VOLUME 10 ISSUE 2 2011

International Review of
Economics Education

The Higher Education Academy Economics Network

International Review of Economics Education

An international journal that promotes research into effective learning and teaching practice in economics higher education

The International Review of Economics Education is dedicated to enhancing learning and teaching in the higher education economics community. It provides a forum for high quality research in the areas of curriculum design, pedagogy, assessment, innovation and evaluation. The journal seeks to promote critical dialogue on educational theory and practice in economics and to demonstrate the relevance of research to good professional practice.

Submissions
Submissions should be sent in electronic form and should be uploaded through our Web interface. Further details of the submission procedure are available at:
www.economicsnetwork.ac.uk/iree/submit.htm

Ashley Lait, Managing and Production Editor
IREE, Economics Network
University of Bristol
8-10 Berkeley Square
Bristol, BS8 1HH, UK

Telephone: +44 (0) 117 331 4333
Email: ashley.lait@bristol.ac.uk

Contents

Editorial Issue 10.2

Ross Guest

In our dual roles as economists and educators, we are interested in the way incentives drive our allocation of time to teaching and other academic activities. The study by Harter, Becker and Watts investigates this question for U.S. academic economists from 1995 to 2005. Their first key result is rather nice in a way: it suggests that U.S. academic economists continue to irrationally spend too much time on teaching given the clear premium in reward for research outcomes. This echoes their own prior studies and those in other countries (Guest and Duhs, 2002, for example); all of which may be welcome confirmation that academic economists are not flint-hard *homo economicus* when it comes to caring about student learning. On the other hand it may simply reflect pressures of large classes and teaching administration that cannot be avoided. The authors' second key finding is perhaps more interesting: that female economists allocate proportionally more time to teaching than do male economists, especially at research universities. Some possible explanations are offered and further work is recommended. It does seem to us an important question, not least for its implications for female salary and career progression.

Hickson and Reid compare the assessment information provided by multiple choice (MC) and constructed response (CR) questions. They define a 'constructed question' as one which requires a student to offer their own short response to a question in contrast to a multiple choice question where students choose between alternative answers offered by the examiner. They find that their CR questions appear to be assessing something which is distinct in comparison to their MC questions and that their CR questions are better predictors of final GPA from other courses/modules. They argue that differences between their results and previous studies are due to differences in method rather than peculiar qualities of the particular questions they used.

The article by Mearman *et al*. returns to the theme of our 2009 Special Issue on pluralism in economics education. The central theme of that Issue was that the actual plurality of the economics discipline, evidenced by its many recognised branches, is in fact not reflected in the mainstream undergraduate curriculum. The new evidence provided by Mearman *et al*. is based on focus group interviews of students. While acknowledging potential bias, the authors found that students liked the plural approach to learning economics - in particular, they appreciated taking a wider perspective in drawing on other disciplines, and are comfortable with ambiguity and the partial state of knowledge. Appealing to what students like and want may be a good strategy for achieving curriculum change, given the increasing importance attached to student evaluations of teaching and student experience surveys.

We continue to publish articles that provide new insights, perspectives and strategies for dealing with particular topics in the curriculum. Three papers in this issue serve this purpose: by Dalziel, Marsden and Sibly, and Kapinos.

Dalziel reminds us that the way students construct knowledge depends on their prior learning, their prior conceptions of the discipline and their approach to learning. He cites Schumpeter's (1949) notion that students have a prior "vision" of the discipline they are learning. Dalziel's concern is with students studying introductory economics as a compulsory unit in degree programs such as environmental

management, landscape architecture or resource studies. He advocates a very applied problem-based approach for these students and presents evidence to show that this leads to improved academic performance. This study adds to the growing body of evidence that applied problem-based learning in economics can be effective for all students, including economics majors (in IREE, for example, see Rigall-I-Torrent, 2011, and a number of papers on classroom games and experiments in several IREE issues, notably the Special Issue, 9.2).

Teachers of intermediate-advanced microeconomics will appreciate the way Marsden and Sibly simplify and integrate the three degrees of price discrimination. The authors argue that the traditional textbook treatment of second and third price discrimination is rather disjointed. They explain a more coherent treatment and illustrate it with a number of real world examples.

An appreciation of the causes and consequences of the ongoing global economic crisis calls for critical reflection of the mainstream macroeconomics curriculum. Have we underplayed the role of debt, financial asset prices, balance sheets, monetary policy targets, liquidity traps? The list goes on. In this context the paper by Kapinos is timely. He presents a new analysis of the liquidity trap in an inflation-targeting framework.

References

Guest, R. and Duhs, A. (2002) Economics Teaching in Australian Universities: Rewards and Outcomes, *The Economic Record* 78, 241, 147-160

Rigall-I-Torrent, R. (2011) Using Problem-based Learning for Introducing Producer Theory and Market Structure in Intermediate Microeconomics *International Review of Economics Education*, 10. 1. 14–28

Schumpeter, J. A. (1949) Science and ideology, *American Economic Review*, 39, 345–59

Time Allocations and Reward Structures for US Academic Economists from 1995–2005: Evidence from Three National Surveys

Cynthia L. Harter, William E. Becker and Michael Watts

Abstract

Using survey data collected in 1995, 2000 and 2005 from US academic economists, in which respondents were asked to indicate what percentage of their work time they allocate to research, teaching and service activities, and also how their departments and schools weight research, teaching and service in determining annual raises and making promotion and tenure decisions, we find these economists were allocating more time to teaching even though perceived departmental and school incentives provided a clear premium for research. The overall samples did not show major changes in their allocation of time from 1995–2005, but there were different responses at different types of schools, with increased time spent on research by faculty at doctoral schools while at masters' and baccalaureate schools more time was devoted to teaching. We use regression analysis to investigate factors that affect how different faculty members allocate their time between teaching and research. In addition to Carnegie school classifications and related school characteristics, faculty members' gender and rank were significant predictors of how economists allocate their time. Male economists, particularly among assistant professors at research universities, spent less time on teaching and more time on research than female economists.

JEL classification: A20, A22

1. Introduction

In this paper we report survey data on time allocation decisions made by US academic economists, and investigate factors that explain those allocations. We combine data from national surveys conducted in 2005 (Watts and Becker, 2008) and in 1995 and 2000 (Becker and Watts, 1996, 2001), which they used to identify teaching methods in undergraduate economics courses. Our focus here is on questions from the background sections of these surveys, in which respondents were asked to indicate the percentage of time they allocated to teaching, research, and service, and the weightings they felt their departments assigned to each of these activities for decisions on annual raises and, separately, for promotion and tenure decisions.

The limited research on factors affecting how faculty members allocate time to different activities has been noted before (Gautier and Wauthy, 2007; Toutkoushian, 1999). The earlier research almost always considers faculty members across fields and departments, focusing on rank/tenure or other variables

that presumably affect all faculty members in much the same way, or on general problems and issues such as the difficulty of accurately measuring output in research, service and especially teaching. A large share of this work has been done by economists, who not surprisingly focus on the effects of incentives facing faculty. But despite that, the previous studies rarely deal with faculty members from economics or any other particular discipline, as we do here – although a few make some comparisons across faculty from different disciplines.

By drawing on the three national surveys of US economists, we are first able to see how well these responses match with the responses of faculty members from the cross-discipline surveys that are featured in most of the earlier research. But our focus on economists and the factors that are important in their time allocation decisions also leads us to consider issues that might have different effects from what is or might be seen in most other fields. The most notable such factor is gender, which has been shown to play some (relatively modest) role across faculty from all disciplines, but may have more impact in economics given the historical under-representation of females in economics – at least in the United States – that has been documented both in terms of the shares of undergraduate and graduate degrees awarded, and the share of faculty members at different ranks and types of schools.[1] After briefly reviewing earlier studies on these topics, we use the three national surveys of US economists to develop new findings on these topics and issues.

2. Literature review

Becker (1979) provided an early theoretical model of the expected effects of raising weights assigned to research or teaching, given differences in the ability to quantify and agree upon measures of faculty performance. Flemming (1991) raised additional questions about measures of research output, and how those measures can be subject to different incentive issues and to the mix of pure vs. applied research.

A 1994 report from the U.S. National Center for Educational Statistics (NCES) found full-time faculty reported working more than 50 hours a week, on average, with more time devoted to research at research-intensive institutions. A 1997 NCES study using data from the 1993 National Survey of Postsecondary Faculty (NSOPF) found that full-time faculty spent, on average, 54% of their time teaching and 16% on research, with men spending a higher percentage of time (18% vs. 12%) on research, and a lower percentage (55% vs. 62%) on teaching. Using the same NSOFP data, Bellas and Toutkoushian (1999) found that after controlling for race, experience, marital status, number of children, age, highest degree, rank, field and Carnegie classifications of institutions, women spent only 3% more time teaching than men, and 2% less time on research. They also found that men reported working about two hours a week more on the job than women, but only about one hour a week after controlling for the same factors listed above.

Also using the NSOPF data, Walstad and Allgood (2005, p. 182) concluded that many economics professors at research universities had 'a low regard for teaching and a high regard for research' – and did not find that to be true for professors in other social sciences, the biological or physical sciences, mathematics and statistics, engineering, or business. They found physical and biological scientists were closer to economists' views than respondents from other academic disciplines, 'but not nearly as extreme in the views of the teaching and research tradeoffs as… economics professors' (pp. 182–3). In a later working paper using the NSOPF data, Allgood and Walstad (2006) found a bi-directional but asymmetric substitution effect in faculty allocations of time to research or teaching, with a 10%

[1] See annual reports of the American Economic Association's Committee on the Status of Women in the Economics Profession, published in the annual May *Proceedings* volumes of the *American Economic Review* (or in recent years in the supplementary online *Proceedings*); annual reports on bachelor's degrees awarded in economics drawn from the AEA Universal Academic Questionnaire, published by John Siegfried in the *Journal of Economic Education*; and Ginther and Kahn (2004).

increase in time spent on research leading to a 1.5% decrease in time spent on teaching, and a 10% increase in time spent on teaching reducing time spent on research by 8.5%.

For a sample of US faculty members from the arts and sciences, Singell, Lillydahl and Singell (1996) found that most differences in faculty time allocations were accounted for by structural differences between universities with different research orientations. They also found that faculty characteristics reinforced the different institutional missions, and concluded that self-selection will condition university policies intended to change faculty behaviour – for example in trying to direct more faculty time to teaching at research universities.

Milem, Berger and Dey (2000) found that faculty could devote more time to both teaching and research, and that substitutions between teaching and research might be unidirectional, so that spending less time on teaching could lead to spending more time on research, but devoting less time to research would not lead to spending more time on teaching.

Becker, Lindsay and Grizzle (2003) argued that many universities require faculty to do both research and teaching because stronger students choose to attend universities where faculty are doing more research. They report a strong negative relationship between faculty time devoted to teaching and time doing research.

Laband and Tollison (2003) found a substantial increase in the emphasis on research at US and other universities from 1974–96, tied to stronger incentives for faculty to increase research output (including higher salaries, reduced teaching loads and increased support for travel to conferences). Measured by the share of uncited papers, however, which remained constant at 26%, they found no improvement in the quality of research published.

Gautier and Wauthy (2007, p. 274) argued that faculty time allocation across teaching and research is 'largely a matter of taste and incentives' and support university-wide 'yardsticks' to allocate research funds as a way to improve both teaching and research. Einarson and Clarkberg (2004) found that the time faculty report spending with students outside the classroom is not affected by reported faculty work time constraints, but is affected by faculty having children and by differences in beliefs about the educational role of faculty members.

Link, Swann and Bozeman (2008) reported that time allocations by science and engineering faculty at top US universities are affected by tenure, promotion, and other career path issues, with full professors spending more time on service and less time on teaching and research. Long-term associate professors spend more time teaching and less time on research. Female faculty members appear to allocate more time to service and less time to research.

3. Data

The 1995, 2000 and 2005 mail surveys from which our data are drawn were all five pages long, with few changes in items across the different years. In 1995, 2947 economists were selected either as academic members of the American Economic Association (AEA) or as college/university teachers of economics listed by College Marketing Guide (CMG), a private company that offered mailing lists of US college and university instructors. Unfortunately, for the 2000 and 2005 surveys mailing lists of AEA members by employee type were no longer available. Instead, in 2000 a sample of 3103 economists was drawn entirely from CMG lists. In 2005 the CMG lists were no longer available, so lists of economics instructors were purchased from Market Data Retrieval (MDR), a private company that offered mailing lists of various groups, including college teachers in different disciplines. The 2005 survey was mailed to 3711 academic economists.

In all three surveys fixed-interval sampling was used to identify the questionnaire recipients from the respective source lists. In 1995 the response rate was 21%, in 2000 it was 19%, and in 2005 it was 13%. Results from all three surveys are based on opportunistic samples and self-reported data. There is no way of knowing whether respondents are representative of all US teachers of undergraduate economics courses, but our intuition is that those with greater interest in teaching were more likely to complete and return surveys. Even if this is true, it is not clear that would lead to a predictable bias in responses to the questions of interest in this paper. For example, instructors more interested in teaching might be likely to report higher percentages of time spent on teaching and higher weights on teaching for annual raise and promotion/tenure decisions because they have found positions that reward teaching, either at departments and schools that emphasise good teaching or working as teaching specialists in more research-oriented departments. On the other hand, they might report higher percentages and weights on research if they feel the research demands they face at their schools are excessive. Consequently, we have not attempted to adjust for any possible bias resulting from sample selection issues.

Definitions for variables on faculty members' personal time allocations and departmental incentives – indicating the percentage of time instructors report spending in teaching, research and service, and the weightings these same respondents feel their institutions assign to these activities in awarding annual raises or promotion and tenure – are reported in Table 1. Mean responses and standard deviations are provided for 1995, 2000 and 2005.[2] Percentages do not sum to 100 because some respondents could view their time allocation as including activities other than teaching, research and service, including administration.[3]

As reported in Harter, Becker and Watts (2004), from 1995 to 2000 economists were allocating more time to teaching even though their perceptions of departmental and school incentives (for promotion and tenure decisions as well as annual raises) provided a clear premium for research. The disparity in time allocation and reward structures continues in the 2005 data. Specifically, for the overall sample we see almost no change in faculty time allocations from 2000 to 2005, with US economists spending a little over half of their time on teaching, a little over 20% on research, and about 9% on service activities. This is very much in line with findings from the NSOPF data for faculty from all departments, reported above.[4]

Other than a slight decrease in the perceived weightings assigned to research, the relative weightings on teaching, research and service for promotion and tenure decisions changed very little from 2000 to 2005 in the overall sample. For annual raises the importance of both teaching and research decreased slightly from 2000 to 2005, perhaps reflecting a general funding environment over that period in which most US departments and schools were, in practice if not in word, giving across-the-board raises more often than differentiating on merit. But in general there was very little change in the structure of incentives from 2000 to 2005.

[2] The mean values are slightly lower and the numbers of observations slightly higher here than the basic results reported in Watts and Becker (2008) because we are focusing on a sub-section of the survey and made some minor adjustments. Specifically, if a respondent reported percentages for the weights on promotion and tenure decisions that totalled 100 but left some entries for those variables blank, it seemed clear the blanks represented a zero so we replaced the blanks with zeros. We did the same for variables on department weights for annual raise decisions and faculty time allocation percentages. For some variables this increased the number of observations and lowered mean values.

[3] Tables of *z*-statistics for differences in the mean values reported in Tables 1 and 2, across the three survey periods, are available on request. We make no attempt to draw statistical inferences because of the nature of the survey data.

[4] Guest and Duhs (2002) provide survey evidence for economists in Australia, and conclude that rewards for teaching are too low in Australian schools to promote better student ratings of teaching. They also note the limited use of teaching methods that engage students in learning.

Table 1 Variable definitions and mean values
(standard deviations in parentheses)

	1995		2000		2005	
Variable definition	**n**	**Mean**	**n**	**Mean**	**n**	**Mean**
Proteach – weight in percentage that teaching has in school decisions about promotion and tenure	556	44.13 (25.95)	534	49.00 (24.31)	401	49.85 (22.59)
Proresearch – weight in percentage that research has in school decisions about promotion and tenure	557	43.68 (27.90)	531	37.52 (25.74)	400	35.42 (24.06)
Proservice – weight in percentage that service has in school decisions about promotion and tenure	555	11.94 (9.77)	529	13.15 (10.34)	401	13.89 (9.89)
Annteach – weight in percentage that teaching has in school decisions about annual raises	478	37.53 (27.32)	420	41.02 (26.71)	297	38.09 (26.18)
Annresearch – weight in percentage that research has in school decisions about annual raises	480	40.49 (28.57)	415	36.86 (26.48)	295	35.84 (27.47)
Annservice – weight in percentage that service has in school decisions about annual raises	478	12.19 (12.75)	414	12.36 (11.28)	295	13.16 (12.10)
Teach – percentage of work time devoted to teaching	588	51.96 (22.96)	567	55.85 (21.92)	455	56.45 (23.60)
Research – percentage of work time devoted to research	587	29.54 (22.06)	564	22.82 (18.88)	450	22.52 (20.77)
Service – percentage of work time devoted to service activities	na	na	562	9.09 (8.57)	451	9.32 (8.95)

There are several possible explanations for the disproportionate amount of time spent teaching, compared to the reward structures for teaching and research. Teaching loads and large class sizes in economics courses (both in absolute terms, and compared to class sizes in other disciplines) may require more time than the mix reflected in departmental or school incentives; or additional time spent on research may not reliably lead to more publications, and so have a lower expected return than additional time spent on teaching; or as a matter of tastes economics faculty at most schools may prefer to spend more time teaching than doing research.

The 1995 and 2000 data showed interesting differences in time allocation and incentive structures for baccalaureate and doctoral institutions, however, so to determine whether these differences persisted

in 2005, in Table 2 we break down the time allocation (part A) and incentive results (part B) across different types of institutions using three Carnegie classifications – bachelor's, master's, and doctoral institutions. There were insufficient responses from associate-degree-granting institutions in the 2005 survey to include that as a fourth group.

Table 2 Means of percentages of faculty time allocations and departmental incentives by Carnegie Classification of Institution (standard deviations in parentheses)

Table 2A Faculty time allocations

Faculty time variables (percentages)	1995 values	2000 values	2005 values
Baccalaureate Institutions			
Teach	47.08 (19.19) n=98	61.13 (17.79) n=113	64.29 (19.82) n=101
Research	32.65 (19.89) n=98	17.22 (14.50) n=112	14.60 (12.43) n=100
Service	na	10.00 (9.26) n=112	9.76 (7.86) n=100
Masters Institutions			
Teach	56.03 (20.16) n=134	57.60 (19.83) n=193	59.97 (21.32) n=157
Research	24.63 (16.93) n=134	19.80 (14.14) n=190	19.87 (17.45) n=157
Service	na	9.60 (8.34) n=191	9.80 (8.01) n=157
Doctoral Institutions			
Teach	64.70 (20.95) n=99	48.51 (19.02) n=72	45.69 (23.57) n=162
Research	16.48 (15.40) n=98	29.01 (18.85) n=72	33.38 (24.06) n=159
Service	na	9.98 (8.11) n=72	8.71 (10.36) n=160

Table 2B Departmental incentives

Departmental incentives variables	1995 values	2000 values	2005 values
	Baccalaureate Institutions		
Proteach	34.02 (16.48) n=96	59.81 (17.69) n=110	59.43 (16.31) n=100
Proresearch	54.91 (19.25) n=96	25.75 (16.30) n=109	24.87 (14.74) n=99
Proservice	10.92 (6.90) n=96	15.93 (11.43) n=108	15.64 (8.74) n=99
Annteach	31.92 (15.92) n=83	49.58 (26.91) n=77	42.06 (28.21) n=71
Annresearch	52.60 (20.87) n=83	24.59 (19.25) n=76	21.72 (20.57) n=70
Annservice	11.72 (8.94) n=83	14.49 (10.81) n=76	16.06 (13.18) n=70
	Masters Institutions		
Proteach	50.78 (17.23) n=131	50.23 (17.48) n=179	54.84 (16.06) n=137
Proresearch	33.32 (16.97) n=131	32.81 (17.84) n=179	29.93 (16.11) n=136
Proservice	15.69 (9.16) n=131	16.02 (8.89) n=178	15.34 (7.91) n=137
Annteach	40.09 (26.26) n=110	41.40 (25.16) n=142	44.49 (24.43) n=90
Annresearch	28.77 (22.32) n=110	30.57 (21.82) n=142	30.04 (21.68) n=89
Annservice	13.94 (13.55) n=110	13.34 (11.95) n=141	14.03 (10.11) n=89

Table 2B Departmental incentives (cont.)

Departmental incentives variables	1995 values	2000 values	2005 values
	Doctoral Institutions		
Proteach	61.08 (18.93) n=96	39.71 (15.04) n=68	30.56 (16.23) n=130
Proresearch	22.52 (15.90) n=96	49.29 (18.41) n=68	57.57 (21.07) n=131
Proservice	16.35 (11.57) n=96	11.28 (7.32) n=68	10.52 (8.46) n=131
Annteach	53.44 (27.62) n=77	40.98 (16.97) n=58	28.56 (17.04) n=113
Annresearch	21.91 (18.30) n=77	45.31 (19.86) n=58	56.09 (23.24) n=113
Annservice	16.53 (14.79) n=77	12.31 (7.15) n=58	11.06 (9.04) n=113

We find only small changes in time allocations in 2005 for any of the different types of schools, but the direction of changes are different at different types of schools. At bachelor's and master's institutions there is a small increase in time devoted to teaching, and at bachelor's institutions there is a small decrease in time devoted to research. Conversely, at doctoral institutions we see a small decrease in time spent on teaching and a small increase in time spent on research. This probably signals that faculty at the different types of schools are being held to different kinds of performance standards.

That is supported by comparing changes in incentives structures from 2000 to 2005, when there is a decrease in the weightings for both teaching and research in determining annual raises at bachelor's institutions and an increase in the importance of teaching for both promotion/tenure decisions and annual raises at master's universities. Conversely, at doctoral institutions the relative weighting for teaching declined while the importance of research increased. Although faculty at the doctoral schools still continue to report spending nearly half of their time on teaching – and considerably more time than the perceived weights for teaching in departmental and school incentive structures – over time the faculty at these schools do seem to be responding to a rising premium on research. Those internal incentives from departments are no doubt reinforced by a growing difference in compensation levels for economists at doctoral/research schools, compared to other schools with more of a teaching mission, as reported annually in the May *American Economic Review Papers and Proceedings* volume.

