, for example, the UK, the USA, Australia and Norway, teaching is managed centrally at the institution-level (Shulman 1993; Smeby 1996). The knowledge and content of the curriculum are discipline specific, but the surrounding framework of the degree structure, assessment regulations, validation of courses and the professional certification of teachers, is centralised. In many disciplines, this leads to pedagogy being regarded as a universal set of activities and routines, which are not contextualised in the discipline, thereby disconnecting teaching from the intellectual community of its discipline (Shulman 1993). A generic view of teaching and learning masks the nuanced and complex ways in which disciplinary academics approach pedagogy. Academics tend to form disciplinary *research* communities, where research interests are shared through conferences and publications in the public sphere, whereas disciplinary *teaching* communities tend to be implicit or privatised (Shulman 1993; Becher and Trowler 2001; Krause 2014). Jones (2009) makes a case for ‘re-disciplining’ the scholarship of teaching and learning on the grounds that ‘an understanding of the cultures of the disciplines is germane, as is an understanding of the ways in which these disciplinary cultures shape teaching’ (97).

There is a small but growing body of literature specifically examining the relationship between learning, teaching and the disciplines (Smeby 1996; Neumann 2001; Neumann, Parry, and Becher 2002; Shulman 2005; Lindblom-Ylance et al. 2006; Wareing 2009; Pryor and Crossouard 2010). Research literature points to the tension between generic theories of teaching and learning, and disciplinary claims to distinctiveness, or even superiority. Several articles describe typical characteristics and related teaching and learning approaches adopted by different discipline groups – so for example, hard pure sciences tend to be more knowledge-driven, content-focused, cumulative, quantitative and ‘teacher-centred’ (Lattuca and Stark 1994; Neumann, Parry, and Becher 2002; Lindblom-Ylance et al. 2006), whereas soft pure domains are rich in language, and tend to adopt more interpretive, critical, divergent and ‘student-centred’ approaches. Shulman (2005) provides theory and evidence of ‘signature pedagogies’ in the applied disciplines (both ‘hard’ and ‘soft’), showing how particular teaching routines have developed in certain professions, for example the dialogic medical round for doctors, and case arguments in Law.

The ‘role of disciplines in shaping teaching’ is a ‘relatively new focus’ (Neumann 2001, 136). While a few studies make reference to disciplinary assessment and feedback regimes (Warren Piper, Nulty, and O’Grady 1996; Neumann 2001; Neumann, Parry, and Becher 2002; Pryor and Crossouard 2010), observing characteristics of assessment in the four main disciplinary groups, there is relatively little evidence about the influence of different disciplinary assessment patterns on student learning. This is surprising given that assessment has been described as a key driver of student learning, focusing students’ attention and effort to the extent that ‘from our

Table 1. Distinctive disciplinary patterns of assessment.

Discipline group	Assessment emphasis	Types of assessment used
'Hard pure'	Memorisation; application of course content; fact retention and solving logical problems	Examinations; practical work; laboratory reports; numeric calculations and Multiple Choice Quiz
'Hard applied'	Practical competence; application of theory to practice and factual understanding	Multiple Choice Quiz; examinations; simulations and case studies
'Soft pure'	Analysis and synthesis of course content and continuous assessment	Essays; oral presentations; short answer papers and project reports
'Soft applied'	Emphasis on personal growth and intellectual breadth and application of theory to practice	Simulation and case studies

students' point of view, assessment always defines the actual curriculum' (Ramsden 1992, 187). Similarly, wide-ranging studies underscore the importance of feedback for student learning (Black and Wiliam 1998; Hattie 2009). Research on disciplinary assessment and feedback seems to stop at the categorisation stage, for example in the National Report on Examination Practices in Australian Universities, as illustrated in Table 1 (Warren Piper, Nulty, and O'Grady 1996).

In this study, the researchers categorise assessment patterns by discipline and then explore the connection between disciplinary assessment patterns and student learning responses. The research contributes new knowledge to the field by triangulating evidence of student perceptions of learning with data about disciplinary assessment patterns. The authors go on to explore plausible explanations for the disciplinary effects on student learning from the literature. These hypothetical explanations provide fertile ground for further research.

Methodology

This paper provides recent data on assessment and feedback regimes from research-intensive and 'teaching-oriented' universities. The study used data from the 'Transforming the Experience of Students through Assessment' project, funded by the UK's Higher Education Academy (Jessop, El Hakim, and Gibbs 2011, 2013). The researchers categorised assessment and feedback data into three disciplinary groups: Sciences – pure and applied (combining 'hard pure' and 'hard applied'), Humanities ('soft pure') and Professional programmes ('soft applied').

