

Variations in higher education portfolio assessment

Discussion of quality issues based on a survey in Norway across institutions and disciplines

Introduction	1
Background: recent changes in Norwegian Higher Education	2
The 'Bologna process' – what does it involve?	2
The Quality Reform of higher education in Norway	2
Why did portfolio assessment 'take off' after the Quality Reform?	3
A survey study of portfolios in four HE institutions	4
Method	4
Key findings	5
Discussion	9
Summary of survey findings and our interpretations	10
Quality issues regarding feedback	12
The grading process and the use of explicit assessment criteria	14
Conclusion	15
References:	16

Introduction

While 'portfolio' was an unknown concept for most teachers and students in higher education in Norway five years ago, a nationwide survey in 2004 showed that portfolio assessment is now used in all types of educational institutions and across disciplines. This is surprising since the established view of assessment says that it is more difficult to change than any other aspect of education. So what had happened? In the first section of this article we will briefly sketch the background for the changes in assessment in higher education in Norway necessary to situate the research study. In the second section we present findings from a survey in four institutions that documents the popularity of the portfolio as well as great variations in portfolio practices. In the third section we discuss a few of the many quality issues related to portfolio assessment in our context. The first question we raise is what concepts of portfolio the variations in our data signal, and how much variety is possible to sustain for the term 'portfolio assessment' to be meaningful and useful (across disciplines, institutions, countries and stakeholders). We particularly focus on whether reflective texts are a defining element of all portfolios. Second we discuss feedback and quality issues involved in teacher and peer feedback, and different stakeholders' perspectives (White & Ostheimer, 2006). The third issue we raise is the grading process, particularly the use of explicit assessment criteria in relation to quality.

Although a major finding is that there is a great variety in portfolio concepts and use, there are some common characteristics across institutions and disciplines: the portfolios are 'disciplinary content portfolios' used for assessment at the end of a course, but the learning aspect is strongly emphasized and feedback is seen as important. All portfolios consist of written texts, but students' content knowledge and understanding are the primary focus of assessment, not their writing competency, although this indirectly will affect their grade.

Background: recent changes in Norwegian Higher Education

Norway is a country with only 4.5 million inhabitants. It has six state universities, five scientific colleges, 25 state colleges (which call themselves ‘university colleges’) and 26 private colleges. The majority of students are in the state system. Higher education has recently undergone a major reform, called ‘the Quality Reform’, which forms the background for the changes in assessment practices in general and portfolio assessment in particular. Since all European countries are now heavily influenced by the Bologna process we will first give a brief overview of what this involves. Changes in the higher education sector in Europe can only be understood in light of the Bologna Declaration.

The ‘Bologna process’ – what does it involve?

When 16 European education ministers met in Bologna in 1999 to discuss a common European education policy for the future, few had foreseen the consequences. The Bologna Declaration is not a treaty that is ratified by parliaments or signed by the governments that were involved in formulating it. Nevertheless it has already exerted considerable influence on educational policies in many European countries. Its clear goal is the creation of a coherent European Higher Education by 2010 in order to ensure mobility within Europe and to make Europe more competitive in the international arena. The objectives of the Bologna Declaration are specific:

- a common frame of reference for comparing diplomas from all the European countries
- an alignment of programs at undergraduate, graduate and postgraduate level: 3 year Bachelor + 2 year Bachelor, followed by 3 year PhD
- implementation of the European Credit Transfer System (ECTS) ¹
- quality assurance systems
- student and teacher mobility

The Quality Reform of higher education in Norway

The recent reform of Norwegian higher education was strongly influenced by the internationalization in the higher education sector in general and the Bologna Declaration in particular. Norway, although not a member of EU, has been in the forefront of implementing the Bologna principles.² The reform which formally was introduced through Stortingsmelding-27 (2000-2001)(Parliament Proposition 27/2001), is comprehensive and represents an attempt to achieve a higher degree of efficiency through devolution of authority to the higher education institutions, the provision of stronger leadership, increased emphasis on internationalisation and the formation of an agency for quality assurance and accreditation. A new study structure and grading system are introduced, as well as new pedagogical designs and a new funding model that is supposed to provide stronger incentives for improvement.

¹ The European Credit Transfer System (ECTS) is the EU system for transfer of study credits and grades between countries. The system is meant to supplement, not replace national systems, and plays an important role in creating mobilization between European institutions and create a European education area: (http://www.europa.eu.int/comm/education/programmes/socrates/ects_en.html).

² The University of Bergen hosted the third conference of European education ministers in June 2005.

The Bachelor/Master study structure (3+2 years) was implemented at all levels of the Norwegian universities, scientific colleges and state colleges in the autumn of 2003.³

The new study structure represents a radical break with many of the traditions in Norwegian higher education. It affects both structure and length of undergraduate and graduate studies, our assessment system, teaching, supervision and student learning. Norwegian students will now get their Bachelor's degree in 3 instead of 4 years, the new credit point system (in line with ECTS) is introduced and our grading system has changed from a very detailed numerical scale to a letter scale (ABCDEF). All courses are modularised (most courses are 10 or 15 ETCS) and the use of external assessors at undergraduate courses is reduced. New types of courses are created, although many of the new programmes build upon the old ones.

The pedagogical expectations of the reform were clearly formulated in the official documents (i.e. NOU 2000 and the Parliamentary Proposition 27/2001), and can briefly be summarized as follows: 1) more use of teaching methods requiring active student participation, 2) closer follow-up of each student and regular feedback on papers, 3) closer connection between teaching and assessment, 4) more emphasis on formative assessment and alternatives to traditional exams, for instance, portfolio assessment, 5) increased use of information and communication technology. Educational institutions also have to make agreements or contracts with students concerning courses, clearly outlining the rights and responsibilities of the institution and the student in relation to each other. These measures are clearly in line with international trends in higher education.

It has been a major political concern that the Quality Reform should lead to better teaching and learning. The first report from the evaluation research group of the Quality Reform indicates that change in assessment forms is one of the most salient features of the pedagogical changes, particularly the introduction of portfolio assessment and continuous assessment (Michelsen & Åmodt, 2006). Increase in compulsory student writing and more regular feedback to students are to some extent interconnected with the use of portfolios (Dysthe, (in press)).

