

Document downloaded from:

<http://hdl.handle.net/10251/62640>

This paper must be cited as:

Josep Domenech; Desamparados Blazquez; E. De la Poza Plaza; Ana Muñoz-Miquel (2015). Exploring the impact of cumulative testing on academic performance of undergraduate students in Spain. *Educational Assessment, Evaluation and Accountability*. 27(2):153-169. doi:10.1007/s11092-014-9208-z.



The final publication is available at

<http://dx.doi.org/10.1007/s11092-014-9208-z>

Copyright Springer Verlag (Germany)

Additional Information

The final publication is available at Springer via <http://dx.doi.org/10.1007/s11092-014-9208-z>

Exploring the impact of cumulative testing on academic performance of undergraduate students in Spain

Josep Domenech¹ · Desamparados Blazquez¹ ·
Elena de la Poza¹ · Ana Muñoz-Miquel²

Abstract Frequent testing provides opportunities for students to receive regular feedback and to increase their motivation. It also provides the instructor with valuable information on how course progresses, thus making it possible to solve the problems encountered before it is too late. Frequent tests with noncumulative contents have been widely analysed in the literature with inconclusive results. However, cumulative testing methods have hardly been reported in higher education courses.

This paper analyses the effect of applying an assessment method based on frequent and cumulative tests on student performance. Our results show that, when applied to a microeconomics course, students who were assessed by a frequent, cumulative testing approach largely outperformed those assessed with a single final exam.

1 Introduction

Motivated students have a good basis for learning. Since the assessment method used in a course can drive students' motivation, it is of paramount importance to design assessment methods that properly motivate students to learn. As expressed by Biggs and Tang (2007, 163), "what and how students learn depends to a major extent on how they think they will be assessed".

Testing, as a part of the assessment procedure, is more than simply taking a sample of what students have learned. It promotes learning, even in the case in which no explicit feedback is provided (Dempster, 1991). However, providing prompt and proper feedback increases motivation and learning to a greater extent (Race, 1995; Dochy, 2008; Kuo and Simon, 2009).

When testing is performed frequently, students can receive regular feedback, which enhances their motivation. By increasing frequency of testing, class attendance also improves since students might fear missing a test, and consequently part of the course grade (Leeming, 2002; Kling et al, 2005). Moreover, a higher attendance rate helps improve student learning (Chen and Lin, 2008). However, it

¹Universitat Politècnica de València · ²Universitat Jaume I

should be bear in mind that testing might involve unintended consequences such as increased test anxiety and stress (Dochy, 2008).

Frequent testing also generates feedback the other way round. The student-to-instructor feedback provides educators with valuable information on how the course progresses, and allows them to identify students' weaknesses and strengths, and to distribute class time accordingly.

A wide variety of research studies has analysed the effect of different test frequencies on students' achievement in higher education courses. Although some of these works found that test frequency is positively related to student academic performance (Miller, 1987; Casem, 2006), most described an ambiguous relation (Fulkerson and Martin, 1981; Deck, 1998; Haberyan, 2003; Kling et al, 2005). The common ground for these studies is that they focus on tests that cover non-cumulative contents. This way of designing assessment activities contrasts with the perspective of the psychological theory, which considers that learning improves when the same material is repeated in all the learning activities (Kuo and Simon, 2009). For this reason, an assessment method based on frequent, cumulative testing could potentially improve learning. However, this methodology applied to a higher education course has hardly been explored in the literature.

The main purpose of this article is to analyse the effect of frequent testing when implemented in a cumulative fashion, and also its influence on higher education students' performance. Cumulative means that each test includes all the materials from the beginning of the course. This method was applied to an intermediate microeconomics course and was put into practice as weekly cumulative tests. The analysis was performed by considering three different points of view: i) comparing the academic performance of frequently tested students with others who were less frequently tested; ii) collecting students' impressions about the assessment method; and iii) testing the validity of the assessment method.

2 Literature Review

2.1 Feedback and motivation

The behaviourist approach bases the learning process on the reward expected after following a behaviour (Skinner, 1974). Grades can be viewed as extrinsically motivating drivers since they are material rewards offered by lecturers to reinforce good studying behaviour (Adelman and Taylor, 1990). Thus the assessment and grading play an essential role in the learning process.