Two sets of data were triangulated: the Programme Audit and the Assessment Experience Questionnaires (AEQ). These two research methods are described in more detail in the following section.

Programme audit

The programme audit has been shown to distinguish between different whole programme assessment environments (Gibbs and Dunbar-Goddet 2009). The audit consists of a discussion with the programme leader over programme and module documentation. The purpose of the audit is to distil elements of the assessment

process which relate to students' learning from assessment, their study behaviour and approaches to learning (Gibbs and Simpson 2004). The audit represents the planned curriculum, providing 'official' data about the experience of a student over the course of a three-year undergraduate degree. It captures data about the:

- number of summative assessments;
- number of formative-only assessments;
- variety of assessment methods;
- volume of oral feedback;
- volume of written feedback;
- timeliness of feedback;
- proportion of examinations to coursework and
- explicitness of goals, criteria and standards.

For the purposes of the audit, summative assessments are defined as graded tasks which contribute towards the degree classification. Formative assessment refers to compulsory tasks which elicit feedback, without allocating marks. This is in line with research by Sadler (1989) and Black and Wiliam (1998) who maintain that grading limits students' capacity to learn from feedback by shifting focus to the mark rather than the developmental aspects of feedback. Oral feedback has been quantified by estimating the proportion of students taking up formal tutorial opportunities about assessed work, and the frequency and length of these tutorials. The amount of written feedback students receive is quantified by sampling cover sheets and on-script comments.

Assessment Experience Questionnaire

The second research method was AEQ Version 3.3 (Gibbs and Dunbar-Goddet 2012) which has been used widely to measure how students experience various conditions of learning. The AEQ consists of 28 statements, including both original items, as well as items from two scales of the Course Experience Questionnaire and the Approaches to Studying Inventory (see Gibbs and Dunbar-Goddet 2012). Students respond on a scale from strongly disagree (1) to strongly agree (5) with higher scores indicating a higher presence of the construct measured by the scale. In line with the analysis of the National Student Survey, the construct of a 'good score' on the AEQ is one where students agree or strongly agree with items in the scale (= and > 4). This applies to all the scales except 'Surface Learning' where a good score is (= or < 2).

An exploratory principal component factor analysis was conducted in order to check the factor structure of the AEQ questionnaire on the present sample of participants. The eigenvalue rule indicated eight factors which together explained 60.08% of variance in the data. A varimax rotation was performed specifying an eight factor solution summarised in Table 2. A few items loaded differently as in the subscales of the original AEQ. Due to reliability issues, further analysis was conducted on the following four factors: Appropriate Assessment; Learning from Examinations; Clear Goals and Standards and Use of feedback.

Data analysis

Data from the programme audit ($n = 18$) were analysed and clustered into disciplinary groups. The researchers amalgamated data from the categories 'hard pure' and

Table 2. AEQ factor analysis summary.

No.	Factor	Eigenvalue ^a	% of variance explained ^a	Cronbach's α
(1)	Appropriate Assessment (AA)	4.37	15.62	.72
(2)	Learning from Examination (LE)	2.14	7.65	.84
(3)	Clear Goals and Standards (CGS)	1.88	6.71	.68 ^b
(4)	Use of Feedback (UOF)	1.58	5.65	.71

^aBefore rotation.

^bValues represent Cronbach's α if one item is omitted.

Table 3. Disciplines under study.

Humanities	Sciences	Professional courses
Theology Religious Studies	Pharmacy	Teaching \times 2
History \times 2	Computing	Nursing
Philosophy	Web Systems	Social Work
Politics	Geography	Law
English	Psychology	
American Studies		
Media Studies		

'hard applied' because the sample contained an insufficient number of pure science subjects. This was justified as there are sufficiently similar assessment emphases (factual retention and memorisation) and methods (examinations, multiple choice quizzes, and practical and laboratory work) between the applied and pure domains enabling this data to be used in combination (for similarities see Table 1). Geography and Psychology were included in the Sciences group, as in both cases the degrees in question were Bachelor of Science degrees with a sufficiently strong scientific orientation to be regarded as sciences rather than social sciences. Table 3 indicates the breakdown of the discipline clusters in our sample. Where data have been collected from two of the same disciplines at different universities, this is shown, for example History \times 2.

The results of the AEQ questionnaire were analysed using the Statistical Package for Social Sciences. The total sample of 762 students was split by discipline group according to the study programme they attended: Humanities ($N=250$), Sciences ($N=239$) and Professional Courses ($N=273$). The average scores of these three groups on AEQ scales were compared using one-way analyses of variance (ANOVA).