4. Determinants of time allocations

To investigate the determinants of time allocations reported by academic economists, we report results from ordinary least squares regressions in which the dependent variables are the percentages of time allocated to either teaching or research. Independent variables include both individual respondent characteristics – gender, rank, whether or not the respondent speaks English as a first language, and membership in the American Economic Association (AEA) – and institutional characteristics such as the Carnegie classification for respondents' schools, the average size of principles classes and average teaching loads for faculty in the respondents' departments, and weightings assigned to teaching and

Table 3: Additional variable definitions and mean values for combined (1995, 2000 and 2005) responses (standard deviations in parentheses)

Variable name	N	Mean
Assoc – dummy variable = 1 for schools with Carnegie classification of Associate	1696	0.10 (0.30)
Bacc – dummy variable = 1 for schools with Carnegie classification of Baccalaureate	1696	0.19 (0.40)
Masters – dummy variable = 1 for schools with Carnegie classification of Masters	1696	0.29 (0.46)
Male – dummy variable = 1 for males	1663	0.80 (0.40)
Instructor – dummy variable = 1 for faculty with rank of Instructor or Lecturer	1636	0.09 (0.29)
Asst – dummy variable = 1 for faculty with rank of Assistant Professor	1636	0.20 (0.40)
Assoc – dummy variable = 1 for faculty with rank of Associate Professor	1636	0.27 (0.45)
Other Rank – dummy variable = 1 for teaching assistants, adjunct professors, emeritus professors, or other miscellaneous ranks	1636	0.03 (0.16)
English1 – dummy variable = 1 for faculty who speak English as their first language	1663	0.89 (0.31)
Class Size – average size of principles classes in the respondent's department	1382	66.34 (90.10)
SemLoad – the average semester teaching load for tenure and tenured-track faculty in the respondent's department	1444	3.10 (1.05)
1995 Dummy – dummy variable = 1 for responses from the 1995 survey	1696	0.37 (0.48)
2000 Dummy – dummy variable = 1 for responses from the 2000 survey	1696	0.35 (0.48)
AEA Member – dummy variable = 1 for respondents who report that they are AEA members	1635	0.64 (0.48)

research activities for promotion and tenure decisions. We also included dummy variables to indicate the year for each survey response – 1995, 2000 or 2005 (the omitted category).[5]

In Table 3 above we report descriptive statistics for the additional variables included in the OLS regressions, based on data from all three surveys. Most respondents are male and speak English as their first language. The largest number of respondents – but not a majority – taught at doctoral institutions.

As expected, the variables for Carnegie classification were highly correlated with other variables measuring institutional characteristics, such as average class size in principles courses and weightings for teaching and research activities in the respondents' promotion and tenure or annual raise decisions.[6] Therefore, to test whether the Carnegie classifications had the expected effects on time allocations for teaching and research, we first used OLS to regress only the Carnegie classification variables on our dependent variables. We report these results combining all three survey datasets in Tables 4A (using percentage of time spent on teaching as the dependent variable) and 4B (using percentage of time spent on research as the dependent variable). The omitted comparison groups are the Carnegie Doctoral and Research classifications – the separate Carnegie classification for Research schools was dropped before the 2005 survey was conducted, so for 1995 and 2000 we combined the Doctoral and Research classifications. It is clear that respondents from the Research and Doctoral institutions spend more time on research and less time on teaching than respondents at schools with the other classifications.

Table 4A: OLS regression for Carnegie Classifications:
Dependent variable = percentage of time spent on teaching

Carnegie Classification	Coefficient	*p* value
ASSOCIATE	31.718	0.000
BACCALAUREATE	12.588	0.000
MASTERS	12.782	0.000
Constant	38.035	0.000

n = 1609
adjusted R-squared = 0.17

[5] The reports by Becker and Watts cited above, reporting responses on teaching and assessment methods from these three surveys, were remarkably consistent over the three survey periods; but of course that does not mean that how faculty members allocated their time across teaching and research, or their perceptions of how their schools and departments were rewarding teaching and research over this 15-year period, would also be so uniform. To test that we estimated the equations reported below in Tables 5–7 (except that initially we did not include the AEA membership variable) for each of the survey years. We then compared the estimated equations using F tests. The null hypothesis of no difference between the pairs of estimates for different years was rejected for all equations with time spent on teaching as the dependent variable. It was nearly rejected for the research equations for 1995 and 2000 (p = 0.07) and 1995 and 2005 (p= 0.09), but not rejected for 2000 and 2005. Because an AEA mailing list was used to draw most of the 1995 sample but not the 2000 and 2005 samples, as noted above, we added the AEA membership variable and re-estimated the annual equations and F tests, feeling that membership in AEA was likely to signal more interest in research activities. Although that lowered some of the F values the same results persisted for the teaching equations, and the null hypothesis was still rejected in the research equations for 1995 and 2005. Therefore, in the final estimates reported here, we add binary variables to indicate the year in which the survey results were provided and include the AEA membership variable.

[6] Many of the simple correlation coefficients with the school classifications were higher than 0.5 (for absolute values), including teaching loads and weightings for teaching and research in promotion and annual raise decisions.

Table 4B: OLS regression for Carnegie Classifications:
Dependent variable = percentage of time spent on research

Carnegie Classification	Coefficient	*p* value
Associate	–29.824	0.000
Baccalaureate	–13.832	0.000
Masters	–13.917	0.000
Constant	35.086	0.000

n = 1599
adjusted R-squared = 0.20

As a robustness test, and because of the change in the Carnegie classifications before the 2005 survey, we also ran regressions for each of the three survey data sets individually, and with the combined data sets for 1995 and 2000. Those results are not reported here but are available on request. The results were quite stable, with signs for all variables unchanged in all of the regressions.

In Tables 5A and 5B we replace the variables for Carnegie classifications with variables for instructors' personal characteristics and the institutional variables such as class size, semester teaching load, and rewards for teaching and research that were correlated with the Carnegie classifications.[7] The first column of results includes observations from faculty at all schools, regardless of the Carnegie classification of the school. The next four columns show results for faculty at each of the four Carnegie Classifications (again combining the Doctoral and Research classifications, which had been merged by 2005).

For the set of academic rank variables, our omitted category was full professors and endowed chairs. We include an 'Other Rank' variable here to identify teaching assistants, adjunct professors and a few others who do not fit traditional rank categories for regular faculty. Not surprisingly, except for the 'Other Rank' group and at the Associate Schools where teaching loads for all faculty members are very high, we find that most respondents spend more time on teaching than the full professors and endowed chairholders, who normally have more experience in teaching their courses. But those differences are only significant for the instructor/lecturer group (and even there not in the Associate schools), and for the Other Rank group at the Doctoral and Research Schools. Instructors and associate professors spend less time on research than full professors, but assistant professors reported spending more time on research (except at the Associate schools). That coefficient is significant for the overall sample and at Doctoral/Research schools – almost certainly a reflection of promotion and tenure incentives.

[7] We lose 300–400 observations by including the class size and semester load variables in the regressions. In an attempt to avoid losing these observations we tried using indicator variables when mean values for these variables were inserted for the missing values. This worked reasonably well for the Semester Load variable but not for the Class Size variable, which exhibited considerably more variation. Using the indicator variables for just Semester Load saved only 127 observations and did not affect signs for other variables, so we dropped all use of the indicator variables.

Table 5A OLS regression for individual and school variables:
Dependent variable = percentage of time spent on teaching

Variable	All Carnegie Classifications – coefficient (*p* value)	ASSOC Carnegie Classification – coefficient (*p* value)	BACC Carnegie Classification – coefficient (*p* value)	MASTER Carnegie Classification – coefficient (*p* value)	DR/RES Carnegie Classification – coefficient (*p* value)
MALE	–2.341 (0.098)	6.530 (0.130)	–3.207 (0.209)	–2.933 (0.241)	–4.356 (0.095)
INSTRUCTOR/ LECTURER	11.279 (0.000)	–0.584 (0.892)	23.451 (0.001)	13.238 (0.016)	19.021 (0.000)
ASSISTANT PROFESSOR	1.865 (0.215)	–0.569 (0.907)	5.181 (0.081)	2.722 (0.272)	–1.703 (0.540)
ASSOCIATE PROFESSOR	1.982 (0.136)	5.578 (0.241)	–0.561 (0.821)	2.303 (0.299)	2.140 (0.388)
OTHER RANK	–5.494 (0.162)	–5.781 (0.398)	–3.113 (0.657)	25.044 (0.173)	–20.097 (0.006)
ENGLISH FIRST LANGUAGE	1.694 (0.363)	12.728 (0.059)	–0.422 (0.917)	1.391 (0.611)	0.596 (0.874)
WEIGHTING FOR TEACHING IN P&T DECISION	0.292 (0.000)	0.258 (0.001)	0.257 (0.000)	0.370 (0.000)	0.130 (0.025)
PRINCIPLES CLASS SIZE	–0.014 (0.031)	–0.083 (0.002)	–0.015 (0.670)	–0.095 (0.021)	–0.001 (0.946)
SEMESTER LOAD	4.732 (0.000)	4.954 (0.172)	3.694 (0.001)	3.991 (0.003)	8.496 (0.000)
1995 DUMMY	2.220 (0.119)	4.236 (0.349)	–3.342 (0.344)	1.941 (0.435)	9.175 (0.001)
2000 DUMMY	1.724 (0.201)	5.500 (0.185)	–3.603 (0.134)	3.589 (0.111)	5.996 (0.028)
AEA MEMBER	–3.666 (0.003)	–4.853 (0.196)	–2.627 (0.273)	–3.739 (0.057)	–4.720 (0.064)
CONSTANT	27.074 (0.000)	15.240 (0.450)	40.471 (0.000)	27.381 (0.000)	20.761 (0.002)
n	1104	115	253	387	349
ADJUSTED R-SQUARED	0.31	0.22	0.25	0.18	0.27

Table 5B OLS regression for individual and school variables:
Dependent variable = percentage of time spent on research

Variable	All Carnegie Classifications – coefficient (*p* value)	ASSOC Carnegie Classification – coefficient (*p* value)	BACC Carnegie Classification – coefficient (*p* value)	MASTER Carnegie Classification – coefficient (*p* value)	DR/RES Carnegie Classification – coefficient (*p* value)
MALE	2.160 (0.056)	–0.533 (0.806)	1.725 (0.399)	3.443 (0.053)	2.300 (0.343)
INSTRUCTOR/ LECTURER	–5.859 (0.002)	0.915 (0.671)	–12.801 (0.022)	–10.706 (0.009)	–12.423 (0.002)
ASSISTANT PROFESSOR	4.567 (0.000)	–1.254 (0.611)	1.853 (0.431)	2.551 (0.149)	10.231 (0.000)
ASSOCIATE PROFESSOR	–1.703 (0.108)	–4.577 (0.057)	–1.090 (0.580)	–0.010 (0.995)	–2.902 (0.207)
OTHER RANK	–2.774 (0.373)	–2.864 (0.404)	–4.252 (0.442)	–5.701 (0.662)	–1.141 (0.866)
ENGLISH FIRST LANGUAGE	–7.990 (0.000)	–3.157 (0.342)	–6.676 (0.040)	–9.212 (0.000)	–6.230 (0.074)
WEIGHTING FOR TEACHING IN P&T DECISION	0.265 (0.000)	0.469 (0.000)	0.251 (0.000)	0.351 (0.000)	0.161 (0.001)
PRINCIPLES CLASS SIZE	0.011 (0.043)	–0.004 (0.776)	0.025 (0.357)	0.040 (0.182)	0.007 (0.388)
SEMESTER LOAD	–4.664 (0.000)	1.238 (0.522)	–4.487 (0.000)	–4.783 (0.000)	–9.259 (0.000)
1995 DUMMY	2.071 (0.067)	2.418 (0.281)	4.980 (0.088)	–0.003 (0.999)	0.349 (0.886)
2000 DUMMY	–0.419 (0.697)	0.797 (0.700)	2.547 (0.182)	–3.194 (0.047)	–1.339 (0.597)
AEA MEMBER	4.477 (0.000)	5.951 (0.001)	2.934 (0.124)	4.514 (0.001)	5.349 (0.023)
Constant	29.864 (0.000)	–0.339 (0.975)	25.562 (0.000)	28.881 (0.000)	45.940 (0.000)
n	1097	113	251	384	349
Adjusted R-Squared	0.42	0.24	0.38	0.35	0.34

Faculty members who spoke English as their first language generally spent significantly less time on research than non-native English speakers. This may be because the native-English speakers were more likely to be at schools that put more emphasis on teaching, while the non-native English speakers were more likely to be at schools that placed a higher value on research. Using a chi-squared test, we found a statistically significant difference in the distribution of native English speakers versus non-native English speakers across the different Carnegie classifications of schools. Specifically, there were fewer non-native speakers at Associate and Baccalaureate schools.

The signs on the variables for departmental weightings of teaching and research activities in promotion and tenure (P&T) decisions – as perceived by respondents – have the expected signs and the coefficients are significant. The same is true for variables on class size in principles courses and faculty semester teaching loads: at schools where principles classes are smaller and faculty members teach more courses, respondents report spending more time teaching and less time doing research – although these results are not always significant for the ranges of class sizes and teaching loads reported within a particular Carnegie group of schools.

The dummy variables indicating which annual survey a respondent completed are only significant in the teaching equation for the Doctoral/Research schools, and in the research equation for the overall sample only for the 1995 survey only at the 0.10 level. As noted above, the 1995 sample was the only group drawn largely from an AEA mailing list, and the AEA membership variable does indicate that economists who are AEA members report spending significantly more time on research and less time on teaching.

Noting that there are both gender and rank effects on how academic economists allocate their time, we used interaction terms to investigate whether women and men at different ranks allocate their time differently. First we looked at the effects of gender and gender*rank interaction terms, while dropping the individual rank variables, to explore the effect gender may have on time allocations if women have different career paths and timelines across academic ranks, compared to males. Then we dropped the individual gender variable and used the different rank variables with the set of rank*gender interaction terms, to investigate the effect of rank if the distribution of male and female faculty members across ranks is different. In both sets of estimations using interactive terms we drop the Other Rank respondents who were included in Table 5, to focus on respondents in regular faculty positions (i.e. ranks of Instructor/Lecturer, Assistant Professor, Associate Professor or Full Professor/Endowed Chairs).

As reported in Table 6A, investigating whether gender makes a difference in how time is allocated to teaching and research by faculty members with different ranks, we find that although males spend less time teaching there are statistically significantly positive interaction effects for males who are instructors. For the results from all types of schools, using an F-test to test the null hypothesis that the coefficients on MALE, MALE*INSTRUCTOR, MALE*ASSISTANTPROF, and MALE*ASSOCPROF are all zeroes, we find an F value of 6.72 (with four degrees of freedom and 1069 observations) with a *p* value of 0.000. Therefore, we reject the null hypothesis and conclude that there is a difference between males and females. In the estimations for different types of Carnegie schools, we reject the null hypothesis at Baccalaureate and Doctoral/Research schools (with *p* values of 0.00) but not for Associate and Masters' schools (with p values of 0.73 and 0.13, respectively).

Table 6A OLS regression with interaction terms investigating gender and career paths: Dependent variable = percentage of time spent on teaching

Variable	All Carnegie Classifications – coefficient (*p* value)	ASSOC Carnegie Classification – coefficient (*p* value)	BACC Carnegie Classification – coefficient (*p* value)	MASTER Carnegie Classification – coefficient (*p* value)	DR/RES Carnegie Classification – coefficient (*p* value)
MALE	–4.214 (0.007)	6.703 (0.206)	–3.585 (0.205)	–4.811 (0.074)	–6.208 (0.030)
MALE*INSTRUCTOR/ LECTURER	12.423 (0.000)	–1.616 (0.736)	27.128 (0.000)	12.387 (0.040)	19.982 (0.000)
MALE*ASSISTANT PROFESSOR	0.442 (0.261)	–2.296 (0.670)	4.886 (0.142)	0.588 (0.833)	–3.046 (0.333)
MALE*ASSOCIATE PROFESSOR	1.548 (0.278)	0.677 (0.897)	–1.268 (0.635)	2.682 (0.263)	2.234 (0.400)
ENGLISH FIRST LANGUAGE	1.775 (0.335)	12.396 (0.066)	–0.885 (0.820)	1.575 (0.566)	0.985 (0.792)
WEIGHTING for TEACHING in P&T DECISION	0.289 (0.000)	0.237 (0.004)	0.249 (0.000)	0.367 (0.000)	0.159 (0.007)
PRINCIPLES CLASS SIZE	–0.015 (0.026)	–0.087 (0.001)	0.007 (0.840)	–0.100 (0.015)	–0.001 (0.888)
SEMESTER LOAD	4.765 (0.000)	5.171 (0.158)	3.923 (0.000)	4.059 (0.003)	8.406 (0.000)
1995 DUMMY	2.431 (0.088)	6.069 (0.200)	–3.240 (0.342)	1.667 (0.505)	8.195 (0.002)
2000 DUMMY	1.700 (0.206)	6.042 (0.151)	–2.769 (0.239)	3.158 (0.163)	5.314 (0.051)
AEA MEMBER	–4.039 (0.001)	–5.834 (0.146)	–2.878 (0.215)	–3.636 (0.065)	–6.127 (0.016)
Constant	29.481 (0.000)	17.050 (0.407)	40.465 (0.000)	29.767 (0.000)	23.345 (0.000)
n	1081	106	247	386	342
Adjusted R-squared	0.31	0.23	0.27	0.18	0.27

We also investigated the gender and rank interactions using the percentage of time spent on Research as the dependent variable. As reported in Table 6B, we found there are statistically significant negative interaction effects for males who are instructors and positive interaction effects for males who are assistant professors. Once again using an F-test to test the null hypothesis that the coefficients on MALE, MALE*INSTRUCTOR, MALE*ASSISTANTPROF, and MALE*ASSOCPROF are all zeroes, for the results from all types of schools we find an F value of 10.65 (with four degrees of freedom and 1062

observations) with a *p* value of 0.000 and again reject the null hypothesis of no difference between males and females. In the estimations for different types of Carnegie schools, we reject the null hypothesis at Master's and Doctoral/Research schools (with *p* values of 0.01 and 0.00 respectively), but not for Associate and Baccalaureate schools (with *p* values of 0.42 and 0.49, respectively).

Table 6B OLS regression with interaction terms investigating gender and career paths: Dependent variable = percentage of time spent on research

Variable	All Carnegie Classifications – coefficient (*p* value)	ASSOC Carnegie Classification – coefficient (*p* value)	BACC Carnegie Classification – coefficient (*p* value)	MASTER Carnegie Classification – coefficient (*p* value)	DR/RES Carnegie Classification – coefficient (*p* value)
MALE	0.855 (0.490)	0.574 (0.833)	0.515 (0.821)	2.537 (0.184)	–0.673 (0.795)
MALE*INSTRUCTOR/ LECTURER	–7.222 (0.001)	–0.409 (0.866)	–9.389 (0.116)	–11.125 (0.013)	–16.213 (0.000)
MALE*ASSISTANT PROFESSOR	5.527 (0.000)	–1.803 (0.503)	1.894 (0.478)	3.596 (0.071)	12.358 (0.000)
MALE*ASSOCIATE PROFESSOR	–1.796 (0.114)	–4.855 (0.068)	–0.105 (0.961)	–0.108 (0.949)	–3.764 (0.119)
ENGLISH FIRST LANGUAGE	–8.201 (0.000)	–3.181 (0.347)	–7.058 (0.026)	–9.406 (0.000)	–6.491 (0.056)
WEIGHTING for RESEARCH in P&T DECISION	0.264 (0.000)	0.477 (0.000)	0.243 (0.000)	0.354 (0.000)	0.155 (0.001)
PRINCIPLES CLASS SIZE	0.012 (0.030)	–0.005 (0.717)	0.008 (0.771)	0.040 (0.177)	0.008 (0.281)
SEMESTER LOAD	–4.704 (0.000)	1.346 (0.499)	–4.620 (0.000)	–4.735 (0.000)	–9.583 (0.000)
1995 DUMMY	2.042 (0.071)	2.142 (0.368)	4.855 (0.091)	0.449 (0.801)	0.498 (0.833)
2000 DUMMY	–0.175 (0.870)	0.616 (0.772)	2.716 (0.150)	–2.802 (0.082)	–0.768 (0.755)
AEA MEMBER	4.545 (0.000)	5.053 (0.011)	3.327 (0.075)	4.504 (0.001)	5.630 (0.014)
Constant	31.164 (0.000)	–0.986 (0.929)	27.532 (0.000)	29.220 (0.000)	49.494 (0.000)
n	1074	104	245	383	342
Adjusted R-squared	0.42	0.22	0.37	0.36	0.36

Finally, as reported in Tables 7A and 7B, we investigate the possible effect of rank that might be related to differences in the distribution of men and women across different ranks. We found that all of the ranks listed in the table spend more time teaching than the omitted category of full professors and endowed chairs, except for Instructors and Assistant Professors at Associate Schools, where teaching loads for all faculty members are very high. In the results for all schools we find statistically significant

negative interaction effects for male assistant professors. Using an F-test to test the null hypothesis that therefore reject the null hypothesis to conclude that rank and gender interact. In the estimations for different types of Carnegie schools we reject the null hypothesis for Baccalaureate and Doctoral/Research schools (with *p* values of 0.00), but not for Associate and Master's schools (with *p* values of 0.43 and 0.07, respectively).