In this paper we focus on the pedagogical aspects of the Quality Reform. Although these affect the daily lives of teachers and students, the structural changes, the new funding model and the establishment of a new Quality Agency have received far more publicity than changing assessment systems. This article is the first attempt to look in some detail at how portfolios are being conceived and used. This is important in our Norwegian context since portfolios represent a considerable break with our higher education assessment traditions. But it is also of international interest because alternative assessment forms are gaining importance in many countries (Segers *et al.*, 2003), and because of the increasingly global education scene.

Why did portfolio assessment 'take off' after the Quality Reform?

While the central aim of the reform was to align Norwegian higher education with the rest of Europe (as foreseen by the Bologna plan), another important aim was to increase effectiveness and throughput and improve teaching and learning. International educational research has shown that feedback and close follow-up of students are very important factors for their academic success. At Norwegian universities, however, compulsory student writing at the undergraduate level was not common, and in most disciplines students were only required to take a traditional end of term exam (Dysthe, 2003b). The central documents of the Quality Reform therefore advocated a change towards assessment forms that tied instruction

³ Following proposals in a report to the Storting (the Norwegian Parliament) on higher education submitted by the Government in March 2001: (Stortingsmelding-27, 2000-2001)).

and assessment closer together, instead of just monitoring what students had learnt at the end of a course. The traditional sit-down-exam had been heavily criticized from educationalists for encouraging short term learning and bad work habits (C. Gibbs, 1994; Shepard, 2001). The backwash effect of assessment on student learning behaviour is well documented, and 'binge studying' before exams was one well known effect. Norway had also been criticized by OECD in a report from 1997 for spending too much of its resources on exams in relation to resources spent on student learning.

The Quality Reform documents seemed to view alternative assessment as a tool to change teaching and learning practice. The fact that NOU 2000 and the Parliamentary Proposition explicitly and positively mentioned portfolios as an example of alternative assessment gave it legitimacy and is an important reason why many chose to try it out. The concept was more familiar for K12 teachers than in higher education, as there had been quite a few development projects involving portfolio assessment since the first one, the Grønåsen writing portfolio project (Dysthe, 1996). However, portfolios had been introduced in the compulsory introductory courses for new university faculty, and some seminars for higher education faculty had been arranged (for instance at the University of Bergen with Kathleen Yancey). Some small scale development and research projects had also taken place, particularly in teacher education, but generally speaking there has been a notorious lack of systematic research on assessment practices in higher education in Norway, as in most other European countries.

A survey study of portfolios in four HE institutions

In this article we report the findings of a survey of portfolios practices at the second largest university and four university colleges⁴. The purpose is to document portfolio practices across different institutions and disciplines. The questions focused on whether there was a distinction between working portfolios and assessment portfolios, the number of assignments and types of genres in the portfolio, what feedback practices were used, how portfolios were assessed and whether written assessment criteria were used. Teacher attitudes towards the usefulness of portfolios relative to student and teacher workloads were also investigated.

Method

The survey is based on an electronic questionnaire to course leaders at the university colleges of Stord/Haugesund (HSH), Vestfold (HIVE), Bergen (HiB), Sogn og Fjordane (HSF) and the University of Bergen (UiB) in the spring semester of 2005. Since the aim of the survey partly was to map the diversity of portfolio models, we formulated the following selection criteria: All courses in the study catalogues assessed by portfolios. They were identified by the study administrators at the departments and institutions, who provided the lists of courses and leaders that met the criteria. This meant, however, that we have no certainty that all administrators interpreted the criteria in the same way. The gross sample was 288. Comparing the gross selection at the university with the university colleges reveals clear differences. At UiB the sample consists almost exclusively of disciplinary courses, while at the university colleges professional courses in teacher education, health and social profession and engineering dominate the sample.

⁴University College i Vestfold (HiVE), University College Stord/Haugesund (HSH), University College Bergen (HiB) University College Sogn og Fjordane (HSF) and University of Bergen (UiB)

Table 1. Gross selection, net selection and response

Institution	Gross sample n	Net sample n	Response rate
UiB	139 (48 %)	81 (41 %)	58 %
HiB	48 (17 %)	33 (17 %)	69 %
HSH	65 (23 %)	57 (29 %)	88 %
HIVE	33 (11 %)	26 (13 %)	79 %
HSF	3 (1 %)	3 (2 %)	100 %
Total	288	200	69 %

The final response rate was 58 % for UiB and 79 % for colleges.

As can be seen from Table 1 above the net sample consists of 200 respondents and the results can therefore be interpreted with an error margin of $\pm 3\%$ to $\pm 7\%$. It is a special sample since gross sample is identical with what is call the ‘theoretical universe’, i.e. the total number of entities that the question concern (Hellevik, 1999). In this case it is all the course leaders for courses using portfolio assessment in our sample of educational institutions. The sample is only representative for the University of Bergen, and for the university colleges of Bergen, Stord Haugesund and Vestfold, not for Norwegian universities and university colleges in general.

The sample at the University of Bergen is dominated by full and associate professors while most of the sample at the university colleges consists of lecturers. 53 % of the respondents have taught this course more than 5 years.

Our data shows that in most cases portfolio assessment was introduced in connection with the Quality Reform. At the University of Bergen for instance, portfolio was introduced at 28 % of the courses in the sample in 2003.

Key findings

In this section we will first present results that show variations in portfolio systems, feedback practices, final assessment and type of assessors. We will look at the degree of systematic variation between UiB and the colleges, as well as differences between disciplines and between educations. We will then present answers to the questions about faculty attitudes and experiences with portfolios and perceived learning gains, relative to the workload of teachers and students.

Content and structure of the portfolio

In literature about portfolio it is common to distinguish between a working portfolio or folder where all the work in a certain period of time is collected and a presentation or assessment portfolio, containing the selection of work to be assessed and possibly graded (Allern, 2005; Dysthe & Engelsen, 2003a; Hamp-Lyons & Condon, 2000). This presupposes that the portfolio is kept over a period time long enough for the students to work on a number of assignments. Whether the portfolio system used makes the distinction between a working and a presentation portfolio, may therefore be an indication of substantial differences. This was therefore one of the survey questions, and we have looked at the distribution of responses from universities and university colleges respectively.