Sampling

The researchers initially sampled using a 'purposive' approach (Lincoln and Guba 1985), with participants selected to give rich and complex insights related to the themes of the research. The researchers sampled seven different discipline fields in the four original partner universities. The partners were small universities (5000–8000 students) with similar disciplinary provision, mainly arts, social sciences,

humanities and professional courses. These could be categorised as ‘teaching-oriented’ universities. The sample expanded through word-of-mouth to a further five programmes at partner universities and to six new programmes at four other universities, three of which were large research-intensive universities. This was an example of snow-ball sampling.

The researchers distributed hard copy AEQ surveys to mainly final year students in the same programmes that participated in the programme audit. Surveys were distributed during lectures and seminars with the consent of students and with the permission of lecturers, who recused themselves during the process of data collection. Ethical clearance was obtained for conducting all aspects of the research. This assured the anonymity of students and academics, the confidentiality and anonymity of the data, and the right to withdraw from the research at any stage.

Findings

The first section provides a descriptive overview from the programme audit about the assessment environment of each discipline’s group. The second consists of a statistical analysis of AEQ data about students’ perceptions of their assessment environment and their learning strategies.

Comparing disciplinary practice: the programme audit

The audit variable which was most consistent across disciplines was the average number of days it takes to return feedback (Humanities 23 days; Sciences 20 days and Professional Courses 23 days). The high level of consistency relates to institutional policies about returning marks and feedback to students, usually within three weeks. However, discipline-group averages masked much wider ranges within the 18 programmes sampled – so for example, the quickest programme to return feedback was a science programme (10 days), and the slowest was in the humanities group (35 days). Three out of the five science programmes took less than three weeks to return feedback; and two out of five took less than two weeks. Prompt return of feedback aligns with the idea of a ‘paradigmatic’ discipline where there is greater consensus and less interpretation in marking assessed tasks, allowing for speedier return times and potentially less written feedback to explain and justify grades.

Discipline groups employed a similar range of assessment methods (Science $n = 15$; Professional $n = 14$ and Humanities $n = 11$), but there was high internal variation within Humanities: for example, two distinct History courses employed as few as 7 different assessment methods, or as many as 17. Humanities and Science courses had similar means for the number of summative assessment tasks (42 and 43, respectively), with professional courses having fewer summative points, averaging at 32.

There was high variation in the number of formative assessment tasks students undertook. A student taking a Science course could expect to have three formative tasks to every one that a student on a Professional course encountered, with a similar ratio for Humanities’ students. Given that science assessments emphasise memorisation, factual retention, applying course content and solving logical problems, it is unsurprising that formative assessment finds a place in helping students to master knowledge, content and factual material in cumulative and sequential ways through the programme, with further scope for laboratory work to help students apply their knowledge.

Table 4. Comparative patterns of assessment for a three-year undergraduate degree.

Audit variables	Humanities (mean)	Sciences (mean)	Professional (mean)
Total number of assessments	54	74	42
Number of summative assessments	42	43	32
Number of formative assessments	12	31	10
Varieties of assessment (<i>n</i>)	11	15	14
Proportion of examinations (%)	14.3	30.6	15
Time to return feedback (days)	23	20	23
Amount of oral feedback	3 hours 17 minutes	4 hours 56 minutes	10 hours 33 minutes
Amount of written feedback (words)	7382	3615	7040

Two areas of major variation in the audit were the amount of oral and written feedback students could expect to receive. Students on Professional courses could expect to receive more than double the quantity of oral feedback (10 hours 33 minutes) compared with Science students (4 hours 56 minutes), and triple the quantity of oral feedback when compared to Humanities' students (3 hours 17 minutes). Science students receive about half as much written feedback (3615 words) when compared to their peers in the Humanities (7382) and on Professional courses (7040). The brevity of written feedback in the Sciences may be explained by the nature of the discipline as 'paradigmatic' or 'convergent' (Biglan 1973b; Becher 1989; Pryor and Crossouard 2008, 2010) with more agreement about what constitutes a correct answer and therefore less need for extensive written feedback. It may be explained further by the widespread practice of not giving any feedback on examinations except by special request at most UK universities, as Science students have double the proportion of examinations compared to other discipline groups.

Table 4 summarises comparative patterns of assessment across the three discipline groups from the 18 audits we conducted.