Table 7A OLS regression with interaction terms investigating gender distribution across rank: Dependent variable = percentage of time spent on teaching

Variable	All Carnegie Classifications – coefficient (*p* value)	ASSOC Carnegie Classification – coefficient (*p* value)	BACC Carnegie Classification – coefficient (*p* value)	MASTER Carnegie Classification – coefficient (*p* value)	DR/RES Carnegie Classification – coefficient (*p* value)
INSTRUCTOR/ LECTURER	7.872 (0.086)	–6.870 (0.360)	6.421 (0.694)	18.180 (0.164)	18.608 (0.012)
ASSISTANT PROFESSOR	6.435 (0.005)	–5.279 (0.598)	8.137 (0.063)	9.273 (0.014)	4.660 (0.241)
ASSOCIATE PROFESSOR	4.461 (0.060)	19.819 (0.052)	1.972 (0.608)	1.719 (0.685)	5.548 (0.204)
MALE*INSTRUCTOR/ LECTURER	4.603 (0.367)	7.509 (0.336)	20.624 (0.245)	–5.834 (0.682)	1.390 (0.870)
MALE*ASSISTANT PROFESSOR	–5.948 (0.019)	5.843 (0.576)	–3.374 (0.489)	–8.637 (0.039)	–8.049 (0.071)
MALE*ASSOCIATE PROFESSOR	–2.854 (0.247)	–16.976 (0.113)	–3.255 (0.415)	0.995 (0.821)	–3.642 (0.420)
ENGLISH FIRST LANGUAGE	1.808 (0.327)	12.007 (0.074)	–0.539 (0.891)	1.648 (0.547)	0.836 (0.823)
WEIGHTING for TEACHING in P&T DECISION	0.291 (0.000)	0.280 (0.001)	0.256 (0.000)	0.370 (0.000)	0.143 (0.017)
CLASS SIZE	–0.015 (0.028)	–0.084 (0.001)	–0.010 (0.764)	–0.095 (0.022)	–0.002 (0.808)
SEMESTER LOAD	4.722 (0.000)	5.130 (0.156)	3.747 (0.000)	4.074 (0.003)	8.577 (0.000)
1995 DUMMY	2.188 (0.124)	5.672 (0.227)	–3.610 (0.291)	1.390 (0.579)	8.629 (0.001)
2000 DUMMY	1.620 (0.228)	6.447 (0.123)	–3.151 (0.179)	3.173 (0.162)	5.639 (0.039)
AEA MEMBER	–4.013 (0.001)	–4.280 (0.267)	–2.896 (0.218)	–3.758 (0.055)	–5.232 (0.041)
Constant	25.301 (0.000)	17.757 (0.375)	36.918 (0.000)	24.533 (0.000)	16.808 (0.008)
n	1081	106	247	386	342
Adjusted R-squared	0.32	0.25	0.27	0.18	0.27

Table 7B OLS regression with interaction terms investigating gender distribution across rank: Dependent variable = percentage of time spent on research

Variable	All Carnegie Classifications – coefficient (*p* value)	ASSOC Carnegie Classification – coefficient (*p* value)	BACC Carnegie Classification – coefficient (*p* value)	MASTER Carnegie Classification – coefficient (*p* value)	DR/RES Carnegie Classification – coefficient (*p* value)
INSTRUCTOR/ LECTURER	–2.434 (0.503)	3.963 (0.299)	–28.634 (0.028)	–10.741 (0.249)	–5.225 (0.431)
ASSISTANT PROFESSOR	0.994 (0.585)	–1.438 (0.817)	1.207 (0.729)	–2.497 (0.353)	4.644 (0.199)
ASSOCIATE PROFESSOR	–2.460 (0.192)	–4.828 (0.347)	–2.967 (0.338)	–1.932 (0.523)	–0.958 (0.809)
MALE*INSTRUCTOR/ LECTURER	–4.749 (0.242)	–3.977 (0.306)	18.826 (0.182)	–0.188 (0.985)	–11.050 (0.148)
MALE*ASSISTANT PROFESSOR	4.612 (0.022)	0.054 (0.993)	0.584 (0.881)	6.258 (0.036)	7.892 (0.052)
MALE*ASSOCIATE PROFESSOR	0.742 (0.704)	0.325 (0.952)	2.397 (0.455)	2.008 (0.522)	–2.668 (0.517)
ENGLISH FIRST LANGUAGE	–8.064 (0.000)	–3.341 (0.326)	–6.466 (0.041)	–9.386 (0.000)	–6.210 (0.068)
WEIGHTING for RESEARCH in P&T DECISION	0.263 (0.000)	0.474 (0.000)	0.261 (0.000)	0.355 (0.000)	0.151 (0.002)
CLASS SIZE	0.012 (0.030)	–0.005 (0.726)	0.006 (0.816)	0.037 (0.216)	0.008 (0.257)
SEMESTER LOAD	–4.725 (0.000)	1.035 (0.603)	–4.653 (0.000)	–4.767 (0.000)	–9.627 (0.000)
1995 DUMMY	2.038 (0.072)	2.240 (0.348)	5.173 (0.069)	0.443 (0.804)	0.159 (0.947)
2000 DUMMY	–0.221 (0.836)	0.653 (0.760)	2.707 (0.147)	–2.917 (0.073)	–1.008 (0.684)
AEA MEMBER	4.440 (0.000)	5.232 (0.007)	2.561 (0.172)	4.489 (0.001)	5.162 (0.026)
Constant	32.003 (0.000)	0.790 (0.942)	27.859 (0.000)	31.828 (0.000)	49.324 (0.000)
n	1074	104	245	383	342
Adjusted R-squared	0.42	0.22	0.38	0.35	0.36

Looking at the same question but using the percentage of time spent on Research as the dependent variable, with results reported in Table 7B, we found significantly positive interaction effects for male assistant professors. In the results for all schools, using an F-test to test the null hypothesis that the coefficients on the rank variables and all of the gender and rank interaction terms are zeroes, we find an F value of 7.46 (for six degrees of freedom and 1060 observations), with a *p* value of 0.000, and

reject the null hypothesis that rank and gender do not interact. In the estimations for different types of Carnegie schools we reject the null hypothesis for Master's and Doctoral/Research schools (with *p* values of 0.03, and 0.00, respectively), but not for Associate and Baccalaureate schools (with *p* values of 0.40 and 0.16, respectively).

Taken together, these results suggest that male and female faculty members behave differently, particularly at the rank of assistant professor, with males reporting that they spend more time doing research than females. The typical ages for assistant professors are the same ages at which it is most common for families to have young children, so female faculty members are perhaps spending more time in child bearing, child care, and other household production activities. They might also be working at different kinds of schools, either due to their own preferences and self-selection of jobs or because of differences in the job offers they receive. We compared the gender distribution across Carnegie classifications using a Chi-squared test and fail to reject the null hypothesis whether the Research and Doctoral classifications are combined or not. Specifically, combining the classifications we find χ^2_3 = 2.44 with a *p* value of 0.487, and not combining the classifications we find χ^2_4 = 7.47 with a *p* value of 0.113.

5. Conclusions, policy implications, and some speculations

Although we noted some periods in which departmental incentives for teaching versus research exhibited modest changes, in general across all types of schools the incentive structures and faculty behaviours are more notable for stability than for change. However, there are important and persistent differences in incentives and behaviours across different types of schools – measured here using Carnegie classifications – and for faculty members with different personal characteristics, including gender and academic rank.

Salary differentials for US economists at research versus teaching institutions have increased sharply over the past decade, as reflected in annual reports that appear in the *American Economic Review: Papers and Proceedings*. Adding that observation to our findings from these survey data raises and highlights an interesting and provocative question: are we witnessing an increasingly sharp demarcation between 'two nations' of economics departments, based on which departments have faculty who regularly publish in established economics journals and which do not? Economists at different kinds of schools appear to perceive these kinds of differences in the incentive structures they face, and allocate their time in different ways in response to those incentives. But at the same time, at all kinds of schools academic economists report devoting a disproportionate amount of time to teaching, compared to their perceptions of how teaching and research are rewarded at their schools.

The finding that rank affects how economists at different points in their academic careers allocate time to teaching and research activities is not surprising or necessarily troubling. But considered together with the apparent differences in career patterns or choices for male and female economists, which affect the distribution of time to teaching and research activities, this may represent an important issue that deserves more investigation and discussion. The gender finding is in some ways the most interesting and challenging issue to face and interpret. But the truth is earlier research in economic education has not been able to provide conclusive answers on why females are less likely to major in economics than males at the undergraduate level (at least in the United States), or to go on to graduate school and faculty positions at all kinds of schools. Some of the possible answers that have been suggested for these outcomes, such as gender role-model effects from female faculty to female students, have been heavily discounted or at least not well or consistently supported.

Similarly, at a very general level it could be suggested that positions in research/doctoral schools require greater time commitments to succeed and remain current with research skills and knowledge, and entail more risk in receiving tenure and salary increases. That could lead many female economists

who want to have children and be very active in child care to seek positions in teaching schools; or female economists might find teaching relatively more rewarding than males, on average. But it is also possible that predominantly male faculty in large, research departments engage in (statistical) discrimination against females in hiring or promotion decisions. Our findings cannot confirm or refute any of these possible explanations.

Moreover, considering the gender issue across academic fields and at other age levels further complicates things. For example, there are technical fields (notably science and medicine) that women today pursue far more than they did in the past, and at much higher levels than we see in economics. And the attitudes and experiences female students have at the precollege level are almost certainly important in shaping their choice of majors in general, and their early decision not to major in economics in particular. At the precollege level, gender differences are already widely observed in economics assessments, and perceptions (including peer perceptions) of economics as a difficult and somewhat technical subject probably discourage some women from taking courses or majoring in economics. But there must be other forces at play, too, because again we do see many women are majoring in other technical and difficult subjects, perhaps because they find the kind of work done in those areas more rewarding or interesting, or because those fields are perceived as being more welcoming and open to female students and young professionals. Those kinds of questions will probably have to be addressed in studies of precollege students, rather than studies of current faculty members. Our findings suggest that it is important for that kind of work to be done, because there are differences in how male and female faculty members in economics departments choose to do their jobs, and those differences may well be larger and more difficult to change than in other fields.

Note: We thank Ross Guest and two anonymous referees for helpful comments on an earlier version of this paper.

References

Allgood, S. and W. Walstad (2006). 'Faculty time allocations to teaching, research, and other work', Working Paper.

Becker, W. E. (1979). 'Professional behavior given a stochastic reward structure', *American Economic Review,* vol. 69 (April), pp. 1010–17.

Becker, W. E. and M. Watts (1996). 'Chalk and talk: A national survey on teaching undergraduate economics', *American Economic Review: Papers and Proceedings,* vol. 86 (May), pp. 448–53.

Becker, W. E. and M. Watts (2001). 'Teaching economics at the start of the 21st century: Still chalk and talk', *American Economic Review: Papers and Proceeding,* vol. 91 (May), pp. 446–52.

Becker, E., C. M. Lindsay and G. Grizzle (2003). 'The Derived Demand for Faculty Research', *Managerial and Decision Economics,* vol. 24 (December), pp. 549–67.

Bellas, M. L. and R. Toutkoushian (1999). 'Faculty time allocations and research productivity: Gender, race and family effects', *The Review of Higher Education,* vol. 22(4), pp. 367–90.

Einarson, M. K and M. E. Clarkberg (2004). 'Understanding faculty out-of-class interaction with undergraduate students at a research university', Cornell Higher Education Research Institute Working Paper.

Flemming, J. (1991). 'The use of assessments of British university teaching, and especially research, for the allocation of resources: A personal view', *European Economic Review,* vol. 35 (April), pp. 612–18.

Gautier, A. and X. Wauthy (2007). 'Teaching versus research: A multi-tasking approach to multi-department universities', *European Economic Review,* vol. 51 (February), pp. 273–95.

Ginther, D.K. and S. Kahn (2004). 'Women in economics: Moving up or falling off the academic career ladder?', *Journal of Economic Perspectives,* vol. 18(3), pp. 193–214.

Guest, R. E. and A. Duhs (2002). 'Economics teaching in Australian universities: Rewards and outcomes', *Economic Record,* vol. 78, pp. 147–60.

Harter, C, W. Becker and M. Watts (2004). 'Changing incentives and time allocations for academic economists: Results from 1995 and 2000 national surveys', *Journal of Economic Education,* vol. 35 (Winter), pp. 89–97.

Laband, D. N. and R. D. Tollison (2003). 'Dry holes in economic research', *Kyklos,* vol. 56(2), pp. 161–73.

Link, A. N., C. A. Swann and B. Bozeman (2008). 'A time allocation study of university faculty', *Economics of Education Review,* vol. 27 (August), pp. 363–74.

Milem, J. F., J. B. Berger and E. L. Dey (2000) 'Faculty time allocation: A study of change over twenty years', *The Journal of Higher Education,* vol. 71(4), pp. 454–75.

National Center for Education Statistics (1994). *Faculty and instructional staff: Who are they and what do the do?* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement (NCES 94-346).

National Center for Education Statistics (1997). *Instructional faculty and staff in higher education institutions: Fall 1987 and Fall 1992*, Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement (NCES 97-470).

National Center for Education Statistics (Various years). *National survey of postsecondary faculty*, Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, http://nces.ed.gov/survey/nsopf.

Singell Jr., L. D., J. H. Lillydahl and L. D. Singell Sr. (1996). 'Will Changing Times Change the Allocation of Faculty Time?', *Journal of Human Resources,* vol. 31 (Spring), pp. 429–49.

Toutkoushian, R. K. (1999). 'The status of academic women in the 1990s: No longer outsiders, but not yet equals', *Quarterly Review of Economics and Finance,* vol. 39 (Special Issue), pp. 679–78.

Walstad, W. B. and S. Allgood (2005). 'Views of teaching and research in economics and other disciplines', *American Economic Review: Papers and Proceedings,* vol. 95 (May), pp. 177–83.

Watts, M. and W. Becker (2008). 'A little more than chalk and talk: Results from a third national survey of teaching methods in undergraduate economics courses', *Journal of Economic Education,* vol. 39 (Summer), pp. 273–86.

Author Biography

Cynthia Harter is the Director of the Center for Economic Education at Eastern Kentucky University. In addition to providing teacher workshops and designing and teaching courses for teachers, Professor Harter has obtained a number of grants to study the effectiveness of economic education in central and eastern Kentucky.

Professor Becker has served as a consultant and visiting scholar at universities, institutes and agencies around the world. He was editor of the Journal of Economic Education from 1989 through 2009, when he also served on the editorial board of the Economics of Education Review. He was 2005-06 president of the Midwest Economic Association and for many years served on the American Economic Association Committee on Economic Education. His full resume is available at http://mypage.iu.edu/~beckerw/.

Michael Watts is a Professor and Director of the Center for Economic Education at Purdue University. He recently completed two terms as Chair of the American Economic Association's Committee on Economic Education.

Contact details

Cynthia L. Harter
Eastern Kentucky University, Department of Economics
106 Beckham Hall, 521 Lancaster Ave.
Richmond, KY 40475
Email: Cynthia.Harter@eku.edu
Tel: 859-622-1390

William E. Becker
Professor Emeritus of Economics, Indiana University
Adjunct Professor of Commerce, University of South Australia,
Research Fellow, Institute for the Study of Labor (IZA)
Fellow, Center for Economic Studies and Institute for Economic Research (CESifo)
325 Blue Ridge Drive, Bloomington IN 47408
Email: beckerw@indiana.edu
Tel: 812-336-9119

Michael Watts
Purdue University, Department of Economics
425 W. State Street
West Lafayette, IN 47907-2056
Email: mwatts@purdue.edu
Tel: 765-494-8543

More Evidence on the Use of Constructed-Response Questions in Principles of Economics Classes

Stephen Hickson and Bob Reed

Abstract

This study provides evidence that constructed response (CR) questions contribute information about student knowledge and understanding that is not contained in multiple choice questions (MC). We use an extensive data set of individual assessment results from Introductory Macro- and Microeconomics classes at a large, public university. We find that (i) CR scores contain information not contained in MC questions, (ii) this information is correlated with a measure of student knowledge and understanding of course material, and (iii) CR questions are better able to 'explain' academic achievement in other courses than additional MC questions. There is some evidence to suggest that this greater explanatory power has to do with the ability of CR questions to measure higher-level learning as characterised by Bloom's taxonomy (Bloom, 1956). Both (i) the generalisability of our results to other principles of economics classes, and (ii) the practical significance (in terms of students' grades) of our findings, remain to be determined.

JEL classification: A22

> 'In sum, the evidence presented offers little support for the stereotype of multiple-choice and free-response formats as measuring substantially different constructs.' Bennett, Rock and Wang (1991)
>
> 'Whatever is being measured by the constructed-response section is measured better by the multiple-choice section…We have never found any test that is composed of an objectively and subjectively scored section for which this is not true.' Wainer and Thissen (1993)
>
> 'The findings from this analysis of AP exams in micro and macro principles of economics are consistent with previous studies that found no differences, or only slight differences, in what the two types of tests and questions [multiple-choice and essay] measure.' Walstad and Becker (1994)

1. Introduction

University principles of economics courses often have enrolments of several hundred students or more. Instructors of these courses face a potential tradeoff when designing tests. On the one hand, constructed-response (CR) questions are thought to assess important learning outcomes that are not well-addressed by multiple-choice (MC) questions.[1] On the other hand, CR questions are much more costly to grade. In addition, the marking of CR questions is less reliable due to the subjective nature of

[1] Multiple choice (MC) questions present students a set of answers and ask them to select the correct one(s). Constructed response (CR) questions require students to provide their own answers. These can range from fill-in-the-blank questions; to definitional or short-answer questions; to questions requiring mathematical solutions; to long essay questions.

the questions. The tradeoff is a very real one to university instructors facing declining budget environments where marking assistance on CR questions may be reduced or eliminated.

Ideally, one would weigh the respective benefits and costs of CR and MC questions to decide the optimal mix of each to employ. However, this is a difficult task, especially given the subjective nature of 'benefits'.[2] Perhaps because of this, much attention has focused on the question, 'Do CR and MC questions measure the same thing?' If this question could be answered affirmatively, it would mean there was no 'tradeoff,' and one could eliminate CR questions. In fact, a number of influential studies claim to demonstrate this result. The implications of this have been well-understood:

> The educational measurement literature suggests that multiple-choice questions measure essentially the same thing as do constructed-response questions. Given the higher reliability and lower cost of a multiple-choice test, a good case can be made for omitting constructed-response questions from a test containing both multiple-choice and constructed-response questions because they contribute little or no new information about student achievement. (Kennedy and Walstad, 1997, p. 359).

Previous research has taken different approaches to this question. Bennett, Rock and Wang (1991) and Thissen, Wainer and Wang (1994) employ factor analysis. Walstad and Becker (1994) regress Advanced Placement (AP) composite scores on MC scores. Kennedy and Walstad (1997) simulate grade distributions using different test formats. Becker and Johnston (1999) utilise two-stage least squares regression. Each of these has its own notion of what it means to 'measure the same thing', and none attempts to reconcile their approach to those of the others. Further, most of this research has focused on AP exams. These results may not be valid for principles of economics classes taught at universities.

Our study takes yet another approach to the MC-CR debate. We use extensive data from principles classes in macroeconomics and microeconomics from a large public university where assessments consist of both MC and CR questions. Our empirical methodology is targeted to an instructor trying to decide whether to use a composite MC-CR assessment, versus an assessment composed of all MC questions. It consists of three steps.

First, we investigate the degree to which CR scores are 'predictable' from MC scores. If a student's performance on the CR component of a test can be perfectly, or near-perfectly, predicted by their performance on the MC component, we could easily conclude that the two components 'measure the same thing'. If that were the case, there would be no reason to use the more costly CR questions, and our hypothetical instructor would be better off using an all-MC assessment. Our empirical analysis does not support this view. We find that the regression of CR scores on MC scores leaves a substantial residual.

The next step consists of determining whether the residual from the CR regressions represents noise, versus information relevant to student knowledge and understanding of course material. To address this question, we use MC data and the CR-residual from the term test to determine whether the CR data can help predict student performance on the final exam. If the residual variable were insignificant, that would suggest that the CR-residual was just noise. In contrast, we find that the CR-residual is large

[2] The only study that we are aware of that attempts such an approach is Kennedy and Walstad (1997). They frame the decision to use CR questions as a tradeoff between reduced 'misclassifications' and higher marking costs. 'Misclassifications' are defined as estimated differences in the grade distribution (beyond natural sampling variation) that would arise on the AP microeconomics and macroeconomics exams from switching to an all-MC format. Unfortunately, in order to categorise these as 'misclassifications,' KW must assume that the mix of CR and MC questions on the AP tests is optimal. If the mix is not optimal, then it doesn't follow that the grade distribution under an all-MC format is worse than under the mixed format. This highlights the practical difficulties of implementing the 'benefits versus costs' approach.

in size and statistically significant. Since the residual represents the component of CR scores that cannot be explained by MC scores, and since it is significantly correlated with final exam performance, we conclude that CR questions contain information about student knowledge and understanding that is not contained in the original set of MC questions.

It is possible that the information provided by the CR-residual supplies the same information that could have been provided by additional MC questions. In other words, our results to this point are not able to help our hypothetical instructor decide whether to use CR questions or additional MC questions. To address this question, the third step constructs a pseudo-counterfactual experiment. We use MC and CR data from the midterm and the final exam to measure whether the CR component of a test provides more information than additional MC questions in explaining students' GPAs in other courses. In each of our 12 sub-samples, the CR component provides substantially more explanatory power than additional MC questions. This suggests that CR questions contain useful information beyond MC questions that may be helpful in assessing students' learning.

The second half of our study investigates why our research obtains results that are at variance with many previous studies. We are able to replicate the key findings of a number of these studies. This suggests that our different results are not driven by differences in the data, but by differences in empirical methodologies. Finally, we recognise that our results reflect the nature and quality of our questions. Therefore, we describe the makeup of the respective MC and CR questions that were used in our research. We conclude with caveats regarding the interpretation and application of our research, and recommendations for future research.

2. Data

Our analysis uses data compiled over a six-year period (2002–07) from approximately 8400 students in two different courses at the University of Canterbury in New Zealand. Introductory Microeconomics and Introductory Macroeconomics are semester-long courses typically taken by business students in their first year of study. Both courses administer a mid-semester term test and an end-of-semester final exam.

Both term tests and final exams consist of a CR and a MC component. While the weights given to these components are different for the term test and the final exam, and change somewhat over the years, the structure of these components has remained constant. For both courses, the term test is 90 minutes long and consists of 25 MC and two CR questions. The final exam is longer at 180 minutes, and consists of 30 MC and three CR questions. There was little change in the coverage of the respective assessments over the years with one exception: in 2007, the final exam gave more coverage to material in the first half of the course. Inasmuch as possible, quality control across assessments was maintained by the fact that the same two instructors taught the classes, and wrote and graded the assessments across the whole time period. The Kruder-Richardson-20 statistics for the MC sections are consistently around 0.7. This indicates a good level of reliability for testing instruments that are measuring multiple dimensions, constructs or areas of interest (Nunnally, 1978).