Assessment has been traditionally categorised into two types: summative and formative (Scriven, 1967). While summative assessment measures student achievements up to a certain point, the main purpose of formative assessment is to provide students with feedback (Taras, 2005; Trotter, 2006) to help them be fully aware of the standards that are expected of them, as well as their achievement level. This feedback is actually two-way (Brown et al, 1997; Boston, 2002; Yorke, 2003; Marriott and Lau, 2008) since it also provides instructors with a useful performance indicator of their teaching techniques. When it is timely, feedback allows students and lecturers to reshape their learning/teaching methods (Yorke, 2003; Nicol and Macfarlane Dick, 2006). In fact, assessment can be used as a means to consolidate and steer learning (Dochy, 2008).

Since feedback gives students the opportunity to reflect on their progression, it should be frequent and timely. That is, feedback has to be provided when it is still useful to identify and correct any deficiencies (Chickering and Gamson, 1987; Crooks, 1988; Gibbs and Simpson, 2005; Marriott, 2009). For this reason, a formative assessment scheme is a powerful practice that underpins high-quality learning and teaching; gauges student progress and provides them with suitable support by means of timely, effective feedback. Likewise, it encourages personal reflection and involvement in the students' learning process (Marriott, 2009).

Motivation should be promoted by good feedback practice (Nicol and Macfarlane Dick, 2006). As mentioned above, designing an assessment method that allows instructors to give regular feedback to students is of capital importance since feedback also plays a powerful motivational role.

However, the provided feedback should be the 'good' one. Just giving marks to students poorly contributes to the learning process and can even prove very damaging as it favours social comparisons, which may discourage students who have problems to perform or weak self-esteem (Gibbs and Simpson, 2005). What students need to know is in which questions they have badly performed and why to have the chance to improve (then they feel motivated because of the expectations of a new opportunity).

Although students may have similar levels of motivation to perform a given task, their internal motivation (i.e., self-generated motivation, interest in the subject, learning goals) may be different in nature (Marriott, 2009). In this vein, lecturers can have a (positive or negative) impact on students' external motivation (Nicol and Macfarlane Dick, 2006) as they may influence their point of view of the subject and their involvement in the assessment process (Marriott, 2009).

2.2 Frequent testing

Assessment frequency needs to be designed to benefit both students and lecturers (Marriott, 2009). Following Kuh (2003), the more the students practice and get feedback on their performance in a given subject (including writing, analysis and problem solving), the more skillful and capable they become. In addition, frequent testing and regular feedback help reduce anxiety related to one shot testing (Crooks, 1988).

The relation between frequency of exams and student achievement has been largely studied in a wide variety of frequencies. However, most of these studies obtained ambiguous results.

Deck (1998) and Kling et al (2005) examined students' performance in final exams of marketing courses. Both research works compared the effects of taking weekly tests versus monthly tests, with no overlapping test content in both frequencies; i.e. non-cumulative contents. Neither work was able to conclude that the frequent-tested group performed better than the control group. Similar results were described by Fulkerson and Martin (1981) when considering two-weekly versus monthly tests for a psychology course. They also considered tests with no overlapping test contents.

The experiment reported by Casem (2006), however, showed a positive relation between test frequency and student performance. In this work, the students in a

biology course that took nine tests per semester better performed than those taking only three. Here tests with no-overlapping content were also implemented.

De Paola and Scoppa (2011) identified two effects related to increased exam frequency: workload division effect and feedback provision effect. Their study analysed two introductory economics courses in which a setting of two non-cumulative exams per semester was compared to a final examination method. The results showed that students who were more frequently tested obtained higher grades than those examined only once. The authors found that this was due exclusively to the workload division effect, whereas the feedback provision effect was insignificant.

The ambiguity as to how test frequency and its related feedback affect students' performance may be caused by the non-cumulative nature of the tests under study. According to Dempster (1991), the use of cumulative questions in tests is essential to learn effectively. In the same vein, Kuo and Simon (2009) mentioned that, following the psychological literature, the positive effect of frequency is found when the same material is repeated in all learning activities. However, higher education experiences of frequent and cumulative tests are scarce.

In short, the literature review on feedback, motivation and testing suggests that frequent testing can increase motivation as students develop self-efficacy, become more engaged and feel less anxious before taking an exam. This assessment method may also help them acquire more profound and long-lasting knowledge. To reinforce these findings, more research in this line has to be done.

3 Assessment method

3.1 Course

The method under study was applied to an intermediate microeconomics course taught at the Universitat Politècnica de València (Spain) during the academic year 2011. This is a mandatory course taught during the second semester of the first year of the Bachelor Degree in Business Administration and Management, which has a scope of 240 credits over four academic years. This microeconomics course includes 6 credits representing 60 hours of teaching activity.