Comparing learning responses across disciplines: the AEQ

Data from 762 AEQ returns were split by discipline group and analysed to find whether there were any statistically significant differences between the groups. Tests of differences were performed by conducting a series of one-way ANOVA on the four AEQ factors that were found reliable in previous analyses (Table 2). ANOVAs revealed significant differences between disciplines on two of the four AEQ factors used in the analysis (Table 5). Bonferroni *post hoc* tests of difference between pairs of groups indicated that students of humanities evaluated the appropriateness of the assessment (AA) they receive as significantly lower than students of the other two disciplines. Students of professional courses, on the other hand, were significantly less clear about goals and standards (CGS) than students of the two other disciplines.

Table 5. Means and standard deviations (in brackets) of AEQ scores among disciplines and one-way ANOVA tests of differences between them.

AEQ subscale	Humanities	Sciences	Professional courses	$F(2, 808)$	p
AA	2.14 (0.69) ^a	2.67 (0.84) ^b	2.78 (0.93) ^b	43.29	<.000
LE	3.12 (1.06)	3.32 (0.98)	3.18 (1.02)	2.46	.09
UOF	3.80 (0.67)	3.72 (0.82)	3.80 (0.70)	1.02	.36
CGS	3.75 (0.73) ^a	3.62 (0.76) ^a	3.41 (0.84) ^b	12.48	<.000

Note: Different superscripts indicate differences significant at $p < .001$.

Discussion

Three main findings from the statistical analysis are discussed. The first was that students across the three discipline groups all derived similar learning benefits from examinations. This is surprising when weighed against the audit evidence, as you would expect science students to learn more from examinations, given that they encounter twice the proportion compared to the other discipline groups. The second finding was that humanities students found their assessment less appropriate than other discipline groups, indicating that the assessment pattern encouraged surface approaches to learning. The relatively low proportion of examinations in the humanities group (14.3%), usually a culprit in driving surface learning, prompts consideration of other factors. The third finding was that students on professional courses were least able to understand the goals and standards of assessment tasks, compared to peers in other disciplines. This was contrary to audit evidence indicating high volumes of oral feedback, which is usually associated with clarifying goals and standards.

In seeking explanations, the researchers have triangulated statistical findings with audit data, exploring contradictions and confirmations. Second, where appropriate, the authors distil explanations from the research literature about assessment and feedback and about discipline groups. This combination of approaches provides the basis for our explanations for the main statistical findings. These explanations are hypothetical, inviting further research to contest or confirm their validity.

Explanations for similar levels of learning from examinations

Explanation 1: examination 'conditions' may override disciplinary differences

Audit data show that students on science programmes encountered twice the proportion of exams, yet experienced similar learning benefits to their peers. The literature suggests that examinations, with their artificial, time-bound, stress-inducing conditions and (mostly) unpredictable content, tend to drive students to pursue more strategic ways of 'playing the game' to achieve good marks (Miller and Parlett 1974). 'Playing the game' includes 'spotting' topics, scanning archival records of examination papers and pursuing cues and clues from lecturers, both in and out of lectures. No matter how many examinations students experience, minds are concentrated on 'just learning the material so that they perform in the exams' instead of 'immersing themselves in the topics' (Newstead 2002, 71). In part, this may be the result of timetabling examinations in quick succession over short-time periods, leading to cramming and 'selective negligence' (Miller and Parlett 1974). At a deeper level, students perceive the artificiality of examinations, and their distance from more authentic modes of assessment, and therefore treat them as a game to be played (Miller and Parlett 1974). The 'deep structure' of

exams (e.g. unknown questions, timed conditions, handwritten papers, anonymity and 'high-stakes' assessment), all may powerfully influence student perceptions, irrespective of discipline. The same conditions and 'deep structure' pertain for all examinations, which may explain students' similar responses to questions about learning gains across the disciplines.

Further research into why students perceive similar learning benefits from examinations across the disciplines could usefully entail a content analysis of examination questions and formats, comparing disciplinary practices. This would help to develop knowledge and understanding as to whether examinations use generic approaches or vary according to discipline. A second research avenue could pursue an empirical investigation of student strategies and perceptions across disciplines in relation to 'playing the examination game', directly measuring and comparing student tactics by discipline.

Explanation 2: feedback on examinations is minimal across disciplines, leading to similar learning gains

Until recently, it has been very unusual for students to receive any feedback on examinations except by direct request. The timing of examinations tends to be at the end of semesters and year end, with marking completed over holiday periods. As a rule, most programmes do not give any written feedback on examinations, although recently there has been pressure from the National Student Union's Charter on Assessment and Feedback (2010), which refers obliquely to the absence of feedback on examinations in Principle 3: 'Receiving feedback should not be exclusive to certain forms of assessment.' Many universities are now revisiting the question of giving feedback on examinations, in response both to the NUS Charter and to student demands based on increased fee regimes. Many lecturers are concerned that feedback on examinations will be used by only a small minority of students because of its timing after module completion and beyond the holiday period. At the time of this study, the audit showed that students received minimal feedback on examinations, contributing to the explanation that learning gains associated with feedback were similar across the discipline areas.