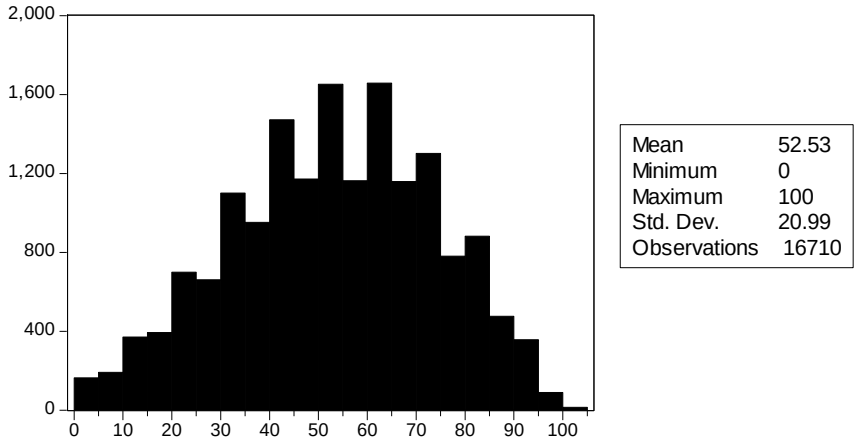
All together, the data set includes assessments from 10 separate offerings of Introductory Microeconomics and eight of Introductory Macroeconomics, for a total of 36 assessments (18 term tests plus 18 final exams). When we eliminate incomplete records and students for whom one of the assessments is missing, we are left with 16,710 observations.[3] By way of comparison, Walstad and

[3] The main reasons for deleting observations were the following: (i) A student received an aegrotat pass. Students apply for an aegrotat pass when they are unable to attend an assessment or their performance has been impaired due to illness or other unforeseen circumstances. (ii) A student had a missing term test or final exam score for some other reason. (iii) A student received a total score for the course equal to zero.

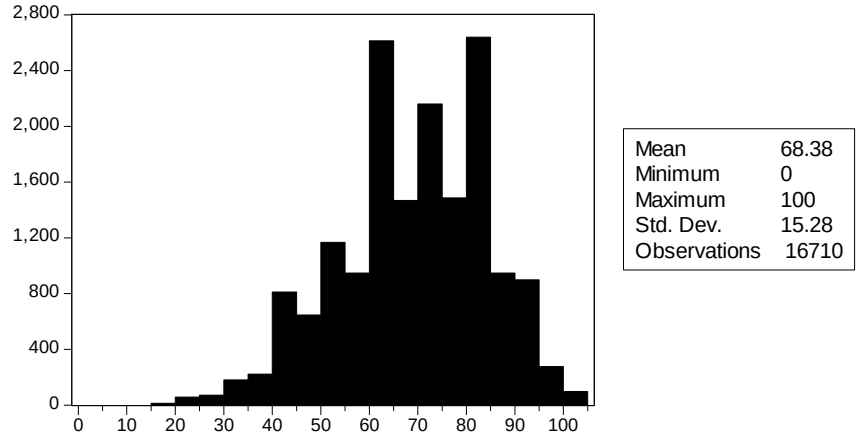
Becker (1994) have a total of 8842 observations and Becker and Johnston (1999) have 4178. Most studies have far fewer.[4]

Figure 1

PANEL A: Constructed-response scores



PANEL B: Multiple-choice Scores



There are two features which make our data set unique. First, we have repeated observations – i.e. both a midterm test and a final exam – for the same student for a given course. Second, we have data about the student's achievement in other courses. We exploit both of these features in our analysis.

The two key variables in our study are student scores on the CR and MC components of their term tests/final exams. These are calculated as percentages out of total possible scores. Panel A of Figure 1 reports a histogram and statistical summary for the full sample of CR scores. The average score is 52.53, and there is evidence of clumping as a result of the way in which the percentage scores are calculated. The lower panel of Figure 1 provides a similar report for the MC scores in our study. These are characterised by a higher mean (68.38) and smaller spread.

[4] For example, Krieg and Uyar (2003) have only 223 observations.

Also noteworthy in Figure 1 is that the distribution of test scores is constrained to lie between 0 and 100. Amongst other problems, this will cause the errors associated with a linear regression specification to be heteroscedastic. We address this problem in two ways. First, we use OLS but estimate the standard errors using the heteroscedastic-robust White procedure. OLS has the advantage of facilitating interpretation of the coefficient estimates. Accordingly, these are the results we report in our paper. However, we also estimate the key regressions using the more statistically appropriate fractional logit procedure. The results are virtually identical.[5]

Table 1 provides a statistical summary of the students represented in our study. Approximately 55% of the sample derive from Introductory Microeconomics classes. By construction, the data set consists of exactly half term test and half final exam results. Table 1 also breaks down the CR and MC scores by term test and final exam. Both components show higher scores on the final exam. This is consistent with the fact that the term test is more time-constrained than the final exam. While the final exam has twice the allotted time as the term test, it is designed to require less than twice the work.

The variable *GPA* reports the student's grade point average for all courses outside of ECON 104 (Introductory Microeconomics) and ECON 105 (Introductory Macroeconomics) in the same year that the student was enrolled in the respective economics class. For example, if a student was enrolled in ECON 104 in Semester 1 of 2005, *GPA* reports their grade point average for all courses they took in calendar year 2005, excluding both ECON 104 and 105.[6] Grade points range from –1 (for a letter grade of E = fail) to 9 (for a letter grade of A+). The variable *COMPOSITE* is a weighted average of the CR and MC components, and is used later in the study when we estimate Walstad and Becker (1994)-type regressions.

Table 1: Statistical summary of data

Variable	Observations	Mean	Minimum	Maximum	Std. dev.
Micro	16,710	0.554	0	1	0.497
Term test	16,710	0.500	0	1	0.500
Constructed-response (Term Test)	8,355	50.0	0	100	20.4
Constructed-response (Final Exam)	8,355	55.0	0	100	21.3
Multiple-choice (Term Test)	8,355	66.8	0	100	15.7
Multiple-choice (Final Exam)	8,355	69.9	16.7	100	14.7
GPA	16,710	3.53	–1	9	2.49
Composite	16,710	63.1	10	100	15.5

While not reported in Table 1, approximately 56% of the sample is male. A little less than half of the students in our sample are New Zealand natives or of European extraction. Approximately 43% of the students are Asian. This high percentage is due to a surge in Asian enrolments that occurred in the early 2000s in New Zealand universities. This tapered off substantially in the latter years of the sample. Maori, Pacific Islanders and Others (primarily Africans and Middle Easterners) account for less than 8% of our sample. With respect to language, a little more than 60% of the sample declared English to be their first language. The great majority of the remainder identified with Chinese.

[5] The fractional logit results are available from the authors upon request.

[6] We chose to exclude both introductory economics classes because of similarities in the way the two classes were assessed. Since the two lecturers work closely together, it is possible that their assessment styles were similar. Correlation in performance across the two classes might represent students' ability to perform well on a particular style of assessment, and not an independent observation about student learning outcomes.

3. Results

Step one

The first step of our analysis consists of determining to what extent performance on the CR component of an assessment is 'predictable' from the student's MC score on that assessment. If the corresponding regressions produce R^2 values close to 1, this would clearly indicate that CR scores added little information to that already provided by the student's MC performance. In this case, our hypothetical instructor would be better off discarding the CR component and using an all-MC assessment.

Table 2 summarises the results of this analysis. We divided our data set into four, mutually exclusive sets of observations: (i) term tests and (ii) final exams from Introductory Microeconomics classes; and (iii) term tests and (iv) final exams from Introductory Macroeconomics classes. For each sample, we regressed students' CR scores on their MC scores for the same assessment. In addition, we aggregated all the observations into one sample. Not surprisingly, we find that MC scores are significant predictors of students' CR scores. An extra percentage point on the MC component predicts an additional 0.7 to 1.1 percentage points on the CR component, depending on the sample.

Table 2: Predicting constructed-response scores using multiple-choice scores

	Sample				
	Micro/Term Tests (1)	**Micro/Final Exams (2)**	**Macro/Term Tests (3)**	**Macro/Final Exams (4)**	**All Observations (5)**
Constant	–7.4980 (–6.72)	–12.1581 (–11.69)	6.1509 (5.79)	–21.2494 (–18.03)	–6.0626 (–10.69)
Multiple-Choice	0.8097 (50.96)	0.9832 (67.81)	0.7143 (43.55)	1.0608 (67.28)	0.8568 (106.63)
Observations	4628	4628	3727	3727	16710
R^2	0.347	0.470	0.318	0.508	0.389
Simple Correlation	0.589	0.686	0.564	0.713	0.624

NOTE: Values in parentheses are *t*-statistics calculated using heteroscedastic-robust (White) standard errors.

On the other hand, we also find that the R^2 values are never close to 1. The R^2 values for the final exam regressions are close to 50%. Those for the term tests are even lower, in the low- to mid-30s.[7] (We discuss this difference between term tests and final exams below.) For the full sample, the R^2 of the regression of CR scores on MC scores is a little less than 40%.[8]

To facilitate comparison with other studies, the last line of the table reports the simple correlation between CR and MC scores. Walstad and Becker (1994, p. 194) report simple correlations of 0.69 and

[7] Conventional wisdom is that CR questions are 'noisier' assessments. This view is supported by the fact that CR scores have greater dispersion (cf. Figure 1 and Table 1).

[8] We also investigated the effect of including higher-order, polynomial terms for the MC variable. This added little to the overall explanatory power of the equations.

0.64 for the Micro and Macro AP tests. Lumsden and Scott (1987, p. 367) report correlations of 0.18 and 0.26 for introductory Micro and Macro courses, respectively. In contrast, they cite a number of other studies where the correlations range higher, though still lower than reported here. Thus, our finding that CR scores are far from being perfectly, or even near perfectly, predictable from MC scores appears to be the norm.

Unfortunately, while an R^2 close to 1 provides strong evidence that CR and MC questions measure the same thing, it is unclear what an R^2 far from 1 implies. Is the unexplained component in CR scores due to the fact that these measure aspects of students' knowledge and understanding that are not measured by MC questions? Or are the two question types assessing the same thing(s) but with noise/measurement error?

Step two

If we had an alternative measure of student knowledge and understanding, we could take the residuals from the regressions in Table 2 and test if they were significant predictors of this alternative measure. If the residuals were unrelated to student knowledge and understanding, say were pure measurement error, then one would expect them to be uncorrelated to this alternative measure. Alternatively, if we could show that these residuals were positively related to this alternative measure, this would provide evidence that the residuals contained information about student knowledge and understanding that was not captured by MC responses.

Unfortunately, we do not have an alternative measure of student knowledge and understanding for the same assessment. We do, however, have a close substitute. Because we have repeated observations for each student, we can test whether residuals from the term test regressions are related to achievement on the final exam. If the residuals represent pure measurement error, one would not expect to find any relationship with students' final exam performance.

Column (1) of Table 3 reports the results of a regression where students' CR scores from the final exam were regressed on (i) their MC scores from the term test, and (ii) the unexplained component of their CR score from the term test (i.e. the residual from the regression specification that was reported in Table 2).[9] We separate the 2002–06 and 2007 final exams because the 2007 final exams included a larger share of material from the first half of the course. We also separate the Introductory Microeconomics and Introductory Macroeconomics final exams. In each of the six samples, the *Residual* variable has very large *t*-values. In addition, the respective coefficients are all positively-signed.[10]

Our results are evidence that CR scores contain information not contained in the existing MC scores, and that this information is correlated with student academic performance. But is this 'information' really related to students' knowledge and understanding of course material? For example, suppose students with bad handwriting receive lower marks on CR questions, *ceteris paribus*. Then a lower score on the term test CR section could be predictive of a lower score on the final exam CR section because it was predictive of bad handwriting.

[9] The residual variables come from term test CR regressions using the same observations as the Table 3 samples (e.g. 'All Observations (2002–2006)', 'All Observations (2007)', etc.). Note that we would get the same coefficient and *t*-statistics for this variable if we substituted the actual CR (Term Test) variable for the associated residual (see Johnston and DiNardo, 1997, p. 82). We use the residual variable to emphasise that this variable contains information that is independent of the information contained in the MC (Term Test) variable.

[10] At the suggestion of a referee who was concerned that our results might be an artifact of a given year's type of assessment or demographic composition of test-takers, we re-estimated the regressions in Table 3, breaking out the observations by year and subject area (Macroeconomics, Microeconomics). The residual from the CR term test regression remained a significant determinant of final exam performance in every case (a total of 22 regressions). The results are available from the authors.

To check this possibility, we also regressed students' final exam MC scores on the same two variables used to predict their final exam CR scores. The qualitative results remain unchanged. For each sample, the *Residual* variable is positively correlated and highly, statistically significant. In other words, the unexplained component of term test CR scores predicts student achievement on both the (i) CR and (ii) MC components of the final exam.

While this latter finding is strong evidence that the CR residuals contain information about student knowledge and understanding, it raises another concern: if CR and MC questions measure something different, why should the term test CR residual have predictive power for the final exam MC score?

Table 3: Predicting final exam performance from term test scores

Variable	Dep. Variable = Constructed-Response (Final Exam) (1)	Dep. Variable = Multiple-Choice (Final Exam) (2)
Sample (1a): All observations (2002–2006)		
Constant	7.5982 (9.55)	37.3361 (60.72)
Multiple-Choice (Term Test)	0.7152 (63.24)	0.4933 (57.12)
Residual from Term Test Constructed-Response Regression	0.5292 (49.49)	0.3092 (38.97)
R^2	0.468	0.410
Observations	7270	7270
Sample (1b): All observations (2007)		
Constant	–12.2469 (–5.97)	25.8495 (14.34)
Multiple-Choice (Term Test)	0.9591 (33.80)	0.6170 (25.09)
Residual from Term Test Constructed-Response Regression	0.6331 (22.03)	0.2198 (11.80)
R^2	0.579	0.415
Observations	1085	1085
Sample (2a): Micro (2002–2006)		
Constant	–0.6955 (–0.58)	27.7901 (30.76)
Multiple-Choice (Term Test)	0.7954 (48.93)	0.5879 (48.29)
Residual from Constructed-Response Regression	0.4710 (31.79)	0.2740 (25.20)
R^2	0.459	0.4424
Observations	3947	3947

Variable	Dep. Variable = Constructed-Response (Final Exam) (1)	Dep. Variable = Multiple-Choice (Final Exam) (2)
Sample (2b): Micro (2007)		
Constant	–12.7999 (–4.93)	23.3048 (11.25)
Multiple-Choice (Term Test)	0.9946 (26.90)	0.6108 (21.07)
Residual from Term Test Constructed-Response Regression	0.6112 (17.21)	0.2547 (11.73)
R^2	0.578	0.454
Observations	681	681
Sample (3a): Macro (2002–2006)		
Constant	9.6417 (8.77)	40.0442 (46.99)
Multiple-Choice (Term Test)	0.7335 (44.66)	0.5055 (39.99)
Residual from Term Test Constructed-Response Regression	0.5757 (33.66)	0.2808 (22.50)
R^2	0.486	0.404
Observations	3323	3323
Sample (3b): Macro (2007)		
Constant	–13.8856 (–4.07)	34.2929 (12.53)
Multiple-Choice (Term Test)	0.9375 (20.68)	0.5685 (15.96)
Residual from Term Test Constructed-Response Regression	0.6663 (13.09)	0.3167 (9.80)
R^2	0.581	0.479
Observations	404	404

NOTE: Values in parentheses are *t*-statistics calculated using heteroscedastic-robust (White) standard errors.

Our explanation recalls a number of previously noted characteristics about our data, and combines this with the educational psychology literature on learning goals. First, both CR and MC scores are lower for the term test than the final exam. Second, the R^2 values from the term test regressions in Table 2 are lower than the corresponding final exam regressions. Third, the term test is more time-constrained than the final exam, as evidenced by lower mean CR and MC scores (cf. Table 1).

Bloom's (1956) taxonomy predicts that MC questions are more likely to test the lower levels of educational objectives (i.e. Knowledge, Comprehension, Application and, perhaps, Analysis). While CR questions test these as well, they are uniquely suited for assessing the more advanced learning goals

(Synthesis and Evaluation).[11] Accordingly, one would expect CR to contain some unique information compared to MC, but also some overlap.

We now attempt to explain both the poorer predictability of MC scores on term tests (cf. Table 2), and the fact that term test CR scores are significant predictors of final exam MC scores (cf. Column 2, Table 3). Given the greater time constraints, we hypothesise that students will devote relatively less time to the MC component on the term test; since MC questions can be answered very quickly, if necessary. However, the cost of this test-taking strategy is that students are less likely to get the more difficult MC questions (Application and Analysis) correct. It is these more difficult MC questions that will test higher levels of learning.

As a consequence, the amount of informational 'overlap' between the MC and the CR questions – as measured by the levels of educational objectives that are assessed – is likely to be lower for the term test than for the final exam. This will cause MC scores to be a worse predictor of CR scores on term tests compared to final exams. It will also cause the MC responses on the final exam to measure higher levels of knowledge and understanding than the MC responses on the term test. Because the CR responses also assess these higher levels, the CR *Residual* will be able to predict final exam MC scores even after controlling for term test MC scores.

The fact that (i) the CR-residual is a significant determinant of MC scores on the final exam, and (ii) the MC variable explains a smaller amount of variation in CR scores on term tests compared to final exams, is consistent with the hypothesis that the CR-residual measures higher-level learning according to Bloom's taxonomy (Bloom, 1956).

Step three

Summarising the above, our results suggest that CR scores contain information not contained in the responses to *existing* MC questions. However, we are still not in a position to help our hypothetical instructor decide whether to use CR questions or additional MC questions: perhaps the additional information provided by the CR questions is merely a substitute for information that could have been provided by including more MC questions. To address this concern, we would like to compare assessments using composite MC/CR questions with those using all-MC questions.

We could empirically address this if we were able to perform the following experiment. Suppose there were two groups of identical students. One group was given a composite test composed of MC and CR questions. Call these variables *MC1* and *CR*. The other group was given a test composed entirely of MC questions, where the first half of the questions was identical to what the first group received. Call these two sets of MC questions, *MC1* and *MC2*. Finally, suppose we had some objective measure of a student's knowledge and understanding of course material. Call this variable *Y*.

Now consider two regression models:

$$Y_i = \beta_0 + \beta_1 MC1_i + \beta_2 MC2_i + \varepsilon_i \quad (1)$$

$$Y_i = \alpha_0 + \alpha_1 MC1_i + \alpha_2 CR_i + v_i \quad (2)$$

If the CR questions contained the same 'information' as the additional MC questions, then the specification of Equation (2) should have approximately the same explanatory power as the specification of Equation (1). Alternatively, if the R^2 value for Equation (2) was smaller than that for Equation (1), that would suggest that the CR questions were less efficient at 'explaining' students'

[11] The six levels of Bloom's taxonomy are sometimes recast as follows (from lowest to highest): (i) Remembering, (ii) Understanding, (iii) Applying, (iv) Analysing, (v) Evaluating, and (vi) Creating.

understanding than the additional MC questions. If the R^2 value for Equation (2) were greater, that would indicate that the CR questions contained information that had greater explanatory power than the additional MC questions.

We could get to this conclusion because we have a counterfactual to compare our composite test results with: One group takes a test composed entirely of MC questions. The other group takes a composite test composed of both MC and CR questions. Unfortunately, our data does not contain a real counterfactual. Instead, we manipulate our data to create a pseudo-counterfactual.

A unique feature of our data is that we have information on students' grades in every course they have taken at the University of Canterbury. As discussed above, we use this information to calculate a GPA value based on their performance in non-introductory economics classes. We use this GPA variable to proxy for *Y* in the experiment described by Equations (1) and (2) and the subsequent discussion. Our working assumption is that GPA in non-economics classes is positively correlated with students' knowledge and understanding of course material in their economics principles class. Our rationale is that students who have a good understanding of course material in one class are also likely to get high grades in their other classes (because better students are more likely to have good knowledge and understanding in all their classes).

For each student in a given principles of economics course, we also have their MC score on the (i) term test and (ii) final exam in that course; and their CR score on the (iii) term test and (iv) final exam. We divide our observations into the same six samples that we used in Table 3. With reference to Equations (1) and (2) above and the corresponding discussion, let *MC1* be the MC component on the term test, and let *MC2* and *CR* be the MC and CR scores from the final exam. Note that *MC2* and *CR* should be from the same assessment to make the comparison as clean as possible.

Using the same logic as above, if the R^2 values are higher from the equations with the *CR* component, that suggests that the CR responses contain more/better information than the *MC2* responses – and not just the same information. Accordingly, we compare the following regression models:

(i) $\text{GPA}_t = \beta_0 + \beta_1 \text{MC(Term)}_t + \beta_2 \text{MC(Final)}_t + \varepsilon_t$, and

(ii) $\text{GPA}_t = \alpha_0 + \alpha_1 \text{MC(Term)}_t + \alpha_2 \text{CR(Final)}_t + \eta_t$.

To recapitulate, the pair of models above proxies for the following thought experiment: Suppose an instructor had given an all-MC term test. Would he or she more effectively assess academic achievement if the final exam consisted of all MC questions, or a mix of CR and MC questions? Specification (i) represents the case where assessment is based solely on MC questions. Specification (ii) represents a composite CR/MC assessment. If MC and CR questions measure the same thing(s), a comparison of the R^2 values from estimating models (i) and (ii) across different samples should show no clear pattern. However, if CR questions measure information not captured by the additional MC questions – such as higher levels of the Bloom (1985) taxonomy – then the R^2 values from Specification (ii) regressions should be consistently higher.

As a further test, we also compare an alternative pair of regression models:

(iii) $\text{GPA}_t = \beta_0 + \beta_1 \text{MC(Final)}_t + \beta_2 \text{MC(Term)}_t + \varepsilon_t$, and

(iv) $\text{GPA}_t = \alpha_0 + \alpha_1 \text{MC(Final)}_t + \alpha_2 \text{CR(Term)}_t + \eta_t$.

Table 4 reports the results of this test. We divide the data into the same six samples used for Table 3. Consider the first two rows of Table 4. For the sample of all observations from 2002–06 (Sample 1a),

the regression of *GPA* on the two MC components produces an R^2 value of 0.424. In contrast, the 'composite' regression of one MC and one CR component has an associated R^2 value of 0.526. The composite 'assessment' does a better job of predicting student achievement. Rows (3) and (4) perform a similar comparison, this time starting with the *MC(Final)* score and adding either the *MC(Term)* or *CR(Term)* score. Once again, the composite 'assessment' does a better job of predicting student achievement. In fact, for every sample and every pair of regression models, a combination of CR and MC scores does a better job of predicting students' *GPAs* than relying solely on MC scores.