Table 2. Differentiating between working portfolio and assessment portfolio? Difference significant. Sig < 0,0005.

Do you differentiate between working portfolio and assessment portfolio?	Yes %	Yes number
Institution		
UiB	28 %	21
University colleges	57 %	64
Disciplinary field		
Math /sciences	16 %	4
Hum, Social Science and Law	35 %	17
Teacher and preschool teacher ed	63 %	50
Health and social worker ed	47 %	7
Engineers	35 %	6
Total	46 %	85

At the University of Bergen only 28 % make a distinction between working and assessment portfolios, while at the university colleges 57 % do. The difference of 29 percentage points is significant. Internally at UiB there is a relatively big difference between on the one hand the humanities, the social sciences and the law faculty with 35 % yes, and the maths-science faculty with only 16 %. At the university colleges teacher education has the highest percentage of making the distinction, then comes nurse education and last engineering. It is difficult to know for sure, however, whether this is a measure of familiarity with the terms or of real differences.

To get more insight in differences we have looked at this question in light of other indicators, first of all what kind of genres do the portfolios contain:

Table 3. Genres of portfolio in % ** = significant at 0,01-level.

Type of work (entries)	Soc+ Hu	Math+Sci	Engineering	Teacher	Health	Total
Expository&argumentative texts**	78	25	28	73	59	62 (120)
Reflection texts**	18	4	22	67	47	40 (77)
Case, project assignments**	10	36	67	50	35	38 (74)
Factual tests**	20	21	56	9	6	18 (34)
Practice related assignments	14	46	44	67	59	48 (93)

Table 3 shows that there are quite large differences between the disciplines regarding which types of work the portfolios contain. Reflective texts are for instance significantly more used in teacher and health educations than in the others. At the social science and humanities faculties students in only 18 % of the courses are asked to write reflective texts for their portfolios, while the corresponding percentage for teacher education at university colleges is 67. This correlates with the division between working and assessment portfolio as the students are usually also asked to reflect over the selection made for the assessment portfolio, if this distinction is made. Together with the question in Table 2 this is a clear indicator of substantial differences in portfolio practices both between educations and between university colleges and the university.

Feedback

Another important aspect of portfolio systems is whether or not students are given feedback and what the feedback practices are like. The results of the question about feedback are unequivocal: 98 % crossed for yes and only 2 % for no. This shows that feedback is a commonality for all courses with portfolio assessment. The picture is more diverse regarding *who* gives feedback and to what extent the comments are made available for others:

Tabell 4. Who gives feedback ** = significant at 0,01-level.

Feedback	Soc +Hu	Math/sci	Engineer	Teacher	Health	Total
Teacher	90	79	94	95	88	91 (176)
Peers**	53	11	33	48	59	43 (84)
Others	8	18	6	6	0	8 (15)
Are comments made available for other students?	58	30	20	40	50	42 (59)
Are students asked to document how they have used the feedback?	14	6	14	25	46	21 (29)

Table 4 shows that in almost all courses the teachers gives feedback during the course. In addition peers give feedback in 43 % of the courses. The question of peer feedback shows significant differences between disciplines, and the tendency is that ‘hard’ subjects (math, science, and engineering) use peer response less than the ‘soft’ disciplines. It is also interesting that 42 % of the total set make teacher comments available for other students

Table 5. Students’ feedback practices

When student feedback is instituted as common practice:	Soc +Hu	Math/Sc	Engineer	Teacher	Health	Total
Have the students got instruction in how to give feedback?	24	0	0	33	56	29 (24)
Is peer feedback compulsory?	54	33	33	50	67	51 (42)
Are peer comments made available to other students?	84	100	80	78	63	80 (62)
Number of respondents	26	3	6	38	9	84

From this table we can see that student comments are made available for peers to a much greater extent than teacher comments. About half have indicated that peer feedback is compulsory, but only 29 % have got any instruction in how to give feedback. None of the differences between the disciplines in table 5 are statistically significant in spite of clear percentage differences. The reason is that the two disciplines that differ most from the average in response profile, maths and engineering, only comprise 3 and 6 respondents.

Final assessment

One of the basic assumptions behind portfolio assessment is that the backwash effect on teaching and learning practices will be positive. It is then important to look at how the final

assessment of the portfolios and the grading is carried out, and what students are assessed on. We needed to get an overview of how the portfolios were assessed in the different systems.

The survey, however, give few clear answers, beyond the fact that there are very different models. Most of them seem to combine assessment of the portfolios with different types of exams, either based on the portfolios or on the curriculum course texts. Regarding grading, 42 % report holistic assessment as basis for the grade, while 35 % say that the grade is based on analytic scoring. There are significant differences between disciplines. At the University of Bergen 50 of the informants in the humanities and social sciences report holistic scoring, while only 13 % in math and sciences. At the university colleges engineering rather surprisingly tops the list of those who use holistic scoring. In teacher education and health and social professions 41 % do the same. Further investigation is needed to interpret these findings.

A central topic in the debate about new assessment forms is the transparency of criteria and the involvement of students in the negotiations of them. This survey only provides information about one aspect: the use of written criteria and whether teachers think this is desirable or not. This was one of the areas of difference between disciplines.

Tabell 6. Are written criteria used for assessing the portfolio? Not answered 9 %.* = significant at 0,05-level.

Are written criteria used?	Soc+Hu	Math+Sc	Engineer	Teacher	Health	Total
Yes	56	35	65	55	87	56 (100)
No	44	65	35	45	13	44 (78)
Total	100	100	100	100	100	100 (178)

56 % report using written criteria when assessing the portfolio. The significant difference is between Math/science discipline where only 35% use written criteria and health sciences where 87% do so.

Assessors

An interesting aspect of assessment practices in higher education in Norway relates to the use of external assessors. Before 2002 all formal exams were required to be assessed by one external assessor in addition to at least one internal. Recent changes in the national law regulating higher education give the institution the right to decide whether or not to use external assessors, and there has been a very steep decrease in such use. Regarding portfolios the survey again shows variations between disciplines. Social sciences and the humanities use external assessors in 80% of the cases, compared to only 36% of Math/science disciplines. At the university colleges it is considerably more common to use more than one internal assessor (59 %) than at the university (27 %).