Future studies into feedback on examinations might pilot different and earlier examination times with feedback given to students within the bounds of the semester. Using an experimental design, researchers could capture student data comparing the learning effects of examinations which elicit feedback within the academic year compared to examinations with no feedback at the end of semester or year. More radically, researchers could pilot 'dry run' examinations within the semester, eliciting peer and/or tutor feedback, or even being peer marked with criteria. This feedback could feed forward to the final examination, and the learning effects from examinations could similarly be measured and compared using an experimental design.

Explanations for lower perceptions of appropriate assessment in the humanities

Explanation 1: there is less authentic assessment in the humanities

Authentic assessment simulates 'real-world' tasks, and is characterised by high levels of challenge and complexity, with progression, sequencing and integration of learning across various elements of a whole programme promoting a deeper approach to

learning (Meyers and Nulty 2009, 567). One of the hallmarks of authentic assessment, linked to the workplace, is that students are involved in self-evaluation and critical reflection. Authentic assessment motivates students and nurtures deep learning through fostering connections across whole programmes of study, and with wider 'real-world' domains; it has profound overlaps with the AEQs concept of 'appropriate assessment'.

The programme audit data showed a slightly lower variety of assessment in the Humanities ($n = 11$), with the most commonly used types of assessment being essays, examinations, critical commentaries and presentations. Of these, only presentations have real-world relevance. In comparison, professional courses contained simulated mock trials in law; oral panels and case studies in social work, classroom teaching and observation, and lesson planning in teacher education, portfolios in both education and social work, and Objective Structured Clinical Examinations in Nursing, as well as professional reflections and presentations in most professional domains. There was also evidence of authentic assessment in the sciences with laboratory based work, for example. While the theoretical and abstract nature of humanities' subjects may not lend itself easily to the authentic types of assessment indicated earlier, there are two areas which could strengthen appropriate assessment in the humanities. The first is within the metacognitive area of developing critical reflection and self-evaluation skills (Ashford-Rowe, Herrington, and Brown 2014, 6); the second is in creating stronger linkages and integration across modules in the whole programme to encourage more synthesis, challenge and higher order thinking from students and to sharpen the concept of progression. Both approaches would develop deep learning among students (Meyers and Nulty 2009; Ashford-Rowe, Herrington, and Brown 2014).

Further research might use content analysis of authentic assessment in other discipline groups which use 'real-world' assessment types, adapting and trialling variations of these in the humanities and determining learning effects. Another research strategy might pilot critical reflection and self-evaluation tasks on selected humanities' programmes and measure effects in comparison to programmes without the same reflective processes.

Explanation 2: 'negative backwash' effect in the Humanities

In the audit data, the Humanities group has the lowest ratio of formative to summative (1:4) and the least oral feedback. Formative assessment, which is not graded, gives students low risk opportunities to be creative and explore alternative ways of constructing knowledge. Both formative tasks and oral feedback tend to be the most collaborative and relational aspects of assessment, prompting reflection and developing 'real-life' skills in students which lead to deep learning (Ashford-Rowe, Herrington, and Brown 2014). The relative lack of oral and collaborative practices, combined with high reliance on essays and text-based formats of assessment which are contained within modular units, may lead to a negative 'backwash effect'. Biggs (2003) describes the 'backwash effect' as being a consequence of assessment design on what students pay attention to, and how they approach learning. A negative backwash effect may lead to reproductive and surface approaches to learning, where students reproduce material fairly mechanically in essays, or place undue emphasis on factual recall in examinations, for example. In the case of humanities, students are repeatedly required to write essays and critical commentaries on narrow modular topics which have little connection or integration with other modules or application to the wider world. The

'backwash' effect may be that students reproduce well-worn formulae for essays, without being challenged to synthesise or connect cross-modular subject domains or to apply knowledge to different contexts. In professional courses, students are regularly required to reflect on the process of learning through portfolios and professional reflections, deepening their engagement with knowledge and context. This was lacking in most of the humanities courses.

Further research might entail piloting more portfolio tasks to determine to what extent these develop students' reflective skills; this could be investigated through the use of appropriate assessment items on a questionnaire, or through qualitative methods on inquiry. A second useful research exercise could pilot more integrated assessments which draw on material from across modules and/or within possible 'real-world' careers, to measure the impact on students' appropriate assessment and deep learning scores.