Table 4: Predicting student GPAs: would an all-multiple choice assessment be better?

	Estimated coefficients			
	Multiple-Choice (Term Test)	**Multiple-Choice (Final Exam)**	**Constructed-Response (Term Test)**	**Constructed-Response (Final Exam)**
I. Sample (1a)				
A. MC(Term) + [MC(Final) OR CR(Final)]:				
(1) R^2 = 0.424	0.0392 (23.25)	0.0811 (46.10)	----	----
(2) R^2 = 0.526	0.0277 (18.44)	----	----	0.0719 (65.76)
B. MC(Final) + [MC(Term) OR CR(Term)]:				
(3) R^2 = 0.424	0.0392 (23.25)	0.0811 (46.10)	----	----
(4) R^2 = 0.485	----	0.0634 (35.53)	0.0491 (37.60)	----
II. Sample (1b)				
A. MC(Term) + [MC(Final) OR CR(Final)]:				
(5) R^2 = 0.490	0.0497 (9.75)	0.0864 (18.01)	----	----
(6) R^2 = 0.593	(0.0328 (7.15)	----	----	0.0732 (25.73)
B. MC(Final) + [MC(Term) OR CR(Term)]:				
(7) R^2 = 0.490	0.0497 (9.75)	0.0864 (18.01)	----	----
(8) R^2 = 0.554	----	0.0764 (17.96)	0.0506 (16.13)	----

III. Sample (2a)				
A. MC(Term) + [MC(Final) OR CR(Final)]:				
(9) $R^2 = 0.434$	0.0472 (18.75)	0.0753 (30.18)	----	----
(10) $R^2 = 0.534$	0.0364 (16.42)	----	----	0.0693 (44.23)
B. MC(Final) + [MC(Term) OR CR(Term)]:				
(11) $R^2 = 0.434$	0.0472 (18.75)	0.0753 (30.18)	----	----
(12) $R^2 = 0.468$	----	0.0671 (24.54)	0.0444 (24.54)	----
IV. Samples (2b)				
A. MC(Term) + [MC(Final) OR CR(Final)]:				
(13) $R^2 = 0.515$	0.0406 (6.39)	0.0989 (16.13)	----	----
(14) $R^2 = 0.587$	0.0291 (4.89)	----	----	0.0723 (19.40)
B. MC(Final) + [MC(Term) OR CR(Term)]:				
(15) $R^2 = 0.515$	0.0406 (6.39)	0.0989 (16.13)	----	----
(16) $R^2 = 0.572$	----	0.0834 (15.06)	0.0450 (11.45)	----
V. Sample (3a)				
A. MC(Term) + [MC(Final) OR CR(Final)]:				
(17) $R^2 = 0.421$	0.0410 (16.08)	0.0792 (29.40)	----	----
(18) $R^2 = 0.533$	0.0296 (13.43)	----	----	0.0702 (45.38)
B. MC(Final) + [MC(Term) OR CR(Term)]:				
(19) $R^2 = 0.421$	0.0410 (16.08)	0.0792 (29.40)	----	----
(20) $R^2 = 0.508$	----	0.0593 (22.79)	0.0550 (29.58)	----
VI. Sample (3b)				
A. MC(Term) + [MC(Final) OR CR(Final)]:				
(21) $R^2 = 0.473$	0.0563 (6.83)	0.0855 (9.84)	----	----
(22) $R^2 = 0.630$	0.0289 (3.95)	----	----	0.0811 (18.22)
B. MC(Final) + [MC(Term) OR CR(Term)]:				
(23) $R^2 = 0.473$	0.0563 (6.83)	0.0855 (9.84)	----	----
(24) $R^2 = 0.522$	----	0.0672 (7.42)	0.0600 (9.67)	----

Taken together, the results from Tables 2 through 4 provide evidence that CR questions measure student knowledge and understanding that is not captured by MC questions. Our evidence is consistent with the hypothesis that the CR variable measures higher-level learning, as defined by Bloom's taxonomy (1956). While other studies, such as Kennedy and Walstad (1997) and Becker and Johnston (1999), provide evidence that CR and MC responses are 'different', our study is the first to link these differences to student academic performance in university principles of economics classes.

Note: Values in parentheses are *t*-statistics calculated using heteroscedastic-robust (White) standard errors. Sample numbers (e.g. 1a) identify the respective sample and are identical to the samples in Tables 3 and 4.

4. Relating our findings to those of previous studies

Our finding that CR scores comprise information not contained in MC scores is at variance with a number of influential studies. In this section, we want to explore whether this is due to differences in our data, or differences in empirical procedures.

Bennett, Rock and Wang (1991) and Thissen, Wainer and Wang (1994) are widely-cited studies from the educational measurement literature. BRW base their analysis from a sample of responses from the College Board's Advanced Placement (AP) examination in Computer Science. TWW re-analyse BRW's data, and add a similar sample from the AP exam in Chemistry. Both employ common factor analysis to study the relationship between 'free response' and MC questions. Both find that a single factor explains most of the variation in the respective questions. They therefore conclude that these two question-types measure the same thing.[12]

While BRW and TWW employ factor analyses, they use somewhat different techniques. BRW use a model in which free response and MC questions are each loaded on a single factor. These two (correlated) factors are then analysed to determine whether they contain unique information. In contrast, TWW employ a more general procedure to decompose the variation in the two types of questions into multiple factors.

The AP exam in Computer Science consists of 50 MC questions, and five free-response questions. The AP exam in Chemistry consists of 75 MC questions and four sections of free-response questions, some of which contain multiple problems. BRW and TWW break up the respective components into multiple 'parcels'. BRW re-organise the 50 MC questions into five sets ('parcels') of 10 questions each. TWW convert the original 75 MC questions into 15, five-question parcels. These parcels become, in a sense, separate variables which are then decomposed into factors.

We attempt to replicate BRW's and TWW's factor analysis results. Unfortunately, our data contain fewer questions than BRW and TWW and are thus less amenable to 'parcelisation'. Instead, we apply principal component analysis (PCA) to students' scores on the CR and MC components. PCA is related to factor analysis in that its 'principal components' are akin to the factors identified by factor analysis. It has the advantage in that it produces a unique decomposition of the correlation matrix.[13] In contrast, factor analysis typically involves a subjective procedure ('rotation') that allows one to generate alternative sets of factors from the same data. A particularly attractive feature of PCA for our purposes is that it yields a straightforward measure of the amount of variation 'explained' by each of the principal components.

[12] While both studies find more than one significant factor, they both conclude that a single factor is able to explain most of the variation in the two types of questions.

[13] Non-unique solutions can arise when two or more eigenvalues are exactly equal, but this is rarely encountered in practice.

Table 5 reports the results of applying PCA to the same five samples we previously analysed in Table 2. As there are only two variables (*Multiple-Choice* and *Constructed-Response*), there are a total of two principal components. By construction, these two principal components explain all of the 'variation' in the correlation matrix.

The first item of interest in Table 5 is the column of 'eigenvalues'. These provide a measure of importance for each of the principal components. In factor analysis, two common approaches for choosing the number of 'factors' are Kaiser's eigenvalue rule and Cattell's scree test. The first of these selects factors having eigenvalues greater than 1. The second of these plots the eigenvalues in decreasing order and selects all factors immediately preceding an abrupt levelling off of the values. Both approaches lead to the conclusion that there is one main factor underlying students' CR and MC responses in each of the samples. This finding is reinforced by the second column in Table 5. 'Proportion' translates these eigenvalues into shares of total variation in the correlation matrix. These range from 78–85% across the different samples.

Table 5: Summary of principal component analyses

Sample (1): All Observations		
Principal Component	Eigenvalue	Proportion
1	1.6236	0.812
2	0.3764	0.188
Sample (2): Micro/Term Tests		
Principal Component	Eigenvalue	Proportion
1	1.5846	0.792
2	0.4154	0.208
Sample (3): Micro/Final Exams		
Principal Component	Eigenvalue	Proportion
1	1.6855	0.843
2	0.3145	0.157
Sample (4): Macro/Term Tests		
Principal Component	Eigenvalue	Proportion
1	1.5636	0.782
2	0.4364	0.218
Sample (5): Macro/Final Exams		
Principal Component	Eigenvalue	Proportion
1	1.7129	0.856
2	0.2871	0.144

NOTE: Samples are identical to the samples in Table 2.

In summary, we find evidence that (i) a single factor underlies students' CR and MC responses in our data, and (ii) this single factor is able to explain most of the variation in the respective scores.[14] In

[14] BRW conclude that one factor explains most of the variation by virtue of a battery of goodness-of-fit measures, finding that the second factor adds little in the way of goodness-of-fit. TWW reach this conclusion by noting that the factor loadings on the second factor are relatively small.

other words, when we use an empirical procedure similar to what BRW and TWW employ, we are led to the same conclusion that they reach.

Walstad and Becker (1994) is another study that has been very influential in the debate over CR versus MC questions. Their study analyses AP Microeconomics and Macroeconomics exams. Each of these has CR and MC components from which an overall composite score is formed, with the components receiving weights of two-thirds and one-third, respectively. WB use these data to regress the composite scores on the MC scores. They find that the MC scores explain between 90 and 95% of the variation in composite scores. WB conclude that there are 'no differences, or only slight differences, in what the two types of tests and questions [multiple-choice and constructed-response] measure'.

We construct composite scores from the MC and CR components using the same weights as the AP exams. We then estimate WB-type regressions using the same five samples we used for our original analyses. Table 6 reports the results. Of interest here are the R^2 from the respective regressions. These range between 85 and 90%.[15] Using the same specification, WB obtained an R^2 of 94% for the Microeconomics exams, and an R^2 of 90% for the Macroeconomics exams. Our macro results are about the same as WB's, while our micro results are somewhat lower.[16]

Table 6: Summary of regressions based on Walstad and Becker's (1994) specification

	Sample				
	Micro/Term Tests (1)	**Micro/Final Exams (2)**	**Macro/Term Tests (3)**	**Macro/Final Exams (4)**	**All Observations (5)**
Constant	–2.4999 (–6.72)	–4.0527 (–11.69)	2.0503 (5.79)	–7.0831 (–18.03)	–2.0209 (–10.69)
Multiple-Choice	0.9366 (176.85)	0.9944 (205.76)	0.9048 (165.51)	1.0203 (194.11)	0.9522 (355.55)
Observations	4628	4628	3727	3727	16710
R^2	0.862	0.891	0.871	0.896	0.876

NOTE: The dependent variable is a composite assessment score created by weighting the multiple-choice and constructed-response components by 2/3 and 1/2, respectively. These are the weights used by the Advanced Placement Economics test that was analysed by Walstad and Becker (1994). Samples are identical to the samples in Table 2.

In summary, the strongest evidence that CR and MC questions measure the same thing comes from factor analysis and WB-style regressions. When we replicate these procedures using our data, we get results similar to the original authors. What can we learn from this? It means that one can get different conclusions from the same data, if one uses different methodologies. We argue that our methodology is more directly applicable for the instructor who is trying to decide whether to use a composite MC-CR assessment, versus an assessment composed of all MC questions.

On the other hand, our results are consistent with two studies that have been influential on the other side. Kennedy and Walstad (1997) use simulation exercises to estimate the effect of moving to an all-MC format for the AP test. They report that the number of students who would receive different AP

[15] These results are very similar to those obtained by Krieg and Uyar (2001).

[16] Conveniently, WB report simple correlations between the CR and MC components of the AP exams. These fall in the same range as the correlations we report for our data in Table 2. Thus, it should not be surprising that we are able to produce WB-type regressions that are very similar to theirs.

grades is small but statistically significant. Further, alternative simulation assumptions produce larger effects.

Becker and Johnston (1999) examine results from the Victorian (Australia) Certificate of Education assessment of high school economics. The VCE assessment consists of both MC and CR components. Like previous studies before them, BJ find a high correlation between MC and CR scores. However, when they instrument the explanatory variable with school-wide performance on that component, they find the correlation becomes small in size and statistically insignificant. They therefore conclude that the MC and CR components measure different dimensions of knowledge.

The KW and BJ studies are complements to ours. Both find differences in what MC and CR responses measure. The unique contribution of our study is that we provide evidence that these differences are related to student academic achievement.

5. A closer look at the CR and MC questions analysed in this study

The debate over CR versus MC questions is to some extent an idiosyncratic one that is course- and instructor-dependent. In this section, we first review the literature on the ability of MC and CR questions to measure higher-order learning outcomes. We then describe the CR and MC questions used in the assessments analysed by this study. This information is useful for determining the extent to which our results may be valid for other university, introductory economics courses.

Bloom (1956) defines the following six levels of learning (our expanded explanations are in parentheses);

1. Knowledge (knowing facts);
2. Comprehension (understanding the importance of known knowledge);
3. Application (putting knowledge and understanding to use);
4. Analysis (using knowledge to breaking down a problem into component parts);
5. Synthesis (combining different parts to form new knowledge and ideas); and
6. Evaluation (determining the worth or usefulness of knowledge, application, analysis or synthesis).

Textbook, MC test banks tend to consist of questions that disproportionately sample from the first two levels of learning. Buckles and Siegfried (2006) conclude that MC questions can be effectively used to assess up through the first four levels of Bloom's taxonomy. In contrast, they argue that while it is possible to use MC questions to assess synthesis and evaluation, these are more reliably measured through CR questions. According to Buckles and Siegfried (2006), the key ingredient for assessing these higher-level learning outcomes is the requirement that students work through a chain of reasoning using a number of logical steps. It is difficult to write a sequence of MC questions that get at this learning dimension, especially when the chain of reasoning can involve a complicated decision tree.

These conclusions find support elsewhere in the literature. As part of a wider study, Iz and Fok (2007) attempt to classify the set of 25 MC questions used in the test for the Higher Diploma of Surveying. They classify 21 of the 25 as levels 1 to 4. The remaining four questions were simply lumped together as 'they were few in numbers… and difficult to discriminate'. Zheng *et al.* (2008) assert that it is 'much more difficult to write multiple-choice questions at the application and analysis levels of Bloom's taxonomy than at the knowledge or comprehension levels'. It is even more difficult to write synthesis and evaluation MC questions. Thus it is no surprise that standard textbook question banks are dominated by recognition-, recall- and understanding-type questions.

Walstad (2006) concurs with Buckles and Siegfried to a large extent, but notes that many CR questions are not well-designed to assess higher-level learning. Despite the best of intentions, CR questions may only be testing recall and recognition. A key issue is whether the student could have memorised the answer in advance.

We next describe the nature of the MC and CR questions used in the assessments included in our data set.[17] The first example is a MC question that was designed to test for Knowledge (Level 1 of Bloom's taxonomy).

> *Which of the following is NOT an impact of inflation?*
>
> 1. *Wealth is transferred from savers to borrowers.*
> 2. *Important price signals become more difficult to read.*
> 3. *The currency loses value.*
> 4. *The value of money assets rises.*

The next example is another MC question, but this one was designed to test for Application and Analysis (Levels 3 and 4).

> *A recession in the rest of the world is likely to cause _____ GDP growth and _____ inflation in New Zealand.*
>
> 1. *higher; higher.*
> 2. *higher; lower.*
> 3. *lower; higher.*
> 4. *lower; lower.*

Assessing higher levels of knowledge becomes much more difficult with MC questions. This is where CR questions provide an opportunity to assess levels of knowledge that cannot, or at least are not, being measured by MC questions.

The following example is taken from the same course as the questions above. It illustrates how a CR question can be written such that higher levels of learning are progressively tested as the student works their way through the question.

> *In 1989, the Government passed the Reserve Bank Act. How would you characterise the NZ economy since that time in terms of growth, inflation and unemployment?*

This question tests Knowledge and Comprehension (Levels 1 and 2). It could be easily rewritten in a MC format. Marks were awarded for stating how economic growth, inflation and unemployment had performed over this period in general terms (Knowledge). Marks were also awarded for answers that commented on the importance of these facts (e.g. recent slowing of growth at that time).

[17] The questions are taken from the term-test and final exam for Introduction to Macroeconomics (ECON 105), Semester One, 2006.

A following CR question is:

> *The Reserve Bank Monetary Policy news release above [not shown here] was issued on 9 March 2006. In this release the Bank identifies a number of factors that are influencing both inflation and growth. Use an AD/AS model to explain how the Reserve Bank currently sees the following factors influencing inflation and growth (remembering that the AD/AS model is a static model so you will need to interpret the results).*
>
> *(i) the slowing (or cooling) of the housing market.*
>
> *(ii) labour costs.*
>
> *(iii) business confidence.*

This question tests Application, Analysis and some Synthesis (Levels 3, 4, and 5). Students are required to break down the economic factors identified in the Reserve Bank news release and to use the AD/AS model to analyse the question. The student needs to have a good working knowledge of the AD/AS model because the question does not explicitly identify how AD/AS are affected by the respective factors. Further, the student must bring these factors together to determine their overall impact on growth and inflation. The latter involves extending results from the static model (price and GDP level) to a dynamic world (inflation and growth).

The next CR question follows up the previous one and moves to Synthesis and Evaluation (Levels 5 and 6):

> *If the three influences analysed above were the only factors impacting the NZ economy, what conclusions would you make about the outlook for inflation and growth?*

Students must combine all three answers into one overall judgement. From the answers to the previous question there is no ambiguity about the impact on economic growth but the impact on inflation of these three influences is ambiguous. Students need to recognise this and answer accordingly. The question and the resources provided with the question contain little guidance for the student. Further, students must provide a consistent answer based on their previous answer.

Typically, students who have learnt some facts will achieve a good score on the first CR question. Students who have learnt the mechanics of the AD/AS model will earn at least some of the marks for the second CR question. The most able students will earn marks for the last CR question.

These latter examples are designed to illustrate the difficulty with writing MC questions to assess the highest levels of learning. These levels of learning are best assessed when the student is asked to analyse a complex economic question that requires them to assemble a chain of logical arguments. Consider the problem of assessing such a problem with MC question(s). If a single MC question is used to assess a problem of great complexity, fairness would dictate that it be worth many more points than simple recognition, MC questions. But the all-or-nothing marking of MC questions makes this a risky measure. In contrast, if a sequence of MC questions are used to assess the different parts of the logical chain, it is difficult to not lead the student into the answer by virtue of asking the question(s). The

combination of their free-response nature, along with partial-credit marking, endows the CR question format with the potential to better assess higher-level learning while maintaining fairness to students.

6. Conclusion

This study provides evidence that constructed response (CR) questions contribute information about student knowledge and understanding that is not contained in multiple choice questions (MC). This finding may be useful to university instructors of principles of economics classes trying to decide whether to use constructed response (CR) questions on assessments, with their higher marking costs; or to employ all multiple choice (MC) questions.

To address this issue, our study empirically investigates the relationship between CR and MC questions using a data set compiled from several years of university introductory economics classes. Similar to other studies, we find that MC questions are able to explain, at best, about 50% of the variation in CR scores. However, unlike other studies, we are able to provide evidence that the corresponding residuals are related to student knowledge and understanding. Specifically, we find that the component of CR scores that cannot be explained by MC responses is positively and significantly related to performance on a subsequent exam in the same course.

However, the key issue for instructors considering a switch to an all-MC format is whether CR questions provide information that could not be obtained by expanding the set of MC questions. We exploit the panel nature of our data to construct a quasi-counterfactual experiment. We show that combining one CR and one MC component always predicts student achievement better than combining two MC components.

A final contribution of our study is that we demonstrate that empirical approaches that rely on factor analysis or Walstad–Becker (1994)-type regressions lead to different conclusions about the relationship between CR and MC questions when applied to our data. We argue that our methodology is more directly applicable for an instructor trying to decide between a composite MC/CR assessment, and an assessment composed entirely of MC questions.

We have two sets of cautions with regard to interpretation and application of our results. First, although this study employs a large number of observations, these all come from two courses at a single university. It is difficult to determine how generalisable these results may be. While we have attempted to give the reader an understanding of the type of MC and CR questions used in these courses, the most direct way to establish external validity is to replicate our methodology using data from other principles of economics classes.

Second, while this study presents evidence that CR questions contain information not contained in MC questions, it does not address the practical importance of this additional information. For example, if the use of CR questions resulted in only a small modification of students' grades, then our hypothetical instructor trying to decide between a MC/CR and an all-MC assessment might well choose the latter. The only study that has attempted to measure the effects of switching from a composite to an all-MC test is Kennedy and Walstad (1997). Their study focuses on AP test results, and they found little difference in outcomes between MC/CR and all-MC assessments. There are no studies that attempt to do the same for university classes. We hope this study will stimulate further work on this topic.

References

Becker, W. E. and Johnston, C. (1999). 'The Relationship between Multiple Choice and Essay Response Questions in Assessing Economics Understanding', *Economic Record,* vol. 75, pp. 348–57.

Bennett, R., E., Rock, D., A. and Wang, M. (1991). 'Equivalence of Free-Response and Multiple-Choice Items', *Journal of Educational Measurement,* vol. 28(1), pp. 77–92.

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals, handbook 1: Cognitive Domain,* New York: McKay.

Buckles, S. and Siegfried, J. J. (2006). 'Using Multiple-Choice Questions to Evaluate In-Depth Learning of Economics', *Journal of Economic Education,* vol. 37, pp. 48–57.

Iz, H. B. and Fok, H. S. (2007). 'Use of Bloom's Taxonomic Complexity in Online Multiple Choice Tests in Geomatics Education', *Survey Review,* vol. 39(305), 226–37.

Johnston, J. and DiNardo, J. (1997). *Econometric Methods,*. New York: The McGraw-Hill Companies, Inc.

Kennedy, P. E. and Walstad, W. B. (1997). 'Combining Multiple-Choice and Constructed Response Test Scores: An Economists View', *Applied Measurement in Education,* vol. 10(4)*,* pp. 359–75.

Krieg, R., G. and Uyar, B. (2001). 'Student Performance in Business and Economic Statistics: Does Exam Structure Matter?', *Journal of Economics and Finance,* vol. 25(2)*,* pp. 229-241.

Lumsden, K.G. and Scott, A. (1987). 'The Economics Student Reexamined: Male-Female Differences in Comprehension', *Journal of Economic Education*, vol. 18(4), 365–75.