Table 7. Use of external assessor Not answered: 6 %. ** = significant på 0,01-level.

Do you use external assessor in the assessment of the course?	Soc+Hu	Math/Sc	Engineer	Teacher	Health	Total
Yes	80	36	87	60	50	63
No	20	64	13	41	50	38
Total	100 (46)	100 (28)	100 (15)	101 (79)	100 (16)	101

A table not shown here shows 13 % of respondents using one internal and no external assessor.

Teacher attitudes towards portfolios as a tool for learning

We also asked the respondents a number of questions about how they evaluated the effects of introducing portfolio in their course, and we will briefly summarize some findings because in some respect they show the same pattern of variation. The main finding is that most lecturers (50 to 60 %) thought that the introduction of portfolios had positive consequences for student learning and student motivation for learning. The remaining 40 % to 50 % had not observed changes or were not sure. There are only small variations between the members of the different disciplines in their attitudes towards portfolios, even though there is a certain tendency for engineers to be less positive and for lecturers from the health and social workers' courses to be more positive than the mean.

Since workload has been a hot topic in the wake of the Quality Reform, we asked the teachers how they evaluated student learning in relation to their own workload, and 37 % percent agreed, while 41 % disagreed with the statement: "all taken into account portfolio assessment demands too much work of me in relation to students' learning benefit". The most obvious and direct interpretation of this is that portfolio means extra work and many teachers are not sure whether the pay-off in student learning is worth it. But the response can also be interpreted in relation to the fact that higher teaching work load comes at the expense of research, which has higher status in the academic system. There are significant differences in attitudes between the disciplines regarding learning benefits for students when compared with the workload involved for students or teachers. Engineers are here significant more negative and health and social workers more positive. Whether this indicates that portfolios are more suitable in some disciplines than in others or whether the findings mirror differences in views of how much time should be spent on teaching compared to research, these data cannot tell us.

Discussion

We will limit our discussion to three of the many quality issues related to portfolio assessment:

- A general finding in our survey is that portfolio practices are diverse and that a common framework or understanding seems to be lacking. This gives rise to questions of how much variety is possible to sustain for the term 'portfolio assessment' to be meaningful and useful across different institutions and disciplines and whether different disciplines need to define portfolio differently in order to establish what they think is a high quality assessment system.
- The survey shows that feedback is a common element of portfolio practices, involving both teachers and peers. What quality issues are involved in these feedback practices?
- The use of written criteria, on the other hand, is, according to the survey, not common practice in Norwegian portfolio assessment, even though there seems to be agreement in the international portfolio research community about the necessity of explicit criteria and even rubrics. We discuss this issue from the perspectives of teachers, students and governing bodies.

According to E. M. White the major stakeholders in assessment are teachers, students, researchers and assessment theorists, testing firms and governing bodies (White *et al.*, 1996a:

pp 11-23). Although White writes specifically about writing assessment, his stakeholder groups are general. In Norway, however, we have never had the testing culture that pervades American education and testing firms are no major players in this field. For our purpose we will limit ourselves to teachers, students and governing bodies and the concerns of each of these stakeholders in relation to the issues listed above.

Variations in conceptions and use of portfolios

In this section of the article the underlying questions are: What concepts of 'assessment portfolio' lie behind the varied practices? Are differences determined by different contexts and goals or do they simply reflect a lack of shared understanding of what portfolios are? Should we endorse normative definitions of 'portfolio' as a basis for quality judgements or accept all or some variations as contextually valid?

Summary of survey findings and our interpretations

Central characteristics associated with portfolios in the international literature are documentation of learning over a period of time, genre variety, student choice, evidence of reflection and self assessment and postponed assessment (Davies & LeMahieu, 2003; Hamp-Lyons & Condon, 2000; Klenowski, 2002; Paulson *et al.*, 1991; Yancey & Weiser, 1997). Reflection has been particularly high on the list as a defining quality of portfolios (White & Ostheimer, 2006; Yancey, 1998). This raises the question of normativity and whether these characteristics are equally valid in all educations and disciplines.

The portfolio practices we see behind our survey findings vary from advanced reflection-based models with flexible feedback-practices to portfolios that consist of facts oriented texts, without reflections and with rudimentary feedback procedures. Our data revealed some systematic differences *between the university colleges and the university*: In the university colleges a higher percentage (than in the university) 1) make a distinction between working and assessment portfolio 2) include reflective texts 3) use holistic assessment or use the portfolio as a basis for an oral exam. This may indicate that the differences in portfolio practices are contextually determined and related to specific differences in the two kinds of educational institutions.

Some differences are clearly a consequence of the time frame for the portfolio. In Teacher and Health Ed in university colleges students can collect texts for their portfolios over extended periods of time and even combine several courses, while the courses at the University of Bergen are stand-alone courses, regulated to max 10-15 ECTS (students are expected to take 3 courses or 60 ECTS per semester). The "collection-reflection-selection"-model (Hamp-Lyons & Condon, 2000) which involves a distinction between working and assessment portfolios, makes more sense with a longer time frame. If assignments are substantial and require several revisions, it is common that all of them are included in a course portfolio. The time frame may also affect reflection practices, as development is more visible after a whole year than after a short course. More important, however, is that reflection has been a very central aspect of the training in professional schools for two decades, and therefore portfolio reflective texts fit well into these cultures. Reflection on the theory-practice nexus, for instance, is a central part of the learning process in professional schools, while this is not a prominent feature in most university disciplines.

Our findings also show salient differences between disciplines within the same institution. In the university as well as in the university colleges, 'soft' disciplines have more reflection based and varied portfolio models than the 'hard' disciplines (maths, sciences and engineering). This same tendency goes for peer response, which is less used in hard than in soft disciplines. *Does this indicate that definitions of portfolios need to be deeply situated in*

institutional and/or disciplinary contexts and should we therefore abandon normative statements and overarching requirements for what can be called portfolios? The issue of reflection will serve as an example of what is at stake.