The frequent cumulative testing method described below accounted for 85% of the course grade, and the activities done in labs represented the remaining 15%. These activities consisted in four sessions where students had to develop a project related to the course contents, which had to be presented orally during the last lab session. The grading system in Spanish universities follows a numerical scale, for which a minimum of 50% of the total score is required to pass the course. For this reason, grade and score will be used interchangeably in the remainder of the paper.

3.2 Frequent Cumulative Testing (FCT)

The frequent cumulative testing (FCT) method was designed to replace the final exam. Students were assessed by hand-written in-class tests that were taken most of the weeks throughout the semester. By the end of the semester, ten tests had

been taken by students, and the average score accounted for 85% of the course grade. Additionally, two extra tests were scheduled for allowing students to make up two missed or low-scored tests. This evaluation system was compulsory for all the students, who had no option to sit a final examination to pass the course.

These tests had three specific characteristics: firstly, they lasted 30 minutes. The aim of this short duration was to promote the development of the ability to solve problems quickly, and to prevent tiredness.

Secondly, tests were prepared to cover all the content taught to date. The intention of this cumulative content nature was to help students consolidate the knowledge acquired as they had to revise learned material week after week. Roughly, half of the test dealt with recent contents (taught in the last 2 weeks), while the other half dealt with content taught earlier. As more contents were covered, fewer questions per already tested content were asked to maintain the 30-minute duration in all the exams. Students were aware of this test structure in advance.

Finally, two types of tests were prepared: multiple choice and problem or essay question tests. These test types were taken on alternative weeks. Hence both theoretical and applied learning were tested. Multiple choice tests, which included between 10 and 12 questions with four options, were automatically corrected with optical mark recognition (OMR) software. Problem or essay question tests included one to three questions.

To allow students to receive prompt feedback, suggested answers were immediately handed out in class and posted online to facilitate their revision. Marks were also made available on the same day the test was done or the next day at the latest. The fact that answers and marks were quick published provided students with timely feedback on their performance, and enabled them to identify and amend their weaknesses in good time.

Feedback was also useful in the students-to-instructor direction since it allowed us to quickly detect weaknesses. For this reason, the questions that the majority of students got wrong, as well as related topics, were tackled in class to ensure that they were well understood.

4 Empirical Analysis

As mentioned above, the main objective of this study was to evaluate the influence of applying a frequent cumulative testing (FCT) scheme on students' performance. For this purpose, an empirical analysis taking three different perspectives into account was performed.

In the first part, students' academic performance was evaluated by a statistical analysis. Data from the academic year in the FCT scheme were compared to data from the previous year in which a final exam (FE) method was followed. The analysis particularly focused on the impact that the assessment method had on exam scores and pass rates.

In the second part, student opinions on the assessment method were analysed to obtain a wider view of their perspective. To this end, students in the FCT group were asked to freely and anonymously write their opinion on the assessment method on the last day of class. Written opinions were analysed by outlining the main ideas of each student and then they were classified by a thematic approach. In this way, it is possible to draw general conclusions on each topic.

Finally, the third part analysed a questionnaire that was completed by the students assessed by the FCT scheme. It was designed following the guideline provided by Dochy (2008) to evaluate the edumetric quality aspects of educational assessment. Specifically, the questionnaire evaluated the validity of the assessment tasks, the validity of the assessment scoring and the consequential validity of the assessment method.

4.1 Comparing student academic performance

This subsection aimed to provide a quantitative assessment of the outcomes of the FCT method described above. As mentioned earlier, academic performance of the students under FCT treatment was compared to the performance of those students that took the course during the previous academic year, in which a final examination (FE) method was followed. Particularly, the analysis focused on how the assessment method affected the probability of passing the exam and the exam scores obtained by students.

Students following the final examination method were required to pass a single exam covering all the course contents to earn the credits. This exam, which took the same structure as that of the tests prepared for FCT students, accounted for 90% of the course grade. It combined essay and multiple choice questions retrieved from the same test bank. Lab sessions were organised in the same way in both years. If a student failed his/her first chance to pass the course, he/she had only one chance to retake it 3 months later. All the results presented herein refer to the exam result of their first opportunity. Accordingly, these results were compared to the first chance of passing under FCT; that is, excluding the two make-up tests.