Explanations for professional students' lack of clarity about goals and standards

Explanation 1: the nature of oral feedback does not help students to reflect on goals

Audit data showed that students on professional courses received an average of more than 10 hours of oral feedback on a three-year degree, averaging at more than three hours per year of dialogue with lecturers or practitioners. This is a high volume of oral feedback, but there may be a gap between the quantity of feedback, and its nature and value for students' development (Blair and McGinty 2013, 468). Akcan and Tatar (2010) examined the content of oral feedback in supervision conferences between student teachers and their mentors and found that there were two main approaches: prescriptive or collaborative. Prescriptive oral feedback was directive, telling student teachers what they had done wrong and how to improve using practical advice. This form of feedback did not help students to reason and reflect about their practice and had limited capacity for clarifying goals and standards. Collaborative feedback was characterised by helping student teachers to reflect and reason and to 'build a bridge between their theoretical knowledge base and their practical experience at the schools' (Akcan and Tatar 2010, 154). Reflective feedback which questions students and develops links between theory and practice is much more likely to help students clarify goals and standards than expert-driven, teacher-centred advice. One explanation, therefore, is that the kinds of oral feedback students experienced may have been prescriptive and therefore diminished their capacity to clarify goals and standards. Further research which used content analysis of feedback conversations between students and mentors might help to develop these categories, analyse power relationships and distil principles for helpful oral feedback.

Explanation 2: university academics and practitioners have differing goals and standards

Students on professional courses are usually required to undertake work placements, where they are likely to receive immediate and plentiful feedback on their performance. Feedback dialogues between practitioners and students are opportunities for goals and standards to be clarified, but they may confound students if academic (university) and

practice (placement) goals differ. In the social work context, the problems of fair and standardised assessment are acknowledged: ‘there have been doubts raised as to the ability of practice teachers to competently perform their assessment duties’ (Crisp, Green Lister, and Dutton 2006, 726). This has been compounded by the introduction of service-users in assessing ‘fitness to practice’ among social work students, where there is a tendency for ‘niceness’ to prevail over criticality: ‘clients may be overly positive about the student and consequently service user feedback may reflect a student’s popularity rather than their competence’ (Crisp, Green Lister, and Dutton 2006, 729). In teacher education, differences surface around practising teachers being more preoccupied with classroom management as the measure of effective teaching in contrast to university teachers’ focusing on the quality of activities and pedagogic interactions. A study of practice and university teacher feedback dialogues in Turkey suggested that these differences ‘could be explained by the difference of the two parties in their understanding of what effective teaching is’ (Akcan and Tatar 2010, 165). Variations in understanding of goals and standards across university-placement contexts may be responsible in part for students’ confusion about goals and standards.

Future research might compare staff perceptions of goals and standards across the academic–practitioner divide, using interviews or focus groups, and determining the sources of confusing messages to students, in order to address them. Qualitative data from students about conflicting goals and standards, through interviews, focus groups or participatory analysis of assessment feedback, would help to reveal any inconsistent messages and could lead to improvements in the consistency of goals and standards between academics and practitioners.

Explanation 3: professional students need more formative tasks to clarify goals and standards

Professional students have a third of the formative assessment opportunities of their science peers, and slightly more formative assessment than students in the humanities, with an average of 10 formative tasks over three years, and a ratio of 1 formative:3 summative tasks. Given that professional students experience high varieties of assessment (an average of 14 varieties over 32 summative points), it is not surprising that they are confused about goals and standards. This is particularly the case within professional courses requiring reflective journals/logs, which students find difficult to negotiate in the relative isolation of the placement, and with the sophisticated balance required between the professional and the personal. The case for more low-stakes, formative assessment, to help students clarify the standard aimed at has been well argued in the literature as a powerful tool which short circuits ‘trial-and-error learning’ in understanding goals and standards (Sadler 1989, 120). Further research which piloted assessment patterns with a higher ratio of formative to summative assessment, and compared effects on student learning to results from this study, for example, would be helpful.

Conclusion

This paper has taken a data-driven approach to exploring disciplinary differences between traditional categories established by Becher (1989). It has emphasised the role of discipline groups in shaping learning and teaching by examining one area of practice, namely assessment and feedback, using the Programme Audit, and the AEQ.

One limitation relates to the nature of our data which is numerical and statistical. More fine-grained qualitative data in future studies would deepen the comparative analysis. In the discussion we have pursued lines of explanation for the two problematic areas of difference, using audit data and assessment literature to support our theories. Each of these leading theories beckons further research, inviting more textured, qualitative data from students, lecturers and practitioners to provide explanation and/or counter-evidence to the theories.