The status and use of reflection and reflective texts in portfolio assessment

There is general agreement in the literature that reflection is crucial to portfolio assessment and that it needs to be built into the portfolio process. The issue here, however, is *whether reflective texts are a defining element of portfolios*. These usually take the form of a cover letter explaining the rationale for the texts in the portfolio and/or a reflective letter about the student's learning process and trajectory of understanding over a certain period of time (Yancey, 1998) or an argument showing how the goals set for the course or program have been met in the portfolio (White, 2005). While such reflective letters seem to be common practice in teacher and health education in Norway, this is not the case at the university and in engineering. Reflective texts are opposed by many university teachers who feel strongly that assessment should be restricted to content learning, understanding and use of concepts and theories as well as students' mastery of thinking and discourse in the discipline. In a study at the Department of History at the University of Bergen, a department that has used portfolio assessment over many years and where the faculty are strongly committed to this form of assessment it was found that although reflection on course content might be strongly focused as a means of fostering critical thinking, the writing of meta-cognitive reflective texts was dismissed by the informants in this study as an intrusion from what they termed as 'psychologically oriented pedagogues' (Dysthe & Tolo, in press). Teachers had tried out the use of reflective texts but found that students wrote very superficial and formulaic texts in this genre, and they abandoned it as part of the portfolio requirement, because the learning benefit of training students to write such texts was not deemed worth the time it would take.

We may disagree with this standpoint, but in our view it would be counterproductive in a context like this to insist on a normative definition of portfolio where a reflective letter is obligatory. Instead, we think it is important to recognize that making the change from final exams to portfolios is in itself a major step forward. At the same time we think it is important to try to engage university teachers in continuous discussions about different ways of conceptualizing the portfolio, about the need for clear goal statements and also about the role reflective writing may play in each disciplinary context, using research evidence from the respective disciplines (White, 2005: pp 587). Quality issues involved in the assessment of reflective letters is a complicated topic in itself that will not be dealt with here.

What role do disciplinary differences play in portfolio variations?

In the international research literature disciplinary differences in portfolios have not been much in focus. A large proportion of the literature deals with portfolios in teacher education or writing (i.e. (Dysthe & Engelsen, 2004; Klenowski, 2003; Yancey & Weiser, 1997; Zeichner & Wray, 2001)), and accounts of portfolio use in the disciplines are often focused on program evaluation (i.e. (Shay, 1999; White & Ostheimer, 2006). (Barrett & Carney, 2005) discuss variations in portfolio concepts in the article: "Conflicting paradigms and competing purposes in electronic portfolio development" and distinguish between portfolios for accountability, for learning and for marketing, a distinction that is not particularly useful for our data set. Their conclusion is that differences in view of learning and differences in purpose of the portfolio determine how it is defined. Other researchers have voiced the same (Callahan, 1999; Klenowski & Askew, 2005; Murphy & Underwood, 2000). Our data indicate that different disciplines favour different approaches to portfolio assessment. It seems likely that there is a relationship between how portfolios are conceptualised and what kinds of knowledge that is privileged, typical ways of working and learning and types of assignments

used in different disciplines. But it is also possible that teachers with different views of learning within the same discipline will use portfolios in very different ways. Comparative studies of portfolios across disciplines are needed in order to get a deeper understanding of this complex issue. This question is very topical in our country at this moment of time when portfolios are in their first phase. Our research group was asked by a highly placed administrator in the Ministry of Education to tell them in which disciplines portfolios were suited as an assessment tool and where they were not. Critics of portfolio assessment in Norway tend to think that it is a fad in the wake of the Quality Reform that has invaded disciplines where it does not fit, and that it actually represents a threat to quality assessment in all subjects. It would therefore be useful to distinguish what is general and what is discipline specific about portfolio assessment.

The wide variations in our study correspond to findings in a more fine-grained study at the University College of Oslo (Wittek & Havnes, 2005). The varied portfolio conceptions may indicate on the one hand that they are seen as a flexible assessment tool that may be customized to fit all needs. Another interpretation is that 'portfolio' is being used in Norway as a new term for basically old and traditional forms of assessment, maybe out of ignorance, maybe in order to be seen as up-to-date in the reform climate. On this background (Wittek & Havnes, 2005: 45) raises the question: Is it in the interest of stakeholders to continue using portfolio assessment about very different practices?

Stakeholders' views of portfolio variations

For *students* it may be problematic if the term 'portfolio assessment' is used about very varied practices within the same institution because then they will meet very different expectations from course to course. Familiarity with an assessment form is important for students' feeling of mastering. On the other hand, from a learning point of view, students may benefit from assessment that is customized to each course, and it may be that the main issue is the clarity and the quality of the information given to students.

Teachers who want to use portfolios for learning, prefer flexibility, and one of the attractions of portfolios for teachers is that it can be made to fit very different courses and very different styles of teaching. *Administrators and governing bodies*, however, who want to compare courses, would prefer a clear definition of what portfolio practices entail and what an assessment portfolio looks like. Standardizing portfolios seems to them the only sensible way to go. An additional force in this direction in our post Bologna period, is that *European education authorities* are very much concerned with student mobility and therefore comparability of courses within the European higher education sector. This presupposes a more common definition of what is meant by 'portfolio assessment', and preferably also standardization. This is the route taken by the European Common Framework for Language Competence, where standardized descriptions of competency levels are supposed to be accepted throughout Europe.

Quality issues regarding feedback

The importance of teacher and peer response in formative assessment

From a socio cultural perspective on learning, which emphasizes social interaction and co-construction of meaning, feedback is a crucial element of portfolio practice. According to our survey almost all informants say that students get feedback on their portfolio assignments: 90% of course leaders say feedback is given in the course, and 40% say they use peer feedback extensively. Consistent findings in international educational research have shown that feedback has positive effect on student learning (Black & William, 1998; G. Gibbs *et al.*, 2003).

The importance of the quality of feedback for *formative* assessment was also underlined in the final report of a very comprehensive study of undergraduate teaching and learning in several disciplines and across several universities in Great Britain.⁵ It shows that feedback and supervision were the two issues that concerned students most in their learning environment and where satisfaction was lowest (Hounsell & Entwistle, 2005). Three aspects of good feedback was timing, the use of examples and information about assessment criteria underlying the feedback. Our study does not, however, say anything about the quality of the comments, whether actually use them to improve their work or whether students are expected to show that they have revised. Because so much teacher and student time is invested in feedback, this is an issue that needs further investigation.