Nunnally, J. (1978) *Psychometric Theory*, New York: McGraw-Hill.

Thissen, D., Wainer, H. and Wang, X. (1994). 'Are Tests Comprising Both Multiple-Choice and Free-Response Items Necessarily Less Unidimensional Than Multiple-Choice Tests? An Analysis of Two Tests', *Journal of Educational Measurement,* vol. 31, pp. 113–23.

Wainer, H. and Thissen, D. (1993). 'Combining multiple-choice and constructed response test scores: Towards a Marxist theory of test construction', *Applied Measurement in Education,* vol. 6, pp. 103–18.

Walstad, W. (2006). 'Testing for Depth of Understanding in Economics Using Essay Questions', *Journal of Economic Education,* vol. 37, pp. 38–47.

Walstad W. and Becker, W. E. (1994). 'Achievement Differences on Multiple-Choice and Essay Tests in Economics', *American Economic Review,* vol. 84, 193–96.

Zheng, A. Y., Lawthorn, J. K.,Lumley, T. and Freeman, S. (2008). 'Application of Bloom's Taxonomy Debunks the "MCAT Myth"', *Science,* vol. 319, pp. 414-15.

Author Biography

W. Robert Reed is Professor of Economics at the University of Canterbury. His previous appointments were at the University of Oklahoma and Texas A&M University. His main areas of research are public economics and applied econometrics, with a particular research interest in taxes and economic growth. He has published in the Journal of Political Economy, the Journal of Public Economics, the Journal of Labour Economics, and elsewhere. This is his first foray into economics education.

Stephen Hickson is a Teaching Fellow at the University of Canterbury (UC). He has been at UC full time

since 2003. Prior to that appointment he worked full-time for Statistics New Zealand and part-time for UC. His main areas of research are in economics education particularly assessment. Stephen also teaches on the University of Canterbury MBA programme.

Contact details

W. Robert Reed and Stephen Hickson
University of Canterbury
Private Bag 4800
Christchurch 8140
Tel: +64 3 366 7001
Fax: +64 3 364 2635
Email: bob.reed@canterbury.ac.nz and stephen.hickson@canterbury.ac.nz

Does Pluralism in Economics Education Make Better Educated, Happier Students? A Qualitative Analysis

Andrew Mearman, Tim Wakeley, Gamila Shoib and Don Webber

Abstract

This paper contributes to the debate on pluralism in the economics curriculum. Here pluralism means a diversity of theoretical perspectives. One set of pedagogical arguments for pluralism are those found in 'liberal' philosophy of education. To this end, the first part of the paper presents arguments for pluralism based on 'liberal' pedagogical arguments. The paper also notes more instrumental arguments for pluralism and the barriers to such an approach. Finally, the paper considers new primary evidence from focus groups on student perceptions of economics. This evidence shows support for the arguments that a pluralist curriculum is popular and develops cognitive capacities of criticism, comparison and analysis – exactly those argued for in (liberal) pedagogical discussion – as well as judgement, understanding and writing skills. However, pluralism as a teaching strategy may be more difficult for those delivering it.

JEL classification: A20, A13, A12, B40

1. Introduction

This paper aims to contribute to the debate on pluralism in economics education. It builds on arguments, for example in Clarke and Mearman (2003), that in constructing curricula, the educational aims of a programme must be considered (and preferably expressed) and that content and teaching method should be organised as far as is practical with those aims in mind. In particular, 'liberal' pedagogy suggests specific aims. Clarke and Mearman (2003) argued indirectly for pluralism via an argument for the teaching of Marxism. This paper builds on that work by developing these arguments for pluralism directly; and by presenting new primary evidence on student perceptions of economics.

The first part of the paper examines specific liberal educational arguments for a pluralist curriculum. The main argument is that students will be better educated in pluralist curricula. The second part of the paper then considers barriers to this pluralist approach. The third part presents evidence from research into student perceptions of economics via focus groups (and some student evaluations). This evidence shows tentative support for the arguments that a pluralist curriculum is popular and develops cognitive capacities of criticism, comparison and open-mindedness – exactly those argued for in liberal pedagogy – as well as judgement, understanding and writing skills. Naturally, these results are preliminary and suggest several directions for further empirical work.

2. Some educational arguments for pluralism

Here pluralism refers mainly to a diversity of theoretical perspectives. There are several arguments for pluralist curricula. These include (ontological) claims that the nature of the world is such that no single theory could explain it (see Dow, 1996, 1997, 2008; Mäki, 1997; Holcombe, 2008); and (epistemological) claims that no single standard exists for adjudging one theory as being the best one, and that all theories are fallible (see Budzinski, 2008; Mearman, 2008). These claims could *per se* create an argument for a pluralist pedagogy. If the goal of education were to prepare students for the world, educating them in a way which inculcates ways of thinking suited to its actual state would seem sensible. For example, students might be taught about complexity, given the considerable evidence that the world operates like a complex system. However, arguments such as these for the superiority of one position run into the epistemological arguments about the fallibility of theories. Further, it is possible to be a pedagogical pluralist without being either ontologically or epistemologically pluralist. Recently, other authors have made pedagogical arguments for pluralist curricula. Many of these are present, and are surveyed, in Reardon (2009). Fullbrook (2009) draws on arguments similar to those above. Van Dalen (2003) offers arguments which suggest that through economic mechanisms such as competition and creativity, a better economics would be produced. It would be more useful to students as a result.

Linked to that point is another general pedagogical argument that should be made which suggests the advantages of pluralism. It utilises the distinction between the 'cognitive' and 'affective' domains of education (from Bloom, *et al.,* 1964, p. 57). To summarise, in order for the cognitive dimension of the student to develop, they must be engaged with the material. This can be achieved in many ways, e.g. by the use of examples. We may also appeal to the instrumentalism of students by suggesting that concept *x* is worth learning because it will help them earn more. Another way to gain students' interest is to relate to contemporary real problems. Indeed, at a recent event at one of the authors' institutions, 250 students attended an extra-curricular 90-minute event discussing the 'credit crunch'. Moreover, the importance of realisticness is implicit in several pedagogical literatures, including experiential research, Inquiry-Based Learning, Problem-Based Learning, and Action Research (see Laurillard, 2008; Fry and Love, 2007). Research on computer games suggests verisimilitude is an important feature in appealing to players (for example, see Schultze and Rennecker, 2007). Gruene-Yanoff (2009) suggests credibility of theories is also important.

Klamer and Colander (1987) suggest that students find relevance of theory to be important to them; and that economics graduate students found the irrelevance of much Economic theory disengaging (see also Johnston, *et al*, 2000). For many students, excessive theorisation (and a lack of realisticness) can be significant in reducing interest in economics (Mearman, 2008). Analysis of qualitative data collected also suggests that unrealistic assumptions and models are a serious problem in achieving student interest. Some argue that some perspectives in economics tend to display these (undesirable) traits less than others (Lawson, 2003) and thus it can be argued that they can be more engaging. The growth of the Post Autistic Economics movement suggests this to be the case (Fullbrook, 2003). However, a potentially disengaging facet of so-called heterodox economics is that for some it appears incoherent. Nonetheless, teaching pluralistically, perhaps through current events, may be more engaging than simply learning one set of theories. Recent economic events have led to a resurgence in interest in several non-mainstream economists, including Keynes, Marx and Minsky. Usually, too, these current events involve complex debates, which our evidence suggests can be engaging.

3. The role of educational aims

From an educational point of view, though, many of the above arguments are moot because they miss – or perhaps assume away – the essential point, which is to arrive at a set of goals of education which then condition the content and process of provision. It is necessary to know the aims of education before deciding on its process and content. In debates on the aims or goals of education, an analytical distinction is often made between 'liberal' and 'instrumentalist' approaches to education. The

dichotomy rests on a further distinction between 'intrinsically' and 'instrumentally' beneficial education. This paper now goes on to examine arguments for pluralist economics based on this distinction.

4. 'Liberal' aims of education

This type of education is also referred to as 'intrinsic education' in the literature, because it suggests that education is intrinsically valuable *per se*. For Bridges (1992) its central feature is ''to equip people to make their own free, autonomous choices about the life they will lead'' (p. 92) which implies:

> i) an ability to treat *critically* and of course also informedly, ideas and beliefs put forward by other people;
> ii) an awareness of the wider alternatives … available upon which one may *exercise choice*;
> iii) a level of *personal independence or autonomy* which gives one the will, courage or confidence to act on one's own beliefs.

(Bridges, 1992, p. 92, emphasis added)

These three goals can be more succinctly classified as critical and analytical (evaluative) thinking; comparative thinking; and intellectual open-mindedness. They collectively aim at the achievement of intellectual capacities, i.e. at the process of thinking within the individual. These aims mean that curricular content is only relevant in achieving outcomes that are (thought) processual – and content should be assessed according to its ability to achieve these outcomes; and 'facts' and 'knowledge' are de-emphasised.

We would argue that a pluralist approach would achieve these intrinsic aims *better* than a monist one. Analytical thought, arguably, can be achieved within a single approach. Indeed, mainstream economics, as it is taught, arguably scores highly on analytical thought. Its emphasis on modelling, mathematics and precision, diagrammatic expression, and concepts such as opportunity cost all help students develop analytical capacity. However, non-mainstream approaches have similar qualities. For example, Keynesian economics also develops analytical thinking, through the above techniques, and through a type of 'human logic' (see Dow, 2004). Austrian economics forces students to engage with non-equilibrium thinking. Thus, a pluralist approach may be beneficial in terms of analytical approaches in that it furnishes students with a *wider set* of analytical capacities. This may help them think flexibly in a complex world and equip them with the tools to apply different approaches to different situations.

Similar arguments can be made about critical thinking. Critical thinking has been considered explicitly in economics pedagogy by, for example, Earl (2002), Feiner (2002) and Guerrien (2002, 2009). Here, critical thinking means simply that the student interrogates the material (and perhaps the world and their perception of it). Clearly, any perspective may be taught critically in this way and any perspective could be the vehicle for critical thinking. Indeed, in the primary research reported below, on many occasions, students respond that by studying economics they have learned to think differently, and they have learned to question prior beliefs. The key to achieving critical thinking would appear to be that students are exposed to a critical attitude on the part of their lecturers (by not merely using critical thinking exercises). Criticality could clearly be achieved within any programme, but a pluralist approach may achieve critical thinking more effectively than a monist approach. This may occur in several ways. Given that all heterodox theories are partly critiques of the mainstream, they are inherently (albeit not exhaustively) critical. Critique can of course be either internal (in the sense that evidence might all be from one theoretical perspective, using the same methodology, etc.) or external. The process of critique we have been discussing above has been primarily internal. However, of course external critique is available in economics through a variety of channels. For instance, one can confront students with theory or evidence which is distinctly critical of one view or another. Indeed, in a module delivered by two of the authors, students are explicitly required to criticise articles they read and confront them with evidence which either supports or undermines the paper's view to some degree. Student evaluations

on this module comment that it involves deep thought and analysis, and encourages them to think independently and make decisions.

The point of pluralism is to make these processes more effective. It is to emphasise the role of external critique. It is to propose that students are forced to consider and perhaps reconcile diametrically opposed views. Thus, an approach of teaching multiple perspectives is desirable, because it increases the likelihood that the critical faculty of the student will be augmented. A critical curriculum also insures against the possibility that a teacher eschews or does not promote effectively internal critique.

At the same time, by definition, by studying different views on the same topic, the students' comparative thinking faculty would be exercised effectively. Thus, it can be argued that analytical, critical and comparative thinking can be achieved more effectively in a pluralist rather than monist curriculum. Similar arguments may be made in terms of instrumental aims and outcomes.

5. Instrumental(ist) aims

The pedagogical case for pluralism can also be made in terms of so-called instrumental benefits. Instrumental benefits are those concrete, identifiable skills, such as the ability to solve certain types of problem, know formulae or techniques, remember and perhaps apply theory, or possess 'knowledge' of a topic. In general, instrumental benefits involve the achievement of specific narrow learning outcomes. Clearly, all education will involve instrumental outcomes, even if they are not intended or explicitly stated. In order to engage in critical evaluation of a theory, it is necessary to know and understand it: both are instrumental outcomes. Learning basic arithmetic is obviously useful in many facets of life but it is also an aim of education that students are able to do arithmetic. A student can therefore achieve intrinsic and instrumental benefits simultaneously.

An education which is *geared towards* such instrumental goals may be regarded as 'instrumental*ist*'. An example of instrumentalist education is one in which a student is indoctrinated into a particular view or behaviour. For Hobsbawm (1997) state education was begun with indoctrination in mind. More broadly, though, any educational process can be regarded as indoctrinatory if its content is delivered uncritically: contrary to the tenets of 'liberal' education discussed above.

How might pluralism serve the needs of instrumental/ist education? On one hand, there may be an obligation to provide knowledge which is useful. If the world is complex, then students need to learn multiple theories, all of which are fallible. If theory X can partially illuminate phenomena, it is useful to learning. Also, as Clarke and Mearman (2001) note, learning theory X may mean that theory Y is understood better. Both theories may provide policy proposals, seem highly relevant, inspire debate and argument, and link to other disciplines. Moreover, in open, complex environments, there may also be a need for multiple methods and techniques: pluralism equips students with different methodological approaches to problems and different tools with which to solve them (see Downward and Mearman, 2007). Further, by teaching in terms of debates, students learn to negotiate their way through difficult issues on which there are multiple perspectives. In short, they learn judgement. Thus, students become more creative, better problem-solvers, and, by dealing with multiple perspectives, which have different methodological and even ethical bases, students learn to better negotiate complex, difficult situations.

These are all crucial qualities for potential employees to have. All could make workers more productive. Such considerations are relevant in the light of an increasing focus on employability of graduates. O'Doherty *et al*. (2007)'s survey of employers of economics graduates employed as economists suggested that employers value communication, complex problem solving, the ability to think systemically and debating skills. All of these suggest a role for pluralism; most obviously debating skills. From the arguments presented above, these demands suggest a role for pluralist curricula. Ironically perhaps, then, students trained in pluralist curricula could be better educated in an instrumentalist

sense than those in a monist scheme. However, employability could conflict with intrinsic aims of education, depending on how the teaching is done (see Clarke and Mearman, 2004). If workers are trained to be compliant and robotic process followers, liberal goals will be confounded. At this point education becomes instrumental*ist*. Similarly if the focus of the instructor is on the learning of any theoretical concept rather than its interrogation, the liberal aims may be lost. A pluralist approach may again insure against this outcome.

6. Barriers to pluralism

Some may argue that the case for pluralism is obvious and that the interesting question is why pluralism has not been adopted. Here, briefly, one reason is discussed: the pedagogical arguments that students struggle with ambiguity. An approach in which students deal with debates or deal with multiple perspectives concurrently may have many benefits, but a danger is that students get too confused, nihilistic or disengaged. Earl (2009) shows that an instructor who tries to push their students too quickly in this way could come unstuck and lose them. A chief benefit of the pluralist approach is that it allows students to make up their own minds and to make tentative commitments to theoretical positions, whilst acknowledging that others exist and that all have merits and problems. However, as Earl notes, this way of thinking does not occur overnight, nor can students be dragged to that level. Most start off as what Earl (from Perry) calls 'dualistic', i.e. right and wrong, thinkers: one theory must be the whole truth, or it is useless.

The extent of contrast used may be of crucial importance. Indeed, this is an important question for justifying the inclusion of heterodox approaches generally. From a Piagetian pedagogical perspective, a strong contrast is necessary: dissonance is necessary and the development of the student is shown in its resolution or *equilibration* (see Smith, 2004). In that regard, strong contrasts – even if between caricatured positions – may be desirable. Where parallel perspectives also involve different ethical or political bases – or indeed any other example of interdisciplinary content – this effect could be magnified. However, in Vygotskyian pedagogy, also, it is necessary that students are not exposed to concepts too far outside their *zone of proximal development*, and that therefore trying to understand the contrast between, say, mainstream and radical viewpoints may be too difficult. A mainstream teacher could argue that comparative thinking and debate could be achieved within a mainstream curriculum. For instance, there may be a debate over the specification of a production function, or the value of its parameter(s). It might be that all of the benefits of the comparative approach could be achieved and the potential confusion could be avoided.

It is essential then that the tutor employing pluralism is careful in their delivery. Students should be encouraged to think comparatively and critically, and to do so early in their study. It is imperative that the lecturer communicates to the students early on and repeatedly what they are trying to do (Earl, 2009). The design of assessment may be important here. For example, essays of increasing length and significance in terms of marks can ease students into the habit of thinking critically and openly (Mearman, 2007). A stress on the need to make an argument and develop a position can be similarly beneficial. Questions in the style of 'compare and contrast' may be particularly beneficial. One student evaluation suggested that a class debate before writing the essay would be useful to clarify the issues pertinent to it. Earl further suggests that tutors need to spend more time offering feedback on exactly how the argument could have been improved. Through detailed feedback, it may be possible to build confidence in students, enabling them to take positions on issues. Crucially, students must feel able to reach the conclusion that if they find a theory or model counter-intuitive or contradictory, that might be because the theory is indeed so, and that their confusion is not merely their own fault. Mearman (2007) discusses these issues further. As a final point, it is perhaps worth noting that most introductory macroeconomics courses are taught in terms of debates (and history), which confounds the view that students early in their economics study cannot cope with ambiguity.

7. Some empirical evidence

Barone (1991) holds that students also exposed to heterodox material at Dickinson College developed from dualistic thinkers, displayed greater understanding, were more effective critical thinkers, displayed improved judgement and confidence and were better performers in policy debates. However, besides that there is little evidence on the efficacy of teaching pluralist economics. Most of the small amount of the empirical work on teaching (even if interest in the subject is higher) focuses on the effective teaching of mainstream economics. Work that has been done by heterodox economists tends to be in making the case for teaching heterodoxy and developing materials which can be used. Most of the heterodox work focuses on content: they argue that economics student numbers fell in the 1980s and stayed low (cf. Salemi and Siegfried, 1999; and Siegfried, 2008) because *economics* is problematic (see Knoedler and Underwood, 2003). This work crowds out empirical work even further. Systematic empirical study of the effectiveness of pluralist approaches is required. We are encouraged by recent developments in this area, for instance as discussed in Garnett and Mearman (2011)'s retrospective on Barone's work.

Mearman *et al*. (2008) and Webber *et al.* (2009) analysed an online survey of students. Students were found to be heterogeneous; however, some more general findings are worth noting. The results suggest that confusion creates negative perception of economics, particularly when confusion is also positively associated with frustration, and negatively associated with the perception of economics as helping people make better decisions. However, the survey analysis to date does not tell us the source of the confusion; nor does it tell us which type of students might be relatively more confused. The survey finished with two open-ended questions in which students were asked first to name three concepts from their current economics units which added most to their understanding of the real world and, second, to list topics which they would have liked to see covered but which were not, respectively. The word ‘debate’ did not occur frequently either as something students liked or wanted more of. However, there were some other indicators of a demand for pluralism, for instance in the references to specific schools of thought.

The remainder of the paper draws on new primary data, mainly from focus groups conducted during 2007 and 2008. Before discussing the focus group research, it is worth summarising some of the findings we might expect to see in that data. A pluralist approach should, it has been argued, produce greater engagement on the part of students, the development of intellectual capacities and practical skills, greater capacity for judgement, greater understanding, and make students more employable, but with the danger that it generates confusion. We shall see that all of these are suggested to some extent in the focus group data.

8. Focus groups

Focus groups are an established means of collecting qualitative data (see Flick, 2006, Ch. 15; Table 16.1). They are a group discussion led by a moderator on a topic of interest. They mimic social situations, albeit somewhat artificially, and provide insights into beliefs, attitudes and understanding. As with most qualitative data, they allow participants to speak in their own language, and they allow surprising findings to emerge. They are not associated with testing hypotheses, although the researcher typically enters them with some key research questions (see previous paragraph) in mind. In each focus group, the question put to students was ‘how effective is economics in creating understanding of real world issues?’ This question is general enough to generate a number of strands of discussion, which indeed happened.

The process of organising the focus groups yielded variety in their composition. Two of the universities were middle-to-low ranking, and the other three were middle-to-high ranking (according to the THES). The sample could thus have benefited from the inclusion of a high-ranking and a low-ranking institution. One of the universities was in Scotland, the others in England. One of the universities

teaches mainly by distance learning to mature students. In terms of individual group members, there is considerable variety in terms of gender, nationality, age, main occupation and course of study. The average age of the participants was 27 (likely higher than average); 12 out of 23 (52.2%) were male (compared to a UK average of 65.1% (source: HESA, 2006/7)); 15 of 23 (65.2%) were full-time students (compared to a UK average of 88.6% (source: HESA, 2006/7)); 14 of 21 (66.6%) were UK students (compared to a UK average of 61.6 per cent (source: HESA, 2006/7)); 7 of 23 (30.4 per cent) were studying economics and of the others (two unassigned), all but one was studying some combination of economics and another subject; 17 of 21 (81.0%) (two unassigned) were in their final stage of undergraduate study. The students were not asked about ethnicity.

One clear source of possible bias in the organisation of focus groups is that the students may have been unusually predisposed to heterodoxy or pluralism. Given that the initial known points of contact tended to be colleagues of the author(s) and thus sympathetic to heterodox thought, this could have led to sample selection bias. However, in each case, the local contact approached all students in their teaching group and took all the volunteers who came. Another way in which heterodox bias could creep in is that the students could have been taught heterodox material. Indeed, in all five cases, to varying degrees, that was the case. However, in four of the cases, all students were taught predominantly mainstream material. In one of those, though, the students had been selected from a history of thought/methodology group, which might bias them towards heterodox concerns. In the fifth case, the students were taught an explicitly pluralist programme. Thus, the focus group members were almost certainly not representative in terms of their exposure to pluralist curricula. However, it is not inevitable that the groups would be in favour of pluralism: they might well be very confused by it, as discussed above.

Focus groups were recorded and transcribed, and analysed using *Nvivo*. A variety of findings emerged from them. Some concern and expression was noted about the emphasis in teaching on mathematics; however, often students recognised the importance and usefulness of mathematics as a part of economics. They also recognised the importance of evidence of other types, including data of different types. Students also spoke approvingly about using different types of logic for different situations. Included in that category was a recognition of the value of looking at problems from different perspectives: the vast majority of participants expressed a preference for studying a range of perspectives and understanding the debates which occur. Students also noted that in studying economics, they drew helpfully on other disciplines, namely ethics, politics, philosophy, psychology, history, ecology and sociology. This highlights the possible benefits of another form of pluralism: interdisciplinarity.