4.1.1 Data

The data set used in these analyses included graded students of three microeconomics course classes. One of the classes (48 students) was assessed by the FCT method, while the other two (101 students) were assessed by an FE method. The three classes were taught by the same instructor to avoid biases in teaching. A few exchange students attended the course both years, but they were not included in the data set because most of the control data for them were missing.

Table 1 provides descriptive statistics for the variables included in the study. *Score* is expressed as a percentage of the highest potential score, the final exam score for students assessed by the FE method and the average of the ten in-class test scores for students assessed by the FCT treatment. To allow a fair comparison, make-up tests were not considered in both methods. Therefore, tests missed by FCT students were considered to score 0 points. In this case, the average grade for the considered cohort was about 53%. *Pass* is a dummy which takes a value of 1 if the student passed the exam. According to the Spanish higher education grading system, this occurs when the exam score is more than 50% of the highest possible score.

Some control variables were also included to capture the individual characteristics that might affect academic performance, apart from the assessment method. All the control variables were retrieved from university records and are expressed as follows: *Female* is a dummy to control for students' gender that was included

Table 1 Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Pass	149	0.611	0.489	0	1
Score	149	52.998	17.46	12	95.1
FCT	149	0.322	0.469	0	1
Average	149	66.832	6.316	52	90
Female	149	0.611	0.489	0	1
Enrolled Credits	149	60.97	10.169	30	94.5
Acquired Credits	149	41.265	11.983	6	84
Retake	149	0.027	0.162	0	1

because results on testing frequency may differ depending on gender. For instance, the literature suggests that males perform better in final exams and worse in continuous assessment methods than females (McNabb et al, 2002). This suggestion falls in line with the females' preference for continuous assessment, as found by other research work (Furnham and Chamorro-Premuzic, 2005; Cano, 2011). *Average*, which ranges from 50 to 100, represents the students' grade point average (GPA). This variable encompasses data for all the courses taken during the first academic year, except for the course which is being analysed in this paper. *Acquired credits* represents the credits for the courses passed, apart from the course under study. Both *Average* and *Acquired credits* are expected to positively relate to the course grade. In contrast, *Enrolled credits*, which account for students' workload during the year, are expected to negatively affect student performance. *Retake* is a dummy that equals 1 when students are retaking the course. As Eikner and Montondon (2001) suggested, retaking the course may have an impact on students' expectations and performance. According to their study, which consisted in finding factors of success in an intermediate accounting course, students who were retaking the course better performed in pretests than first-time students, but worse in final grades. However, other studies have found evidence that students retaking the course achieved higher course grades if compared with first-time ones (Emerson and Mencken, 2011).

Since student assignment to classes was not randomised, it is important to check that individual characteristics are similar across treatments. Table 2 compares the control variables for both assessment methods. As observed, the proportion of female students is higher in those groups assessed with the final exam method (64% vs. 54%), as was the proportion of students retaking the course (3% vs. 2%). The students in the FCT group acquired more credits (42 vs. 40.9) with a slightly higher GPA at the end of their first academic year (67.3% vs. 66.6%), whereas they were enrolled in fewer credits (58.75 vs. 62). Although none of the described differences was statistically significant at the 5% level, this should be cautiously interpreted because the sample was not random and some differences in unobservable variables might exist.

4.1.2 FCT effects on exam scores

This section aims to quantify the effect of applying a frequent cumulative testing method on exam scores. For this purpose, several linear regressions in the course

Table 2 Student characteristics across the FCT and FE groups

	Means		Differences
	FCT	FE	FCT vs. FE
Average	67.333	66.594	0.739 (1.109)
Female	0.542	0.644	-0.102 (0.086)
Enrolled Credits	58.750	62.025	-3.275 (1.768)
Acquired Credits	42.000	40.916	1.084 (2.112)
Retake	0.021	0.030	-0.009 (0.029)

Standard errors in parentheses

exam score were tested. These regressions are specified as follows:

$$Score_i = \beta_0 + \beta_1 FCT_i + \phi X_i + \varepsilon_i \quad (1)$$

where $Score_i$ is student i 's exam score, FCT_i is a dummy that equals 1 when student i followed the FCT method, and X_i is a vector of the control variables for student i . The estimation results are summarised in Table 3.

The regression shown in column 1 includes the assessment method as the only regressor on the exam score, with no other controls. According to the estimations, the application of FCT significantly affected students' scores, which increased by 12 percentage points if compared to the final exam score method. The effect of FCT was still strong and statistically significant when the control variables were included in the regression, as column 2 shows.