As to the quality and the use of *peer response* in connection with portfolios, our survey tells us that peer response is widely used, but that only in 29 % of portfolio assessed courses have students got any kind of instruction or training in how to go about giving feedback. An empirical study of electronic portfolio assessment in history shows that students' value both teacher and peer feedback, but use the comments sparingly in revision (Dysthe & Tolo, in press). Other studies have shown that giving feedback is a new task for students and they are uncertain about how to do it (Wittek, 2003). Empirical studies from other countries have shown that there is an improvement potential in increasing instruction and training (Sluijsmans, 2002; Zhu, 1995).

Open (public) feedback a quality issue?

An interesting finding from our study is that almost 80 % of student comments were given in fora open for others than the authors of the text. This could either be in response groups meeting face to face or by electronic commenting in group rooms in Virtual Learning Environments. The latter is more and more common, and one of the major changes in higher education pedagogies is the change from private to public feedback, i.e. teacher and/or student feedback is made available for the whole group, not just the individual student. We see this as a quality issue supported by both theory and empirical evidence. From a socio-cultural perspective on learning (Dysthe, 2003a; Säljö, 2001; Vygotsky, 1986) it can be claimed that when students get access to different ways of solving an assignment and different response voices, it may further their ability to self assess. It can be argued that the formative effect of open feedback is more general and that the connection to improvement of student papers is more indirect. Many students also want the authoritative voice of the teacher telling them what to do and what to correct (Dysthe & Breistein, 1999).

We have not seen empirical studies comparing the quality of the comments in open and individual feedback, but the teachers interviewed in our case study of portfolios in history (Dysthe & Tolo, in press) indicate that the quality improved considerably when comments were published electronically and thus accessible for all students and teachers in the course.

Criteria for what constitutes good feedback are both general (across disciplines) and discipline specific (G. Gibbs, 1999; G. Gibbs & Simpson, 2003). Teachers need to ask: What constitutes good feedback *in this discipline, in this course, at this level?* As good feedback is time consuming, it is also necessary to ask: Considering *the resources available in our particular context right now*, how can we increase the professionalism and effectivity of feedback? How do we balance collective and individual feedback? Where is a realistic level of ambition for teachers and level of expectation for students? The question of teachers' increased workload as a result of the Quality reform has been given much attention by the universities in Norway, and the needs of different stakeholders have to be balanced (Dysthe *et al.*, 2006).

⁵ Enhancing Teaching-Learning Environments in Undergraduate Courses (2001-2005)

Different stakeholders' view of feedback

While students see good feedback as one of the major quality issues of the portfolio process, and teachers agree in principle but have to consider time constraints, those who use assessment results, view the opportunities for extensive feedback as one of the factors that make portfolios most suspect as assessment instruments. It is summed up in the question: "Whose work is this anyway?" Last year it was even raised to the floor of the British Parliament in an attempt to reduce the weight given to coursework (a version of portfolio assessment used in England for many decades). Keeping the problem of outright plagiarism aside, feedback was seen to contaminate the assessment results. One way of dealing with this dilemma, is to ask students to document the process in the portfolio, including feedback and how they have used it. The problem with this is that increases the workload of assessors if they have to deal with multiple drafts, unless the reflective letter is reintroduced where students explain how they have dealt with major critical comments from teachers and peers.

The grading process and the use of explicit assessment criteria

The grading process of essay exams in Norwegian higher education has traditionally not been formalized and this has made it difficult to change to more reliable methods of scoring student papers even after the Quality Reform (Dahl, 2006). The use of explicit criteria and rubrics is even more necessary when portfolios are being assessed, and it has been defined as an important quality issue (Smith & Tillema, 2006). In our study only 56 % report using explicit, written criteria, and many of the informants did not think that these would be useful (Table not included). This corresponds to our earlier findings as well as international studies (Dysthe & Tolo, in press; Ecclestone, 2001) (add other ref). Teachers argue that tacit criteria developed through years of assessment experience and that explicit criteria may be too rigid and limiting. There is some truth in the last assertion as international literature warns against too instrumental use of criteria and documents the importance of involving students in the development and interpretation of them (Dysthe *et al.*, in press; Elwood & Klenowski, 2002; Rust, 2002; Rust *et al.*, 2003; Sadler, 1983; Woolf, 2004).

Before the Quality Reform the compulsory use of an external assessor (from another institution) for all HE exams (together with the teacher of the course or an internal assessor), was seen as a safeguard for reliable assessment. Dahl (2006) has argued that the lack of shared, explicit criteria and rubrics constituted a major flaw in our assessment system. If so, this is even more the case after the Quality Reform when the use of external assessor is voluntary and it has been abolished to a great extent. In such a context it is alarming that almost half of the teachers do not favour the use of explicit criteria.

Stakeholders' views of this issue usually coincide, but our survey has shown that not all teachers agree. Students generally want to have explicit criteria, both in order to know what is expected of them and as a help when they are asked to give feedback to peers or to self assess (Wittek & Dysthe, 2003), but mature students argue against instrumental use of criteria (Dysthe *et al.*, in press). Assessment experts and portfolio researchers have agreed that the use of explicit criteria and rubrics for assessing portfolios is one of the means of making assessment more reliable (Camp, 1996; Murphy, 2003; White *et al.*, 1996b; White & Ostheimer, 2006; Yancey & Huot, 1997). White's account of writing portfolios is relevant also to portfolio assessment across the disciplines as in our survey:

Assessment of writing by portfolios has gone through several developments since it emerged in the early 1990s. ... Portfolios appeared to resolve many of the problems that had become evident with essay testing: the validity problem of using only one or at most two impromptu writing samples, the absence of opportunities for the writer

to reflect and revise, the lack of context and audience for the writing, inappropriate or banal writing prompts, and so on (White, 2005: pp 582).