Our paper has revealed difficulties in two discipline groups' learning from assessment and feedback. The potential to draw on cross-disciplinary practice to achieve the aim of better assessment environments for students is high, particularly given student perceptions of more appropriate assessment in science and professional courses, and better understandings of goals and standards within science and humanities courses. Factors facilitating learning are contained within each of these distinct disciplinary community's assessment environments, suggesting further, detailed exploration and research.

An area of similarity was students' common perception of learning from examinations across all three discipline groups. The discussion explored reasons for this, including the structure of examinations, and lack of feedback. Comparative research on the nature of examinations across the disciplines, investigating alternative and authentic 'real-world' examinations, which encourage deep learning, would be valuable. The policy of not giving feedback on examinations is currently being challenged by the UK's National Union of Students, but discussion about policy changes invites deliberation about the timing of examinations if giving feedback is not to become a rhetorical exercise.

The central question guiding our research was whether disciplinary differences in assessment and feedback practice *matter* for student learning. Our analysis has shown three problematic areas of assessment and feedback practice among disciplines: first, recurrent practice of examinations does not necessarily improve student perceptions of learning from examinations; second, students' perceptions of the appropriateness of their assessment in the humanities is lower than that in other discipline groups and finally, professional students have more difficulties in clarifying goals and standards related to their assessment – the 'gold standard' of what counts as 'good' academic practice. We have focused on these three areas in need of further research, rather than on the positive messages: parity of learning from examinations across the disciplines; significantly better perceptions of the appropriateness of assessment in the sciences and professional domains; significantly better concepts of goals and standards in the sciences and humanities.

Acknowledgements

This research was made possible by an award of funding by the Higher Education Academy to conduct research on programme-wide assessment through the 'Transforming the Experience of Students through Assessment' (TESTA) National Teaching Fellowship project. The underlying research materials for this article can be accessed at www.testa.ac.uk.

References

- Akcan, S., and S. Tatar. 2010. "An Investigation of the Nature of Feedback Given to Pre-service English Teachers During their Practice Teaching Experience." *Teacher Development: An International Journal of Teachers' Professional Development* 14 (2): 153–72.

- Ashford-Rowe, K., J. Herrington, and C. Brown. 2014. "Establishing the Critical Elements that Determine Authentic Assessment." *Assessment & Evaluation in Higher Education* 39 (2): 205–22.
- Becher, T. 1987. "Disciplinary Discourse." *Studies in Higher Education* 12 (3): 261–74.
- Becher, T. 1989. *Academic Tribes and Territories*. 1st ed. Maidenhead: SRHE and Open University Press.
- Becher, T. 1994. "The Significance of Disciplinary Differences." *Studies in Higher Education* 19 (2): 151–61.
- Becher, T., and P. Trowler. 2001. *Academic Tribes and Territories: Intellectual Enquiry and the Cultures of Disciplines*. Buckingham: Open University Press.
- Biggs, J. 2003. *Teaching for Quality Learning at University*. Buckingham: SRHE and Open University Press.
- Biglan, A. 1973a. "The Characteristics of Subject Matter in Different Academic Areas." *Journal of Applied Psychology* 57 (3): 195–203.
- Biglan, A. 1973b. "Relationships between Subject Matter Characteristics and the Structure and Output of University Departments." *Journal of Applied Psychology* 57 (3): 204–13.
- Black, P., and D. William. 1998. "Assessment and Classroom Learning." *Assessment in Education: Principles, Policy and Practice* 5 (1): 7–74.
- Blair, A., and S. McGinty. 2013. "Feedback-dialogues: Exploring the Student Perspective." *Assessment & Evaluation in Higher Education* 38 (4): 466–76.
- Crisp, B. R., P. Green Lister, and K. Dutton. 2006. "Not Just Social Work Academics: The Involvement of Others in the Assessment of Social Work Students." *Social Work Education: The International Journal* 25 (7): 723–34.
- Fanghanel, J. 2009. "The Role of Ideology in Shaping Academics' Conceptions of their Discipline." *Teaching in Higher Education* 14 (5): 565–77.
- Gibbs, G., and H. Dunbar-Goddet. 2009. "Characterising Programme-level Assessment Environments that Support Learning." *Assessment & Evaluation in Higher Education* 34 (4): 481–89.
- Gibbs, G., and H. Dunbar-Goddet. 2012. "A Research Tool for Evaluating the Effects of Programme Assessment Environments on Student Learning: The Assessment Experience Questionnaire (AEQ)." Accessed November 29, 2012. <http://www.testa.ac.uk/resources/publications>
- Gibbs, G., and C. Simpson. 2004. "Conditions under which Assessment Supports Students' Learning." *Learning and Teaching in Higher Education* 1 (1): 3–31.
- Hattie, J. 2009. "The Black Box of Tertiary Assessment: An Impending Revolution." Accessed October 12, 2013. <http://ako.aotearoa.ac.nz/ako-aotearoa/ako-aotearoa/resources/pages/black-box-tertiary-assessment-impending-revolution>
- Jessop, T., Y. El Hakim, and G. Gibbs. 2011. "TESTA: Research Inspiring Change." *Educational Developments* 12 (4): 12–16.
- Jessop, T., Y. El Yakim, and G. Gibbs. 2013. "The Whole is Greater than the Sum of its Parts: A Large-scale Study of Students' Learning in Response to Different Programme Assessment Patterns." *Assessment & Evaluation in Higher Education*. Published online 29 April 2013. <http://dx.doi.org/10.1080/02602938.2013.792108>
- Jones, A. 2009. "Redisciplining Generic Attributes: The Disciplinary Context in Focus." *Studies in Higher Education* 34 (1): 85–100.
- Kolb, D. 1981. "Learning Styles and Disciplinary Differences." Chap. 10 in *The Modern American College*, 232–256. San Francisco: Jossey-Bass.
- Krause, K. L. 2014. "Challenging Perspectives on Learning and Teaching in the Disciplines: The Academic Voice." *Studies in Higher Education* 39 (1): 2–19.
- Lattuca, L. R., and J. S. Stark. 1994. "Will Disciplinary Perspectives Impede Curricular Reform?" *The Journal of Higher Education* 65 (4): 401–26.
- Lincoln, Y. and E. Guba. 1985. *Naturalistic Inquiry*. Beverly Hills, CA: Sage.
- Lindblom-Ylänne, S., K. Trigwell, A. Nevgi, and P. Ashwin. 2006. "How Approaches to Teaching Are Affected by Discipline and Teaching Context." *Studies in Higher Education* 31 (3): 285–98.
- Meyers, N. M., and D. D. Nulty. 2009. "How to Use (five) Curriculum Design Principles to Align Authentic Learning Environments, Assessment, Students' Approaches to Thinking and Learning Outcomes." *Assessment & Evaluation in Higher Education* 34 (5): 565–77.