The marginal effects of the control variables displayed the expected behaviour. Variables related to student performance in other courses (*Average* and *Acquired credits*) positively affected student performance in the course under study. The variable relating to students' workload during the academic year (*Enrolled credits*) had, as expected, a negative impact on their exam scores. The effect of the *Female* variable was ambiguous a priori since the tests included both multiple choice and essay questions (Lumsden et al, 1987). Similarly, no hypothesis was assumed about the effect of retaking a course.

Columns 3 and 4 examine the effect of FCT on males and females separately. The coefficients on the treatment variable apparently show that the effect of FCT was stronger on females (their grades increased by 10.4 points) than on males (their grades increased by 7.6 points). This difference could be in line with females' higher preference for continuous assessment, which has been found by other research work (Furnham and Chamorro-Premuzic, 2005; Cano, 2011). However, if the interaction term between the assessment method and students' gender was included in the regression (see column 5), its effect was non-significant. That is, the effect of FCT on exam scores was similar for male and female students.

4.1.3 FCT effects on pass rates

Similarly to what has been previously shown for estimating marginal effects on course grades, this section aims to quantify the impact of a frequent cumulative

Table 3 Estimates of the effect of the assessment method on exam scores. Dependent variable: *Score*

	(1)	(2)	(3)	(4)	(5)
	All	All	Males	Females	All
FCT	12.113*** (3.700)	9.659*** (1.214)	7.601** (3.115)	10.377*** (0.490)	8.292*** (2.738)
Average		1.606*** (0.194)	1.617*** (0.359)	1.563*** (0.233)	1.591*** (0.214)
Acquired credits		0.271*** (0.102)	0.142 (0.178)	0.350*** (0.124)	0.270*** (0.102)
Female		-1.734 (1.629)			-2.516 (2.189)
FCT × Female					2.379 (2.697)
Enrolled credits		-0.327*** (0.089)	-0.404*** (0.136)	-0.325*** (0.117)	-0.329*** (0.0851)
Retake		20.408*** (5.467)	16.787** (6.388)	24.922*** (8.037)	19.940*** (6.190)
Constant	49.160*** (3.700)	-48.248*** (9.740)	-37.982 (25.269)	-50.719*** (7.822)	-46.516*** (11.265)
R-squared	0.106	0.608	0.670	0.587	0.609
Observations	149	149	58	91	149

Robust clustered standard errors in parentheses (clustered by class)

* ($p \leq 0.1$), ** ($p \leq 0.05$), *** ($p \leq 0.01$)

testing method on the probability of passing course examinations. Formally, the logit model used to estimate this probability can be expressed as:

$$\begin{aligned}
 \text{Prob}(Pass_i) &= \frac{1}{1 + e^{-Z_i}} \\
 Z_i &= \beta_0 + \beta_1 FCT_i + \phi X_i
 \end{aligned}
 \tag{2}$$

where $Pass_i$ takes the value of 1 when student i passed the exam, FCT_i equals 1 if student i followed the frequent cumulative testing method, and X_i represents the vector of the control variables. A positive value for the β_1 coefficient indicates a positive relation between the FCT method and the probability of passing the exam. Table 4 reports the estimates for the equivalent specifications to those in Table 3.

Column 1 presents the estimation results for the specification in which no control variables were included. These results show that the odds ratio of passing the exam with the FCT method was 9.30. The estimation of this effect became even better when the control variables were included in the specification (see column 2). As observed, the effects of the control variables were similar to their effects on the course score, and they also presented the expected direction.

Unlike what was found for exam scores, the marginal effect of the assessment method was apparently more marked for male students than for their female counterparts, as columns 3 and 4 show. To further investigate this difference, the interaction term between the assessment method and gender was included in the logistic regression. Column 5 presents the estimation for this equation, in which the interaction term had a non-significant coefficient. Therefore, the assessment method similarly affected both genders.