White then discusses the problems with scoring portfolios holistically. This system was developed for essay testing by the Educational Testing Service in the 1960s to solve the twin problems of unreliable (inconsistent) scoring and high cost. By ensuring that all readers were trained and used the same set of scoring guidelines (rubrics) and sample papers, “these methods produced quick and reliable essay scoring and swept the country in a remarkably short time” (p. 584). This has become the standard method of portfolio scoring in the United States even though White argues that “many aspects of holistic scoring work against the principles behind portfolio assessment” (p.583). He suggests two ways of dealing with the problems of holistic assessment that have emerged, neither of which is satisfactory: “Either we grade the entire portfolio by averaging evaluations of the different components, more or less intuitively, or we develop a scoring guide listing traits that must appear somewhere (but not everywhere) in the portfolio” (p 586). White’s proposal of a “phase 2” assessment that reflects the nature of portfolios, where the reflective letter is at the centre of portfolio scoring, is interesting, but not relevant at the early stage of portfolio development in Norway. He also recognizes that portfolios, which is a collection concept and not an assessment concept, “take many different shapes for many different purposes, and therefore will require many variations for scoring” (p1 583). We think that a reasonable place to start in our context in order to improve the quality of the scoring process, is to develop explicit criteria and scoring guides for each particular course portfolio that reflect the goals of the course.

Conclusion

In this article we have documented findings about the recent introduction of portfolios in higher education in Norway and discussed some of the quality issues regarding variations in portfolio conceptualizations, feedback and scoring practices. Since many of the institutions and disciplines are in a very early stage of developing portfolio assessment, very normative standpoints may be counterproductive and halt further experimentation. Quality issues must be taken seriously, however, and we make the following recommendations to all disciplines using portfolio assessment:

- Discuss to what extent the portfolio concept and practice is a result of disciplinary characteristics and reflect the overarching goals of the study programme. Consider how metacognitive and/or critical reflection on course contents can enhance the quality of the portfolios, and whether a reflective letter is useful or not.
- Discuss criteria for good feedback among faculty and introduce students to crucial elements of good feedback practices and design effective training for them at different levels
- Focus on the development of explicit criteria and scoring guides as a means to higher reliability and more transparency in the grading process.

References:

- Allern, M. (2005). *Individuell eller kollektiv læringsprosess? mappevurdering i praktisk-pedagogisk utdanning*. [Tromsø]: Det samfunnsvitenskapelige fakultet, Institutt for pedagogikk og lærerutdanning, Universitetet i Tromsø.
- Barrett, H., & Carney, J. (2005). Conflicting Paradigms and Competing Purposes in Electronic Portfolio Development. *Submitted to Educational Assessment, an LEA Journal, for an issue focusing on Assessing Technology Competencies, July 2005*.
- Black, P. J., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7-73.
- Callahan. (1999). All done with the best of intentions: One Kentucky high school after six years of state portfolio tests. *Assessing Writing*, 6(1), 5-40.
- Camp, R. (1996). New views of measurement and new models for writing assessment. In E. M. White, W. D. Lutz & S. Kamusikiri (Red.), *Assessment of writing. Politics, policie, practices* (pp. 135-148). New York: Modern Language Association.
- Dahl, T. (2006). When precedence sets a bad example for reform: conceptions and reliability of a questionable high stakes assessment practice in Norwegian universities. *Assessment in Education: Principles, Policy & Practice*, 13(1), 5-27.
- Davies, A., & LeMahieu, P. (2003). Assessment for learning: Reconsidering Portfolios and research evidence. In M. Segers, F. Dochy & E. Cascallar (Red.), *Optimising New Modes of Assessment: In search of qualities and standards*. Dordrecht/Boston/London: Kluwer Academic Press.
- Dysthe, O. (1996). *Ulike perspektiver på læring og læringsforskning*. Oslo: Cappelen Akademiske Forlag.
- Dysthe, O. (2003a). Teoretiske perspektiver.[Theoretical perspectives.] In O. Dysthe & K. S. Engelsen (Red.), *Mapper som pedagogisk redskap. Perspektiver og erfaringer*. (pp. 37-59). Oslo: Abstrakt forlag.
- Dysthe, O. (2003b). Writing at Norwegian universities in an international perspective: From indirect strategies of strengthening writing to the Quality Reform. In G. Rijlaarsdam, L. Björk, L. Bräuer, L. Rienecker & P. Stray Jorgensen (Red.), *Studies in writing* (Vol. 12: (Teaching Academic Writing in European Higher Education), pp. 151-165). Amsterdam: Kluwer Academic Publishers.
- Dysthe, O. (in press). How a reform affects writing in higher education. *Studies in Higher Education*.
- Dysthe, O., & Breistein, S. (1999). *Fagskriving og rettleiing ved universitetet. Holdninger blant studentar og lærarar ved Institutt for administrasjon og organisasjon* [Writing and supervision at the university. Attitudes among teachers and students at the Institute of Administration and organizational science.]. Bergen: Report no 1 PLF, Universitetet i Bergen.
- Dysthe, O., & Engelsen, K. S. (2004). Portfolios and assessment in teacher education in Norway. A theory based discussion of different models in two sites. *Assessment & Evaluation in Higher Education*, 29(2), 239 - 258.
- Dysthe, O., & Engelsen, K. S. (Red.). (2003a). *Mapper som pedagogisk redskap. Perspektiver og erfaringer* [Portfolio as Pedagogical Tool. Perspectives and Experiences.]. Oslo: Abstrakt forlag.
- Dysthe, O., Engelsen, K. S., Madsen, T., & Wittek, L. (in press). Explicit criteria in portfolio assessment from student perspectives. A theory based discussion. In A. Havnes & L. McDowell (Red.), *Assessment beyond intuition: raising standards of assessment and learning*: Routledge.
- Dysthe, O., Raaheim, A., & Lima, I. (2006). Pedagogiske endringer som følge av Kvalitetsreformen. In S. Michelsen & P. Åmodt (Red.), *Kvalitetsreformen møter virkeligheten. Delrapport 1. [The Quality Reform meets reality. Evaluation report 1.]*. Oslo: Norsk forskningsråd.
- Dysthe, O., & Tolo, A. (in press). Digital portfolios and feedback practices in a traditional university course. In Kaankanranta (Red.).
- Ecclestone, K. (2001). I know a 2:1 when I see it: understanding criteria for degree classification in franchised university programmes. *Journal of Further and Higher Education*, 301-313.
- Elwood, J., & Klenowski, V. (2002). Creating communities of shared practice: the challenges of assessment use in learning and teaching. *Assessment & Evaluation in Higher Education*, 27(3), 244-251.
- Gibbs, C. (1994). *Beyond testing. Towards a theory of educational assessment*. London: Falmer Press.
- Gibbs, G. (1999). Using assessment strategically to change the way students learn. In S. Brown & A. Glaser (Red.), *Assessment matters in higher education: Choosing and using diverse approaches* (pp. 40-53). Buckingham: SRHE & Open UP.