To sum up the case for pluralism and debate, we can examine one quotation from the focus groups, which expresses several benefits of the pluralist approach:

> '…I found it fascinating studying it …we do have a bit of a difficulty [in that] we've been presented with all these different [views] … But I think it's necessary at this stage to be presented with the alternative views, because we're not studying neoclassical economics, we're studying economics. So I think it is quite broad and I do actually enjoy the slightly conflicting way that some of it's being taught. But I do think than rather than saying this is... "Read this and it will tell you how that works." It's not doing that, it's saying "Read this and it may give you an insight, give you an understanding of some processes."… But you have to bring a lot to it as well, it's not like studying maths where you can just learn the answers, you have to bring a lot to it yourself, I think, and it is a kind of... it is giving the tools to be able to do things with them rather than giving you the answers…'

This is a rich quotation because it conveys that a pluralist approach creates engagement ('I found it fascinating studying it'); the student is happy with ambiguity ('But I think it's necessary at this stage to be presented with the alternative views, because we're not studying neoclassical economics, we're studying economics'); the student recognises the partial nature of knowledge ('Read this and it may give you an insight, give you an understanding of some processes'); the approach stimulates active learning ('You have to bring a lot to it yourself'); and the approach provides ways of thinking about economics and the economy rather than imposing a view ('It is giving the tools to be able to do things with them rather than giving you the answers'). Here it is evident that both intrinsic and instrumental benefits of education are achieved.

One reason why pluralism was considered useful was that it encouraged several cognitive abilities to develop in students. For instance, the use of judgement was recognised as important: and this flowed from a recognition – and, crucially, acceptance – of ambiguity and uncertainty. In terms of Earl's (2009) framework, these students were no longer dualistic thinkers and were cognitively more developed. Participants expressed the belief that studying economics had made them more questioning, critical and able to argue: and that this was independent of their overall maturity and experience. That suggests that students want to be equipped to make better, more informed decisions. Such skills could be valuable *per se*, but also were considered by some participants as desirable for employability. Such concerns also relate to the ability of graduates to make decisions in a business context. Also, the desire to be able to make better decisions relates to participants' preference for and interests in policy; and in turn an expressed preference for Macroeconomics over Microeconomics. Both of those preferences were also found in the survey results. Some illustrative examples of quotations help illuminate this discussion.

Several students emphasised the role of judgement:

> '…like when you look at politics in this country you can go to the right or the left or you can try to stick in the middle and form your own opinion based on the information you're given'.

This quotation also suggests that students felt their ability to debate or argue a case had improved. Other students echoed that sentiment:

> 'A lot of my friends just all make comments and I think "Well, you can't just say that, without, you know, without justifying it"'.

A comment from student evaluations supports that view:

> 'I like a mixture [of views] so you can get a feeling from both sides of an argument.'

Another student argued:

> 'I think with economics there's a lot more critical, like it's not just straight down the line, … from day one whatever you do you know that there's an opposing view, I think everything you learn, you know that someone's going to contradict it and I think that, because you can't look at something at face value I think it's taught me to look at something… not leave it but to look around it before I make my own decision, my own judgement, whereas I don't think I would get that from another subject like, I studied maths and maths is right or wrong… economics is just, grey, the whole thing is grey and you can argue whatever you want really as long as you've got something to back it up.'

This quotation reinforces the contention that pluralism and debate develop judgement, but also criticality, and that the student is happy with ambiguity. However, the student is clear that baseless opinion is not adequate; thus a relativism in which all opinions have equal value irrespective of their logic or evidential base – one potential problem of pluralism – is avoided.

Several students support the claim that pluralism, i.e. having access to multiple viewpoints, increases understanding:

> 'But I think you could benefit a lot, or you do benefit a lot from looking at a situation and saying, "Okay, how would this school of thought solve it and what's this school of thought view's on it" and like that analysing it and it makes you understand the different schools better by actually seeing that in reality if you applied that policy what would actually happen, what would be the consequences and yeah, what could be an alternative policy instead of just saying, solve this problem. It's a different way of looking at it in that sense.'

The quotation suggests that this student feels that multiple theories allow them to see different aspects of a problem and understand it better and that problem-solving may improve. Significantly also, the quotation suggests students are happy with the ambiguity of having multiple theories and no single correct answer. They also seemed happy with the fallibility of theories. This is an important issue because of the common claim that students find multiple perspectives too confusing. Another comment from a student evaluation on an introductory module about an assignment which explicitly demanded a comparative approach reinforces this view: 'I found the paper difficult to write, but yes I feel I was able to express my opinions'. Evidence from student evaluations collected by one of the authors reinforces the view that a pluralist curriculum may help argumentative and writing skills: 'I learned how to write an argumentative paper'; 'I learned to form more concrete opinions and argue them.'

In addition to the benefits outlined above, students also suggested that teaching via debates was more interesting, and engaging. As one student put it: 'Cos in macro we've seen what Keynesians think, or what post-Keynesians think ... and then we make up our mind easily because we see contradiction and then we go like, oh, okay, well actually this one's kind of right, that's kind of wrong ... it probably makes us think easier than the micro last year.' One student evaluation commented that the pluralist module was: '[o]ne of the most interesting modules of the year, due to the requirement to think about economics in different ways'. Another said: 'I think this [debate] is the most interesting way to learn. It lets you see both sides.' These comments explicitly refer to the benefit of pluralism in terms of the increased engagement the students felt.

Of course, not all students are as happy with the approach as those above; there is some evidence that debate confuses students: 'But, [there] appear to be so many schools of thought within economics and different ways of interpreting the same thing that I just find it difficult to say that economics as a body can explain something because two eminently respectable economists may explain something in completely different ways.' This comment was expressing frustration but could also be interpreted as being reconciled to uncertainty. More clear is the following comment: 'you're learning about one certain view and you kind of think, okay, I've understood that and then another economist comes in and tells you something completely different and I kind of find hard to follow all of them'. 'The focus groups echo this sentiment. One source of such confusion might be age, as younger students may be more dualist than older ones; however, this is not the case: although mature students often seemed happy with ambiguity, so too did standard-age students. Some student evaluations also expressed confusion: 'I think sometimes there is too much confusion on which is which, but the stress that is emphasised on the difference between them feels unnecessary. I think that we should [understand] that there is a difference of opinion, but not [in] everything.' Other students commented that at first the comparative approach was difficult, but that they saw the benefits eventually. These comments illustrated that students may need assistance in adapting to a pluralist approach. Thus data from focus groups suggest that for students the positive benefits of debate and engagement outweigh the negatives of confusion. This result appears to be robust for gender and age. Students explicitly spoke of the benefits of learning pluralistically in terms of judgement, analysis and criticality. These sentiments are supported in student evaluations collected by the authors.

9. Conclusions

This paper has investigated the question of pluralism within economics education. It has presented specific pedagogical arguments based on a liberal educational philosophy – which stresses the intrinsic benefit of education – for pluralism: specifically, the intellectual benefits of education may be achieved more easily through a pluralist approach. A pluralist approach means, in this context, that more than one economic perspective is considered, and that proponents of these perspectives are forced to engage in comparison. This is best done if the teacher is open to a pluralist approach; but the nature of the approach insures against excessive bias by the teacher towards or against any specific approach. However, there may also be instrumental benefits of a pluralist approach: students may understand topics better, and be able to cope with decision making and evaluation in complex environments. These benefits may have a pay-off in terms of student employability.

The theoretical arguments are supported here by drawing on new primary evidence from focus groups in UK universities and various student evaluations. The evidence is exploratory and its conclusions are tentative. However, it suggests that students enjoy debate and see the benefits of both in terms of their personal intellectual development. However, there is also evidence of some confusion amongst students resulting from being confronted with multiple perspectives. That highlights the theory, in Earl (2009), that teachers aiming to teach a parallel perspectives approach must do so carefully. One way to do so is via assessment strategies.

References

Barone, C. (1991). 'Contending Perspectives: Curricular Reform in Economics', *Journal of Economic Education*, vol. 22, pp. 15–26.

Bloom, B., Krathwohl, D. and Masia, B. (1964). *Taxonomy of Educational Objectives, Book 2, Affective Domain*, New York: Longman.

Bridges D. (1992). 'Enterprise and Liberal Education', *Journal of Philosophy of Education,* vol. 26(1), pp. 91–8.

Budzinski, O. (2008). 'Monoculture versus Diversity in Competition Economics', *Cambridge Journal of Economics*, vol. 32(2), pp.: 295–324.

Clarke, P. and Mearman, A. (2001). 'Heterodoxy, Educational Aims and the Design of Economics Programmes', *Journal of Economic and Social Policy*, vol. 5(2), pp. 43–57.

Clarke, P. and Mearman, A.(2003). 'Why Marxist Economics should be taught but probably won't be!', *Capital and Class*, vol. 79, pp. 55–80.

Clarke, P. and Mearman, A. (2004). 'Comment on Christopher Winch's "The Economic Aims of Education", *Journal of Philosophy of Education*, vol. 38(2), pp. 249–56.

Dalen, H. van (2003). 'Pluralism in Economics: A Public Good or a Public Bad?', Tinbergen Institute Discussion Papers 03-034/1, Rotterdam: Tinbergen Institute.

Dow, S. (1996). *The Methodology of Macroeconomic Thought*, Cheltenham: Elgar.

Dow, S. (1997). 'Methodological Pluralism and Pluralism of Method', in (A. Salanti and E. Screpanti, eds) *Pluralism in Economics*, Cheltenham: Elgar.

Dow, S. (2004). 'Uncertainty and Monetary Policy', *Oxford Economic Papers*, vol. 56, 539–61.

Dow, S. (2008). 'Plurality in Orthodox and Heterodox Economics', *Journal of Philosophical Economics*, vol. 1(2), pp. 73–96.

Downward, P. and Mearman, A. (2007). 'Retroduction as Mixed-Methods triangulation in Economic Research: Reorienting Economics into Social Science', *Cambridge Journal of Economics*, vol. 31, pp. 77–99.

Earl, P. (2002). 'The Perils of Pluralism in Economics and how to Reduce them', *Post Autistic Economics Review*, vol. 11(January).

Earl, P. (2009). 'In the Economics Classroom', in (E. Fullbrook, ed.), *Pluralist Economics*, London: Zed Books.

Feiner, S. (2002). 'Towards a Post-Autistic Economics', *Post Autistic Economics Review*, vol. 12(March).

Flick, U. (2006). *An Introduction to Qualitative Research*, London: Sage.

Fry, N. and Love, N. (2007) 'Business tutors' perceptions of a virtual learning environment', Business, Management, Accounting and Finance (BMAF) Annual Conference, Aston University, 2–4 May.

Fullbrook, E. (ed.) (2003). *The Crisis in Economics: the Post-Autistic Economics movement – the first 600 days*, London: Routledge.

Fullbrook, E. (2009). 'Relevance of Pluralist Economics for the 21st Century', in J. Reardon (ed.) The Handbook of Pluralist Economics Education, London: Routledge.

Garnett, R. and Mearman, A. (2011) Contending Perspectives Twenty Years on: What have our students learned? International Journal of Pluralism and Economics Education, 2 (1): 2-18.

Guerrien, B. (2002). 'Is There Anything Worth Keeping in Standard Microeconomics?', Post Autistic Economics Review, vol. 12(March).

Guerrien, B. (2009). 'The Need for Critical Thinking', in (J. Reardon, ed.) *The Handbook of Pluralist Economics Education*, London: Routledge.

Hobsbawm E. (1997). *On History*, Frome: Weidenfeld and Nicolson.

Holcombe, R. (2008). 'Pluralism versus Heterodoxy in Economics and the Social Sciences', *Journal of Philosophical Economics*, 1(2), pp. 51–72.

Johnston, C., James, R., Lye, J. and McDonald, I. (2000). 'An evaluation of collaborative problem solving for learning economics', *Journal of Economic Education*, vol. 31(1), pp. 13–29.

Klamer, A. and Colander, D. (1987). 'The Making of an Economist', *Journal of Economic Perspectives*, Fall, 12, pp. 95–111.

Knoedler, J. and Underwood, D. (2003). 'Teaching the principles of economics: a proposal for a multiparadigmatic approach', *Journal of Economic Issues*, 37, pp. 697–725.

Laurillard, D. (2008). 'The Teacher as Action Researcher: Using Technology to Capture Pedagogic Form', *Studies in Higher Education*, vol. 33(2), pp. 139–54.

Lawson, T. (2003). *Reorienting Economics*, London: Routledge.

Mäki, U. (1997). 'The One World and the Many Theories', in (Andrea Salanti and Ernesto Screpanti, eds.), *Pluralism in Economics*, Cheltenham: Elgar.

Mearman, A. (2007). 'Teaching Heterodox Economic Concepts', *Handbook for Economics Lecturers*, Bristol: Economics Network of the Higher Education Academy.

Mearman, A. (2008). 'Pluralism and Heterodoxy: introduction to the Special Issue', *Journal of Philosophical Economics*, vol. 1(2), pp. 5–25.

Mearman, A., Wakeley, T. and Shoib, G. (2008). 'Economics and the real world: students' perceptions of economics and the role of heterodoxy in changing them', AHE conference, Cambridge, July.

O'Doherty, R., Street, D. and Webber, C. (2007). 'The Skills and Knowledge of the Graduate Economist', Economics Network.

Reardon, J. (ed) (2009) *The Handbook of Pluralist Economics Education*, London: Routledge.

Salemi, M. and Siegfried, J. (1999). 'The State of Economic Education', *American Economic Review*, vol. 79: 355–61.

Schultze, U. and Rennecker, J. (2007). 'Reframing Online Games', in (K. Crowston, S. Siber and E. Wynn, eds) *Virtuality and Virtualisation*, New York: Springer.

Siegfried, J. (2008). 'Trends in Undergraduate Economics Degrees', *Journal of Economic Education*, vol. 39(3), pp. 297–302.

Smith, L. (2004). 'Piaget's Model', in (U. Goswani, ed.) *Blackwell Handbook of Childhood Cognitive Development*, London: Blackwell.

D. Webber and A. Mearman (2012). 'Students' perceptions of economics: identifying demand for further study', Applied Economics, print version 44 (9): 1121-1132, published on iFirst 6 Feb 2011.

Gruene-Yanoff, T. (2009) 'Learning from Minimal Economic Models', *Erkenntnis*, vol. 70, pp. 81–99.

Author Biography

Andrew Mearman is Subject Leader for Economics at Bristol Business School. He has published extensively in journals such as Oxford Economic Papers, Cambridge Journal of Economics and the American Journal of Economics and Sociology, on teaching economics, economic methodology, heterodoxy in economics and recently on sustainability and economics.

Tim Wakeley has taught at the University of Bath, and Griffith University, Brisbane. He has pursued research in behavioural and evolutionary economics. Throughout his teaching career Tim has devoted his research work and teaching to the application of economic concepts to real world business issues. In

the summer of 2010 Tim left full-time academia and returned to the UK to pursue his lifelong ambition of self-employment. He is currently developing a subscriber-based online learning resource.

Don Webber is Professor of applied economics at the University of the West of England. Often adopting a quantitative approach, his main research interests lie at the intersection of regional and labour economics, and include drivers of spatial productivity patterns and 'softer' labour productivity enhancers, such as the quality of the work environment, cultural underpinnings of job satisfaction and choice under uncertainty.

Gamila Shoib is an Associate Fellow with the Information Systems and Management Group at the Warwick Business School, University of Warwick. Gamila obtained her Ph.D. in Information Systems from the University of Cambridge. She has worked at the Universities of Cambridge and Bath, and Griffith University, Australia. Her research has focused on the social and organisational aspects of Information Systems, Social Science approaches to research as well as learning and knowledge creation in organisations.

Contact details

Andrew Mearman
University of the West of England
Coldharbour Lane
Bristol
BS16 1QY
United Kingdom
Email: Andrew.Mearman@uwe.ac.uk

Schumpeter's 'Vision' and the Teaching of Principles of Economics to Resource Students*

Paul Dalziel

Abstract

Sixty years ago, Schumpeter's Presidential Address to the American Economic Association discussed the 'vision' underlying the research of individual economists. A similar concept can be applied to different groups of students studying economics. Resource students, obliged to take an introductory principles course designed primarily for commerce students, experienced significantly poorer outcomes than their commerce counterparts. Inspired by Schumpeter's concept, and reflecting the wider movement for problem-based learning, a new course motivated the resource students to engage with the subject by paying careful attention to their concerns and interests. The result was a measurable improvement in the class's relative performance.

JEL classification: A2, A22

1. Introduction

Students from diverse backgrounds come to tertiary studies with different learning needs and styles. In their influential study, Prosser and Trigwell (1999) identified that students whose prior education experiences have prepared them for 'deep learning' do better than students who have not been challenged to move beyond 'surface learning'. Students have different strengths in learning material presented in different ways, such as in visual diagrams, as written text, in spoken lectures or through practised examples (Fleming and Mills, 1992). There is evidence that students benefit from problem-based learning, especially when the problems are directly relevant to their interests (Zuber-Skerritt, 1993; Boud and Feletti, 1997; Bourner *et al.*, 2000; Forsythe, 2002; Savin-Baden and Major, 2004). The literature on economics education also provides a wide range of reflective thought on how to engage students with diverse learning styles; for example, Becker (2000, 2004), Becker and Watts (2001a, 2001b), Colander (2004), Denis (2009), Dynan and Cate (2009), Elzinga (2001), Flores and Savage (2007), Guest (2005), Guest and Vecchio (2003), Hawtrey (2007), Jensen and Owen (2003), Lage *et al*. (2000), Owen (2007) and Ziegert (2000).

This literature includes particular attention on the teaching of introductory economics courses to first-year tertiary students. The curriculum of these courses is very well-defined by a strong consensus in

* A poster on this topic was presented at a special session on Teaching Ideas and Projects at the American Economics Association meetings, San Francisco, 2–5 January 2009, and an earlier version of this paper was presented to a special session on The Education of an Economist at the 50th anniversary conference of the New Zealand Association of Economists, Wellington, 1–3 July 2009. The author is grateful for feedback received from participants on those occasions and also for the insightful comments from Gillis Maclean, Ross Guest and two anonymous referees for this journal.

textbooks, with the powerful conceptual model of competitive market equilibrium at its core (Reimann, 2004). Students are required to learn how to present this model (either graphically or algebraically), how to apply the model to analyse events affecting supply or demand, and how the model is extended to consider imperfect competition, instances of market failure and theories of macroeconomic management. The dominance of this conceptual core in the economics curriculum has led some economists to worry that 'students who come to us to "study economics" instead become experts in mathematical manipulations' and 'their views on economic issues are influenced by the way we teach, perhaps without them even realising it' (Rubinstein, 2006, p. C1; see also Frey *et al.*, 1993, Frank *et al.*, 1996, Frey and Meier, 2003, Cipriani *et al.*, 2009, and Dow, 2009). This concern has particular force when the first-year course on principles of economics is offered to students across the whole campus, both inside and outside the commerce faculty.

In particular, the standard economics curriculum can pose learning challenges for students enrolled in programmes such as resource or environmental management. Students who enter these programmes because they are concerned about negative impacts of market behaviour on environmental or social outcomes are likely to struggle to engage with a course that begins with several week of studying conceptual theories of the benefits of competitive markets. If the examples used to illustrate these theories favour business applications, this is likely to create a further barrier to learning by such students. This issue is described by Jill Caviglia-Harris, who reported that the students in her inter-disciplinary course on environmental perspectives, most of whom had not been exposed to economics previously, held perceptions of the subject that were not always well-founded (Caviglia-Harris 2003, p. 200):

> The first-day survey demonstrated that the majority of the students did not understand economics and its role in environmental policy. Some students stated that economics was the cause of the environmental problems of today and believed that this is what environmental economists studied. For example, one student said that, 'Economists convert nature into money and are one of the factors in the destruction of the environment.'

Caviglia-Harris responded by moving quickly through the core theory (and avoiding its more abstract features) to focus on applications using environment-based examples, in-class experiments and economic games. Thus she was able to help her students learn to recognise the link between prices, markets and environmental policy more effectively than would have been possible for this group of students in a typical principles course. Her article generalises from this experience to advocate a curriculum sequence of 'introductory economics material, economic theory specific to the field addressed, and applications related to the course theme' (Caviglia-Harris, 2003, p. 197).

This present paper argues for an approach that goes one step further, based on an award-winning innovation to improve learning outcomes for resource students enrolled in the introductory economics course at one of New Zealand's eight universities, Lincoln University.[1] Following a proposal made by Nicola Reimann (2004) in this journal, the extra step is not simply to *apply* economic theory to relevant examples, but to present the *theory* itself within a relevant problem-based context. In the Lincoln University example, the problem was explicitly chosen to fit the 'vision' of resource students for their role in addressing environmental issues.

The word 'vision' is taken from the Presidential Address given 60 years ago by Joseph Schumpeter (1949) to the American Economics Association, in which he argued that different scientists come to their research with different visions of the world and their place in it (see McGraw, 2007, pp. 476–84).

[1]The author responsible for the changes described in this paper received an Award for Excellence in Teaching in 2007 (one of two Awards in this category conferred by Lincoln University that year) and was awarded a Good Practice Publication Grant in 2008 from Ako Aotearoa, the National Centre for Tertiary Teaching Excellence, to produce a resource on the innovations described in this paper; see Dalziel (2009, 2010).

This paper argues in the following section that this is also true of students coming to the study of economics. The paper then continues with a description of a course designed to engage the vision of resource students by using a relevant theme from environmental economics (the tragedy of the commons in ocean fisheries) as the *means* for introducing students to the core economic principles involved in competitive markets. The paper finishes with some evidence of the benefits of the change for members of the class, followed by a concluding summary and discussion.