Table 4 Estimates of the effect of the assessment method on the probability of passing the exams. Dependent variable: *Pass*

	(1)	(2)	(3)	(4)	(5)
	All	All	Males	Females	All
FCT	2.230*** (0.314)	3.062*** (0.397)	3.176*** (0.989)	2.717*** (0.101)	3.260*** (0.635)
Average		0.260*** (0.043)	0.263** (0.105)	0.267*** (0.021)	0.264*** (0.264)
Acquired credits		0.028 (0.024)	0.037 (0.054)	0.032** (0.013)	0.029 (0.025)
Female		0.502 (0.583)			0.590 (0.704)
FCT × Female					-0.585 (0.750)
Enrolled credits		-0.035* (0.019)	-0.051 (0.052)	-0.035 (0.024)	-0.035* (0.020)
Retake		3.960** (1.939)	2.191 (1.368)	10.403*** (1.062)	4.405** (2.047)
Constant	-0.0785 (0.3142)	-16.792*** (2.480)	-16.367** (8.312)	-16.709*** (0.858)	-17.180*** (2.033)
Pseudo R-squared	0.222	0.573	0.603	0.568	0.574
Observations	149	149	58	91	149

Robust clustered standard errors in parentheses (clustered by class)

* ($p \leq 0.1$), ** ($p \leq 0.05$), *** ($p \leq 0.01$)

4.2 Students' comments on the assessment method

During the last class of the academic year, students were asked to complete a questionnaire about the FCT method (which is analysed below), and to optionally write their opinion on this assessment scheme anonymously. This was done to obtain a broader view of their perspective. Nine of the 48 students in the FCT group wrote some valuable comments, which were classified into three main themes to make their analysis easier.

FCT contributed to work more frequently. The initial purpose of introducing weekly tests was to prevent cramming exams near the end of the semester, which in previous years led to surface learning and poor academic performance. With FCT, students were forced to work seriously on the course from the very first day of class. Despite students being reluctant to study every week at the beginning, most of them expressed that taking a weekly exam made them study more frequently than otherwise¹.

I think that the followed assessment method is quite correct since it allows us to keep up to date with the course and permits the exam to cover all the materials presented in class (Student 4)

In general, the experience of this new assessment mode is positive. It helped me to learn and to study more frequently (Student 1)

Some of the students also reported that studying frequently made it easier to pass the course, if compared to an assessment method that centres on a final exam.

¹ It should be noted that students' comments were translated into English, and the accuracy in what students wanted to transmit may have been somewhat lost.

If you make some effort throughout the semester, it is not difficult to pass (Student 8)

I think that this assessment method is not as hard as a final exam and, at the same time, it gives us a chance to get higher marks than in a final exam (Student 5)

Weekly tests give you the opportunity to avoid the risk of sitting a final exam (Student 6)

FCT helped increase learning and academic performance. Improving students' academic performance and, therefore, their learning, was the main purpose to implement the FCT scheme. Although the effects of this assessment method on academic performance have been analysed above, improvements in learning and performance were also emerging themes in the students' comments. Despite the fact that students' perception of their increased learning and academic performance was not enough to infer that FCT actually contributes to this improvement, this perception reflected a positive attitude, which undoubtedly contributes to the learning process. Around half the students' comments were in this direction.

I think that the method employed in class led to good outcomes for most students. In the end, I think it is worth making the effort to sit weekly tests (Student 9)

I liked the assessment method. It made it easy to learn and the course seemed easier to pass (Student 7)

The assessment method was pertinent and appropriate to help retain knowledge in your mind (Student 2)

Personally, I very much liked the assessment mode of this course. Sitting tests made me to study every week. I think that this way of evaluating is advantageous for all students. I feel that I have learned lots of things, beyond test scores (Student 3)

Negative effects. Only one student described some cons of being frequently tested. This comment was related to the little grade weight given to each test. In our opinion, obtaining most of the course grade progressively allows students to work only to ensure that they have obtained the desired grade, and then they can spend time on studying other courses with final exams.

I think that sitting so many exams means that, in the end, one makes less effort in it (Student 1)

4.3 Quality aspects of FCT as an assessment method

This section analyses the replies to a questionnaire on the frequent cumulative testing method. It was designed following the guideline provided by Dochy (2008) to evaluate the edumetric quality aspects of educational assessment. It was handed out during the last class of the semester, when some students had already passed

the course, so the attendance rate was lower. This meant that completed questionnaires were collected from 34 students (population=48), which is a response rate of 71%. The questionnaire, which consisted in eight items measuring the validity of the assessment method, is presented in Appendix 1.

The first item related to the validity of the assessment tasks. A set of assessment tasks is valid when the content specifications properly match the competence measured. Questionnaire responses yielded a positive result since over 90% of students agreed/strongly agreed that FCT had completely covered the microeconomics content (see Appendix 1).