- Gibbs, G., & Simpson, C. (2003). *Does your assessment support your students' learning?* [Centre for Higher Education Practice, Open University.
- Gibbs, G., Simpson, C., & Macdonald, R. (2003). *Improving student learning through changing assessment - a conceptual and practical framework*. Paper presentert på EARLI., Padova, Italy Aug 2003.
- Hamp-Lyons, L., & Condon, W. (2000). *Assessing the Portfolio: Principles for Practice, Theory, Research*. Cresskill: Hampton Press.
- Hellevik, O. (1999). *Forskningsmetode i sosiologi og statsvitenskap* [Research methods in Sociology and Political Science.]. Oslo: Universitetsforl.
- Hounsell, D., & Entwistle, N. (2005). *Enhancing teaching-learning environments in undergraduate courses*. Edinburgh: University of Edinburgh.
- Klenowski, V. (2002). *Developing portfolios for learning and assessment. Processes and principles*. London/New York: Routledge/Falmer.
- Klenowski, V. (2003). *Rethinking Assessment in Higher Education*. Paper presentert på 29th. Annual IAEA Conference, 5-10 October, 2003, Manchester.
- Klenowski, V., & Askew, S. (2005). Portfolios for learning, assessment and professional development in higher education. *Assesment and Evaluation in Higher Education*, 31(3), 267-186.
- Michelsen, S., & Åmodt, P. (2006). *Kvalitetsreformen møter virkeligheten. Delrapport 1* [The Quality Reform meets reality. Evaluation report 1.]. Oslo: Norsk forskningsråd.
- Murphy, S. (2003). That was then, this is now: The impact of changing assessment policies on teachers and teaching of writing in California. *Journal of Writing Assessment*, 1(1), 23-45.
- Murphy, S., & Underwood, T. (2000). *Portfolio practices: lessons from schools, districts and states*. Norwood, Mass.: Christopher-Gordon Publishers.
- Paulson, F. L., Paulson, P. R., & Meyer, C. (1991). What makes a portfolio a portfolio? *Educational Leadership*, 48(5), 60-63.
- Rust, C. (2002). How can the research literature practically help to inform the development of departmental assessment strategies and learner-centred assessment practices? *The Institute for Learning and Teaching in Higher Education and SAGE Publications*, 3(2), 145-158.
- Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment and Evaluation in Higher Education*, 28(2).
- Sadler, D. R. (1983). Evaluation and the improvement of academic learning. *Journal of Higher Education*, 54(1), 60-79.
- Segers, M., Dochy, F., & Cascallar, E. (2003). *Optimising new modes of assessment: in search of qualities and standards*. Dordrecht/Boston/London: Kluwer Academic Press.
- Shay. (1999).
- Shepard, L. A. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Red.), *Handbook of research on teaching, 4th edition*. Washington, DC: American Educational Research Association.
- Sluijsmans, D. (2002). *Student involvement in assessment. The training of peer assessment skills*. [Alternate Student involvement in assessment. The training of peer assessment skills.] Dr. dissertation, Open University, Netherland.
- Smith, K., & Tillema, H. (2006). Portfolio appraisal: In search for criteria. *Teacher and Teacher Education (Accepted for publication)*.
- Stortingsmelding-27. (2000-2001). *Gjør din plikt - Krev din rett. Kvalitetsreform av høyere utdanning* [Parliamentary proposition 27/(2000-2001).Do your duty-demand your rights. Quality Reform of Higher Education.
- Säljö, R. (2001). *Læring i praksis. Et sosiokulturelt perspektiv*. Oslo: Cappelen Akademiske Forlag.
- Vygotsky, L. S. (1986). *Thought and language* (A. Kozulin, Trans.). Cambridge, Massachusetts: MIT Press.
- White, E. M. (2005). The scoring of writing portfolios: Phase 2. *College compositions and communication*, 56(4), 581-600.
- White, E. M., Lutz, W. D., & Kamusikiri, S. (Red.). (1996). *Assessment of writing. Politics, policie, practices*. New York: Modern Language Association.
- White, E. M., & Osteimer, M. (2006). Portfolio assessment in an American engineering colleges. *Assessing Writing*, 10, 61-75.
- Wittek, L. (2003). *Læringsarbeid og mappevurdering. En casestudie fra to profesjonsstudier ved det utdanningsvitenskaplige fakultet*. Oslo: Universitetet i Oslo. Det utdanningsvitenskaplige fakultet.
- Wittek, L., & Dysthe, O. (2003). Kvalitetskriterier og vurdering - et utviklingsområde innenfor høyere utdanning. [Quality criteria and assessment - an area for development in higher education.] In O. Dysthe & K. S. Engelsen (Red.), *Mapper som pedagogisk redskap. Perspektiver og erfaringer. [Portfolios as pedagogical tools]* (pp. 129-156). Oslo: Abstrakt forlag.

- Wittek, L., & Havnes, A. (2005). *Vurderingsformer ved Høgskolen i Oslo* [Assessment methods at Oslo University College.]. Oslo: Pedagogisk utviklingscenter, Høgskolen i Oslo.
- Woolf, L. (2004). Assessment criteria: reflection on current practice. *Assessment and Evaluation in Higher Education*, 29(4), 480-493.
- Yancey, K. B. (1998). *Reflection in the writing classroom*. Logan: Utah State UP.
- Yancey, K. B., & Huot, B. (1997). *Assessing writing across the curriculum: Diverse approaches and practices*. Greenwich, CT: Ablex.
- Yancey, K. B., & Weiser, I. (Ed.). (1997). *Situating Portfolios: Four Perspectives*. Logan, UT: Utah State UP.
- Zeichner, K., & Wray, S. (2001). The teaching portfolio in US teacher education programs: what we know and what we need to know. *Teaching and Teacher Education*, 17, 613-621.
- Zhu, W. (1995). Effects of training for peer response on students' comments and interaction. *Written communication, Sage Publications*, 12(4).