- Miller, C. M. L., and M. Parlett. 1974. *Up to the Mark: A Study of the Examination Game*. London: Study for Research in Higher Education.
- Neumann, R. 2001. "Disciplinary Differences and University Teaching." *Studies in Higher Education* 26 (2): 135–46.
- Neumann, R., S. Parry, and T. Becher. 2002. "Teaching and Learning in their Disciplinary Contexts: A Conceptual Analysis." *Studies in Higher Education* 27 (4): 405–17.
- Newstead, S. 2002. "Examining the Examiners: Why Are We so Bad at Assessing Students?" *Psychology Learning and Teaching* 2 (2): 70–75.
- NUS (National Union of Students). 2010. "NUS Charter on Feedback and Assessment and Ten Principles." Accessed October 17, 2013. <http://www.nusconnect.org.uk/asset/news/6010/FeedbackCharter-toview.pdf>
- Pryor, J., and B. Crossouard. 2008. "A Socio-cultural Theorisation of Formative Assessment." *Oxford Review of Education* 34 (1): 1–20.
- Pryor, J., and B. Crossouard. 2010. "Challenging Formative Assessment: Disciplinary Spaces and Identities." *Assessment & Evaluation in Higher Education* 35 (3): 265–76.
- Ramsden, P. 1992. *Learning to Teach in Higher Education*. London: Routledge.
- Sadler, D. R. 1989. "Formative Assessment and the Design of Instructional Systems." *Instructional Science* 18 (2): 119–44.
- Shulman, L. S. 1993. "Teaching as Community Property: Putting an End to Pedagogical Solitude." *Change: The Magazine of Higher Learning* 25 (6): 6–7.
- Shulman, L. 2005. "Signature Pedagogies in the Professions." *Daedalus* 134 (3): 52–59.
- Smeby, J. C. 1996. "Disciplinary Differences in University Teaching." *Studies in Higher Education* 21 (1): 69–79.
- Wareing, S. 2009. "Disciplines, Discourse and Orientalism: The Implications for Postgraduate Certificates in Learning and Teaching in Higher Education." *Studies in Higher Education* 34 (8): 917–28.
- Warren Piper, D., D. D. Nulty, and G. O'Grady. 1996. "Examination Practices and Procedures in Australian Universities." Higher Education Division, Department of Employment Education and Training. Evaluations and Investigations Program. AGPS, Canberra.