The purpose of evaluating the validity of assessment scoring was to identify possible weaknesses when interpreting scoring, for which the subsequent five items (2-6) were used. The results showed that over 80% of students agreed with four of them, which means that the assessment criteria were appropriately used and seemed fair for all the students. Only in statement 4, "FCT lets you show your learning", did the percentage of students who agreed with it lower to 59%. This lower, although favourable value, is consistent with the study of Nowell and Alston (2007) about the overconfidence that students often exhibit in economic courses as they consistently overestimate their grades.

The last two sentences in the questionnaire are related to consequential assessment validity. Student responses revealed that actual assessment consequences are in line with the expected ones. In this context, it is worth noting that, after considering the responses to statement 8, FCT provided students with relevant feedback to make progress.

Overall, the questionnaire responses showed that FCT meets the quality requirements that student-centred assessment modes should follow.

5 Conclusions

This paper has analysed the effect of implementing an assessment method based on frequent, cumulative testing in an intermediate microeconomics course. The results of this research evidence that this method should be considered by economics instructors. Students who were assessed by the frequent cumulative testing approach largely outperformed those assessed with a single final exam. FCT, as the main assessment method in the course, induced students to work on a weekly basis, rather than cramming work in for the final exam at the end of the semester. This evaluation method was also perceived very positively by students, who considered it to be valid and motivating. This fact encouraged us to use it in subsequent years.

Our study also shows the positive effects of FCT on both scores and pass rates. Frequent testing allowed students to receive prompt, timely feedback at a time when which they still had time to react. A cumulative basis allowed them to go more deeply into knowledge and find relations between the concepts explained throughout the semester. Our results support the findings of Casem (2006) and De Paola and Scoppa (2011) about frequent testing, but by considering cumulative rather than non-overlapping tests.

Cumulative testing also enabled students to measure their efforts. This effect can be considered negative since some students, once they were about to achieve

their desired grade, displayed reduced external motivation and academic performance. Although this was not common, very few students seemed to adopt this behaviour.

One of the concerns that instructors had before putting FCT into practice was loss of class time derived from conducting the tests. However, the fact that students revised the whole course content weekly allowed them to form stronger grounds, which made them understand new contents more easily and learn faster. Despite spending 30 minutes each week, the whole course content was covered by the end of the academic year.

The main drawback of FCT is probably the extra effort made by instructors to mark tests and the difficulty to apply it to larger classes, unless computer-based tests are used. This drawback was partially alleviated by the fact that 50% of tests consisted in multiple choice questions, which were marked automatically by optical mark recognition software. In any case, the tests including essay questions or problems should be carefully prepared to make them easy to mark. The prompt availability of test solutions and marks also reduced the number of situations in which students disagreed with their test marks.

This study has several limitations. Firstly, the research was not based on an experimental design. Although appropriate control variables were included for individual student characteristics, it is possible that other factors were confounded when the FCT method was adopted. This, together with the small sample size, limits the generalisation of the results. Another limitation of this work is that, as the only comparison made was between the final exam assessment method results and the FCT method results, no data about another form of frequency in taking exams is available. That is, the study cannot provide evidence that this assessment frequency is the optimum choice to obtain the best gains in student learning. However, this is the first work to explore weekly and cumulative tests as the main assessment method in a higher education course and, despite the limitations, this method yielded better student performance. Further research in this direction could confirm these findings.

References

- Adelman, H. S., & Taylor, L. (1990). Intrinsic motivation and school misbehaviour: Some intervention implications. *Journal of Learning Disabilities* 23, 541 – 550
- Biggs, J., & Tang, C. (2007). *Teaching for quality learning at university*, 3rd edn. Open University Press
- Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research & Evaluation* 8
- Brown, G. A., Bull, J., & Pendlebury, M. (1997). *Assessing Student Learning in Higher Education*, 1st edn. Routledge
- Cano, M.-D. (2011). Students' involvement in continuous assessment methodologies: A case study for a distributed information systems course. *IEEE Transactions on Education* 54, 442 – 451
- Casem, M. L. (2006). Active learning is not enough. *Journal of College Science Teaching* 35
- Chen, J., & Lin, T.-F. (2008). Class attendance and exam performance: A randomized experiment. *The Journal of Economic Education* 39, 213 – 227

- Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *American Association for Higher Education Bulletin* 39, 3 – 7
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research* 58, 438 – 481
- De Paola, M., & Scoppa, V. (2011). Frequency of examinations and student achievement in a randomized experiment. *Economics of Education Review* 30, 1416 – 1429
- Deck, W. (1998). The effects of frequency of testing on college students in a principles of marketing course. PhD thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia
- Dempster, F. N. (1991). Synthesis of research on reviews and tests. *Educational Leadership* 48, 71–76
- Dochy, F. (2008). *The Edumetric Quality of New Modes of Assessment: Some Issues and Prospects*. Assessment, Learning and Judgement in Higher Education, Springer Netherlands, Dordrecht
- Eikner, A. E., & Montondon, L. (2001). Evidence on factors associated with success in intermediate accounting I. *Accounting Educators' Journal* 13
- Emerson, T. L. N., & Mencken, K. D. (2011). Homework: To require or not? online graded homework and student achievement. *Perspectives on Economic Education Research* 7
- Fulkerson, F., & Martin, G. (1981). Effects of exam frequency on student performance, evaluations of instructor, and test anxiety. *Teaching of Psychology* 8, 90 – 93
- Furnham, A., & Chamorro-Premuzic, T. (2005). Individual differences and beliefs concerning preference for university assessment methods. *Journal of Applied Social Psychology* 35, 1968 – 1994
- Gibbs, G., & Simpson, C. (2005). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education* 1(August 5, 2011), 3 – 31
- Haberyan, K. A. (2003). Do weekly quizzes improve student performance on general biology exams? *The American Biology Teacher* 65, 110 – 114
- Kling, N., McCorkle, D., Miller, C., & Reardon, J. (2005). The impact of testing frequency on student performance in a marketing course. *Journal of Education for Business* 81, 67 – 72
- Kuh, G. D. (2003). What we're learning about student engagement from NSSE. *Change* 35
- Kuo, T., & Simon, A. (2009). How many tests do we really need? *College Teaching* 57, 156 – 160
- Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology* 29, 210 – 212
- Lumsden, K. G., Scott, A., & Becker, W. E. (1987). The economics student reexamined: Male-female differences in comprehension. *Journal of Economic Education* 18, 365–375
- Marriott, P. (2009). Students' evaluation of the use of online summative assessment on an undergraduate financial accounting module. *British Journal of Educational Technology* 40, 237 – 254
- Marriott, P., & Lau, A. (2008). The use of on-line summative assessment in an undergraduate financial accounting course. *Journal of Accounting Education* 26,

73 – 90

- McNabb, R., Pal, S., & Sloane, P. (2002). Gender differences in educational attainment: The case of university students in England and Wales. *Economica* 69, 481 – 503
- Miller, F. (1987). Test frequency, student performance and teacher evaluation in the basic marketing class. *Journal of Marketing Education* 9, 14–19
- Nicol, D. J., & Macfarlane Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education* 31, 199 – 218
- Nowell, C., & Alston, R. M. (2007). I thought I got an A! Overconfidence across the economics curriculum. *The Journal of Economic Education* 38, 131 – 142
- Race, P. (1995). The art of assessing 1. *New Academic* 4
- Scriven, M. (1967). *The Methodology of Evaluation*, vol 1, Rand McNally, Chicago, p 39–83
- Skinner, B. F. (1974). *About behaviorism*. Alfred A. Knopf, New York
- Taras, M. (2005). Assessment - summative and formative - some theoretical reflections. *British Journal of Educational Studies* 53, 466 – 478
- Trotter, E. (2006). Student perceptions of continuous summative assessment. *Assessment & Evaluation in Higher Education* 31, 505 – 521
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education* 45, 477 – 501

Appendix 1. Questionnaire to students

Results of questionnaire on frequent cumulative testing

	Strongly disagree		Disagree		Neutral		Agree		Strongly agree	
	No	%	No	%	No	%	No	%	No	%
1. FCT has completely covered the expected microeconomics content					3	9	14	42	16	49
2. FCT is congruent with the received instruction			1	3	5	15	14	41	14	41
3. All students have equal opportunities to demonstrate their capabilities			1	3	4	12	13	38	16	47
4. FCT lets you show your learning			2	6	12	35	13	38	7	21
5. Scoring criteria were clearly defined					2	6	10	30	22	65
6. I understand the criteria used to evaluate my performance	1	3			3	9	10	30	19	58
7. Breath and depth in learning is actively rewarded, instead of by chance			5	15	8	24	15	44	6	18
8. Provided feedback was relevant for progressing					5	15	18	53	11	32

%=% of respondents