2. Schumpeter's 'vision'

Schumpeter (1949, p. 350) observed that the process of scientific research begins with a pre-scientific act of perception and analysis, which recognises a set of related phenomena as having some meaning or relevance that justifies the researcher's interest. This initial mixture of perceptions and pre-scientific analysis Schumpeter called the researcher's 'vision'. His address provided three historical illustrations. Adam Smith's attitude to the land-owning and to capitalist classes 'was the attitude of the observer from outside', whose 'sympathies went wholly to the laborer', and who felt disgust 'at the inefficiency of the English bureaucracy and at the corruption of politicians' (idem, p. 353). Marx conceived history as 'the struggle between classes that are defined as the *haves* and the *havenots*, with exploitation of the one by the other, ever increasing wealth among ever fewer *haves* and ever increasing misery and degradation among the *havenots*, moving with inexorable necessity toward spectacular explosion' (idem, p. 354). Keynes perceived the modern economy as stagnationist, based on his vision of a 'mature and arteriosclerotic capitalist society that tries to save more than its declining opportunities for investment can absorb' (idem, p. 355). Schumpeter suggested that the underlying visions of economist researchers are the source of unavoidable ideological bias (1949, p. 352):

> [T]he original vision is ideology by nature and may contain any amount of delusions traceable to a man's [*sic*] social location, to the manner in which he wants to see himself or his class or group and the opponents of his own class or group. This should be extended even to peculiarities of his outlook that are related to his personal tastes and conditions and have no group connotation – there is even an ideology of the mathematical mind as well as an ideology of the mind that is allergic to mathematics.

Schumpeter was not worried by the biases in these different visions. To the contrary, he argued that ideology is bound to wither over time, not only as a result of changing social patterns but also because scientific 'fact finding and analysis … tend to destroy whatever will not stand their tests' (p. 359). Nevertheless, if Schumpeter was correct to say that economic researchers come to their task with pre-scientific visions, it is also true that *students* of economics arrive with different visions of the world and of the role of economists in understanding that world.[2] Certainly the distinction between a mathematical and a non-mathematical mind made by Schumpeter in his last sentence quoted above will be familiar to instructors of first-year principles courses.

Building on Schumpeter's concept, there are differences between the vision of a typical commerce student and that of a typical resource student in a first-year economics class. The typical commerce student can be assumed to be sympathetic about businesses operating in a market system and is likely to arrive in class with some curiosity about economic principles underlying the actions of consumers and producers. In contrast, the typical resource student may be suspicious about the impact of market-oriented business on the environment and may want to know how public authorities can intervene 'to save the planet'. These differences spill over into diverse visions for economics itself – commerce students are typically more favourably disposed to the study of economics as a way to advance their own ambitions (and indeed is more likely to have studied economics previously), while Caviglia-Harris

[2] This statement and those in the following paragraph are based on the author's 25 years' experience of teaching first-year economics in three different New Zealand universities. Some supporting evidence is offered in Table 1 below.

(2003, cited above) is not alone in finding that some resource students may hold the prior belief that economics is a cause of modern environmental problems.

Further evidence for this difference can be found in Table 1. Soper and Walstad (1983) offer an instrument for measuring the economic attitude sophistication (EAS) of students, comprised of 14 statements for which Soper and Walstad could demonstrate a strong consensus in the economics profession at the time of their study. The instrument asks students to indicate for each statement whether they strongly agree, agree, are undecided, disagree or strongly disagree. These responses are coded 1, 2, 3, 4 and 5 respectively for the statements where the economics consensus position involves strong disagreement, or 5, 4, 3, 2 and 1 for those where the consensus position is the opposite.

Table 1 Results of the economic attitude sophistication survey

	Commerce (101)		Resources (105)		Student t-statistic	
	Mean	SD	Mean	SD		
1. Government should control the price of petrol.	2.77	1.06	2.61	1.09	0.81	
2. Inflation is caused by greedy business and union leaders.	3.21	0.86	2.98	0.72	1.66	*
3. Business makes too much profit.	3.83	0.80	3.46	0.74	2.60	***
4. People should not have to pay taxes.	3.50	1.14	3.51	1.23	–0.08	
5. Free medical care should be provided for all New Zealanders.	2.28	1.12	2.24	1.14	0.15	
6. Banks should not charge interest on loans to customers.	3.53	1.18	3.15	1.04	1.92	*
7. Most people who don't jobs are too lazy to work.	2.81	1.19	3.17	1.12	–1.72	*
8. When a business gets big, it should be controlled by government.	4.10	0.86	3.78	0.65	2.42	**
9. New factories are not needed.	3.90	0.87	3.63	0.94	1.54	
10. People should not be told how to spend their money. (Reverse coded)	3.92	1.03	3.56	1.07	1.86	*
11. If everyone had more money, we'd all be better off.	3.59	0.97	3.66	0.82	–0.43	
12. Profits should not be regulated by government. *(Reverse coded)*	3.55	0.96	3.07	0.82	3.02	***
13. Most unemployed people are lazy.	2.93	1.12	3.15	1.17	–1.00	
14. When a strike occurs, government should step in and settle the dispute.	2.83	0.99	2.71	0.84	0.74	
Aggregate Totals	3.34	0.45	3.14	0.32	3.22	***

1 = strongly agree; 2 = agree; 3 = undecided; 4 = disagree; 5 = strongly disagree

* Significant at the 0.10 level; ** Significant at the 0.05 level; ** Significant at the 0.01 level.

This EAS instrument was administered to commerce degree students and to resource degree students in the first economics lectures at Lincoln University in 2008.[3] Eleven incomplete forms were excluded from the analysis, leaving 105 valid responses from commerce students, and 41 valid responses from resource students. Table 1 reports the mean score and standard deviation for each question, analysed by the two groups, as well as the Student t-statistic for testing the difference between two means. The bottom row of the table reports the aggregate results, which show that overall the commerce students were closer to the economics consensus position than the resource students. This difference was significant at the 1% level. Further, the individual questions with the strongest significance were the three questions on business: the resource students were significantly less likely to disagree that business makes too much profit or to object to the government regulating profits or controlling big business.

[3] Two small changes were made to the wording of the questions: 'gasoline' was replaced by 'petrol' in question 1; and 'Americans' was replaced by 'New Zealanders' in question 5. I also added a question asking if the student had previously studied economics; 61.0% of the commerce degree students reported they had, compared to only 34.1% of the resource degree students.

These results are consistent with the hypothesis that resource students at Lincoln University enter the principles of economics course with a vision different to that of commerce students. As noted in the introduction, this can produce learning barriers for the former in a course primarily designed for the latter. Given a passionate concern for issues such as global climate change, environmental pollution and resource depletion, resource students have less patience to attend several weeks of lectures on theories of market efficiency (following the standard textbook sequence) before the possibility of suboptimal social outcomes produced by externalities and common resources are addressed. This barrier is likely to be reinforced if resource students are predisposed to be suspicious of big business and the profit motive but the instructor's illustrations and case studies are generally drawn from commerce examples (Rubinstein, 2006). The following section therefore explains how a new course was designed to improve learning outcomes by addressing these barriers.

3. Teaching economics to resource students

Prior to 2005, Lincoln University students who were enrolled in degrees in environmental management, landscape architecture, resource studies, social science or tourism management were obliged to study a standard principles of economics course, ECON 101. These students had consistently poorer outcomes in ECON 101 than classmates from the University's commerce programmes in the same course. A nadir was reached in 2004, when their pass rate was 16 percentage points below that of the commerce students, accompanied by a 9 percentage point gap in mean marks. Consequently, a new course, ECON 105, was introduced to replace ECON 101 in the above five degrees. This new course was taught by the commerce faculty at the same level as ECON 101, but was intended to meet the specific learning needs of resource students studying economics for the first time.

ECON 105 shared many features with ECON 101. Assessment in both courses involved two term tests and a final exam with identical formats (although not the same questions). Both courses used the same textbooks. Students in each course had access to a dedicated website where they could download PowerPoint lecture notes (see Chen and Lin, 2008, for the benefits of this), practice tests and exams, and other ancillary learning resources. Instruction took place over 12 weeks, with each week involving three or four hours of lectures and one hour of collaborative learning (small groups working on problem sets and case studies with assistance available from the instructor or tutor). The major difference in learning design lay in the way that the market theory was presented in the two courses: ECON 101 continued to teach the model conceptually, reinforced with practical applications and exercises, while ECON 105 framed the initial presentation of the theory with a pressing real world problem familiar to students from their own environmental concerns.

After a week of similar introductory material, the ECON 101 class continued by studying the textbook derivation of the supply and demand market diagram. In contrast, ECON 105 began its second week with a documentary on the depletion of ocean fisheries: Empty Oceans, Empty Nets (distributed by Video Education Australasia, Bendigo, Australia, 2002). Leet and Houser (2003) explain the power of a shared film in providing context for a class that is learning economics. This was exactly the role performed by Empty Oceans, Empty Nets; New Zealand is a small island nation in the South Pacific, and so management of fisheries is an important issue for its resource students. Thus, the shared documentary motivated students to engage with the ECON 105 course by reassuring them that their concerns for the environment would be taken seriously and by promising that the market theory being studied would be directly relevant to those concerns.

The screening was followed by a discussion of the stylised timeline of Hilborn and Walters (2001, Figure 1.1, p. 7), which shows that sustainable growth in an unregulated fishery is typically followed by sequential periods of over-exploitation, collapse and slow recovery (the key message of the documentary). This became the problem that framed the class's study of the economic model of

competitive markets. In particular, a PowerPoint slide explained to the class that the module would develop an economic model to address four questions:

1. Why do fishing stocks get over-fished in an uncontrolled fishery?
2. Is this a case of market failure? Or will markets fix the problem themselves?
3. In either case, what is the best way for policymakers to help recovery?
4. What is best for people's well-being?

The class explored these questions by learning how to derive and apply the economic model of a competitive market. They studied how the quantity supplied and the quantity demanded respond to a change in market price, and how each curve can shift as a result of other influences. A biological measure of maximum sustainable catch was added to the diagram's horizontal axis, and students were reminded that the documentary had shown firms investing in specialist technologies to improve their ability to harvest a species (shifting the supply curve right) and marketing their product to increase consumption (shifting the demand curve right). The result is shown in Figure 1. When a species is first commercialised, market equilibrium at E_0 is initially sustainable (that is, to the left of the maximum sustainable catch). As the development of new technologies shifts the supply curve (S_0 to S_1), and as increased marketing shifts the demand curve (D_0 to D_1), both to the right, the market equilibrium shifts to E_1. The impact on market price is ambiguous, but the equilibrium quantity moves over time to the right, past the maximum sustainable catch, and the fishery may collapse.

Figure 1 Supply and demand in an ocean fishery

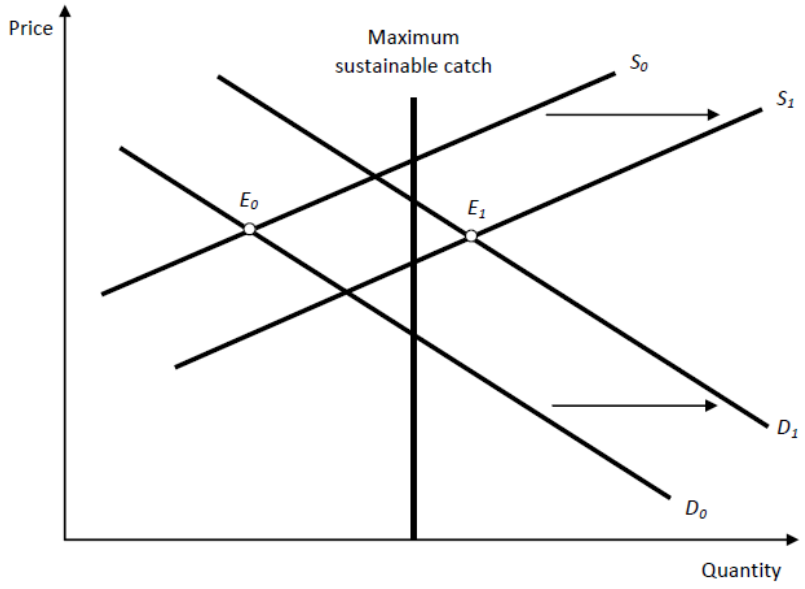


Figure 1 was used to illustrate policy responses shown in the documentary. If the authorities regulate boat owners to raise the cost of fishing, for example, or if they pay subsidies for reduced fishing effort, then the supply curve shifts to the left. If a public campaign by an environmental pressure group encourages a consumer-led boycott of species being fished unsustainably, the demand curve shifts to the left. The policy response adopted in New Zealand involves individual transferable quota permits being issued with a total allowable catch no greater than the maximum sustainable catch estimated by scientists (Lock and Leslie, 2007). This restricts supply to a vertical line representing the total allowable catch on the horizontal axis. This was a good example with which to finish the theoretical presentation,

since it involved students considering an important policy of using a market solution (establishing transferable property rights) to fix an important market failure (the tragedy of the commons).

Thus, by the end of their third week, the resource students in ECON 105 found they had come a long way. In contrast with the more abstract approach of ECON 101, they have been exposed to material they could see as directly relevant to their interests.[4] In particular, the problem-based approach was consistent with their vision as students wanting to address pressing environmental issues. A further advantage was that they had been exposed to core principles of economics in a way that was inherently dynamic (since it is the shifting curves that explain the ocean fishery depletion and recovery). Finally, for some resource students the discovery that markets might be used to fix an environmental problem challenged their negative vision of economics, exactly in line with Schumpeter's hopes for scientific progress.

Table 2 Comparison of ECON 105 and ECON 101 results, 2002–07

	2002		2003		2004		2005		2006		2007	
	RES	**COM**	**RES**	**COM**	**RES**	**COM**	**RES**	**COM**	**RES**	**COM**	**RES**	**COM**
Class size	65	298	86	469	90	561	53	425	58	311	46	274
Dropouts	9.2%	6.0%	5.8%	4.9%	14.4%	4.3%	3.8%	7.1%	6.9%	8.4%	4.3%	10.9%
Failed	27.7%	19.1%	27.9%	20.7%	38.9%	22.5%	24.5%	30.6%	22.4%	34.4%	28.3%	28.8%
Unrestricted	67.7%	76.2%	57.0%	74.8%	57.8%	71.5%	60.4%	64.9%	67.2%	62.1%	63.0%	67.9%
A+ grade	12.3%	17.4%	4.7%	8.7%	3.3%	7.8%	7.5%	4.2%	12.1%	10.3%	10.9%	11.7%
Mean mark	54.5	60.5	52.9	58.9	46.4	55.7	53.1	50.6	53.7	50.8	55.7	53.1
Std. dev.	24.3	21.6	20.3	19.0	24.3	19.2	19.0	20.1	21.5	22.8	19.3	24.9
Difference	–6.0		–6.0		–9.3		2.5		2.8		2.6	
t-statistic	–1.85		–2.54		–3.46		0.88		0.91		0.80	

Notes:
Class size is the number of students enrolled in the class as the end of the second week of the semester.
Dropouts is the percentage of the class who did not receive a grade because they did not sit the final exam.
Failed is the percentage of the class who did not gain credit for the course including those who dropped out.
Unrestricted is the percentage of the class who were qualified to proceed to the next level of economics studies.
A+ grade is the percentage of the class who received a mark of 80 or above (out of 100).
Mean mark and Std. dev. are the average mark (out of 100) and its standard deviation as shown in Figure 1.
Difference is the mean mark for the resource students minus the mean mark for the commerce students.
t-statistic indicates the significance of the difference between the two means, with a critical absolute value at the 5% level of 1.96.

The course then followed the standard textbook syllabus elaborating on the market model: elasticity of demand, consumer and producer surplus, competitive and monopolistic market structures, market externalities, public goods and macroeconomic policy. This was closer to the material in the ECON 101 class, although with a stronger emphasis on cost-benefit analysis and, following the recommendation of Caviglia-Harris (2003), the abstract material was continuously reinforced with environment-based examples, especially in the cooperative learning sessions each week. Thus students learned to practise how to apply economic principles to important environmental policy problems such as choosing endangered species to conserve, reducing congestion in an open-access national park, granting a licence to just one company in a local eco-tourism market, and analysing the impact of a carbon tax to meet commitments under the Kyoto Protocol.

[4] Bartlett (1995, p. 364) has emphasised the importance of applying economic theory to economic problems related to student interests in order to attract bright students to economics.

Table 2 provides an analysis of the change in the learning outcomes of the resource students relative to the commerce students for the three years before, and for the three years after, the introduction of ECON 105. The first measurable impact was on the percentage of resource student dropouts from the class, which almost halved from an average of 9.8% to 5.0% and moved from above the commerce dropout rate in each of the three earlier years to below the commerce dropout rate in the first three years of ECON 105. This contributed to a marked improvement in failure rates, which fell from a peak of 38.9% in 2004 to between 22.4% and 28.3% in the first three years of ECON 105. Both results are consistent with the new course achieving better engagement by its weaker students. At the other end of the scale, the percentage of A+ students increased, from an average of 6.8% to an average of 10.2%. These impacts are reflected in the mean marks of the two groups. In each of the last three years that the resource students sat ECON 101, their mean was six or more marks below that of the commerce students, and this gap was statistically significant in 2003 and 2004. In all of the first three years following the introduction of the new course, the resource group achieved a higher mean mark in ECON 105 than the commerce group achieved in ECON 101 (although this positive difference was not statistically significant).

4. Conclusion

Siegfried *et al*. (1991, p. 213) has counselled that 'instructors need to find the most effective blend of abstract and contextual material to make the powerful ideas of economics accessible to all students.' Bartlett (1996, p. 150) has advised that in an introductory economics course 'student diversity should also be discovered and acknowledged positively on the first day.' More recently, Brewer and Jozefowicz (2006, p. 202) observe that students 'may become critical when they fail to perceive a direct relevance between course content and either their present life or intended career path [and that this] frustration tends to be most evident in required introductory level courses primarily taken by noneconomics majors.' In this journal, Reimann (2004, p. 9) proposed as an observer from outside economics that 'the inductive, problem-first approach is interpreted as one possible way of aligning the teaching-learning environment with students, as it takes the importance of real-world examples and application of theory for student learning into account.'

This paper has provided further evidence in support of these observations. Resource students at a small New Zealand university, obliged to take an introductory principles course designed primarily for commerce students, experienced significantly poorer outcomes than their commerce counterparts. Following Reimann's proposal and inspired by Schumpeter's concept of 'vision', a new course was designed to motivate the resource students to engage with the subject by paying careful attention to their concerns and interests. Thus the new course provided a strongly relevant context (the tragedy of the commons in ocean fisheries) for the students' study of the theory of how competitive markets operate, and the theory was immediately applied to show how a market solution might be used to address an example of market failure. That pattern was followed throughout the course, with abstract material constantly presented within an applied, problem-based context and reinforced with environment-based examples. The result was a measurable improvement in the relative performance of the class (Table 2).

Some caveats are in order. The last empirical result was suggestive but not definitive. It is possible that the improvement came about from statistical chance, or was perhaps due to other associated impacts such as the smaller class sizes of the ECON 105 course or a higher than usual enthusiasm of a teacher and class involved in a deliberate pedagogical innovation (the Hawthorne effect; see Adair, 1984). It is possible that other heterogeneous features of the two groups (including their different levels of prior experience with economics; see footnote 3 above) may have played a role. In a future trial of this type, it would be useful to re-administer the Soper and Walstad (1983) survey to the two groups of students (commerce and non-commerce) at the end of their respective courses to test whether there had been any change in the significant differences found at the course start. Finally, the approach reported in this

paper of offering a separate principles course to a small group of students is unlikely to be viable at many institutions (and indeed proved not to be sustainable at Lincoln University itself). This leaves open the important question of how to address the different visions of economics and the role of economists in a single first-year course.

Nevertheless, this course was designed within the more general movement in tertiary education that advocates problem-based learning. The Lincoln University experience suggests that such an approach is particularly important for students whose vision for their future is as problem-solvers equipped to address pressing global issues. It provides support for Reimann's (2004) proposal that the problem-first or inductive approach is a good way to align the teaching-learning environment with such students.

References

Adair, J. G. (1984). 'The Hawthorne effect: A reconsideration of the methodological artefact', *Journal of Applied Psychology*, vol. 69, pp. 334–45.

Bartlett, R. L. (1995). 'Attracting "otherwise bright students" to economics 101', *American Economic Review*, vol. 85, pp. 362–66.

Bartlett, R. L. (1996). 'Discovering diversity in introductory economics', *Journal of Economic Perspectives*, vol. 10, pp. 141–53.

Becker, W. E. (2000). 'Teaching economics in the 21st century', *Journal of Economic Perspectives*, vol. 14, pp. 109–19.

Becker, W. E. (2004). 'Economics for a higher education', *International Review of Economics Education*, vol. 3(1), pp. 52–62.

Becker, W. E. and Watts, M. (2001a). 'Teaching economics at the start of the 21st century: Still chalk and talk', *American Economic Review*, vol. 91, pp. 446–51.

Becker, W. E. and Watts, M. (2001b). 'Teaching methods in U.S. undergraduate economics courses', *Journal of Economic Education*, vol. 32, pp. 269–79.

Boud, D. and G. Feletti, Eds. (1997). *The Challenge of Problem-Based Learning*, London: Kogan Page.

Bourner, T., Cooper, A., and France, L. (2000). 'Action learning across a university community', *Innovations in Education and Training International*, vol. 37, pp. 2–9.

Brewer, S. M. and Jozefowicz, J. J. (2006). 'Making economic principles personal: Student journals and reflection papers', *Journal of Economic Education*, vol. 37, pp. 202–16.

Caviglia-Harris, J. L. (2003). 'Introducing undergraduates to economics in an interdisciplinary setting', *Journal of Economic Education*, vol. 34, pp. 195–203.

Chen, J. and Lin, T.-F. (2008). 'Does downloading PowerPoint slides before the lecture lead to better student achievement?', *International Review of Economics Education*, vol. 7(2), pp. 9–18.

Cipriani, G. P., Lubian, D. and Zago, A. (2009). 'Natural born economists', *Journal of Economic Psychology*, vol. 30, pp. 455–68.

Colander, D. (2004). 'The art of teaching economics', *International Review of Economics Education*, vol. 3(1), pp. 63–76.

Dalziel, P. (2009). 'Schumpeter's "vision" and the teaching of principles of economics', Poster Session on Teaching Ideas and Projects, organised by the AEA Committee on Economic Education, Allied Social Science Associations meetings, San Francisco, 3–5 January.

Dalziel, P. (2010). 'Improving learning outcomes by seeing vision: Teaching economics to resource students', *Good Practice Publication Grants* e-book, Wellington: Ako Aotearoa, available at http://akoaotearoa.ac.nz/gppg-ebook.

Denis, A. (2009). 'Editorial: Pluralism in economics education', *International Review of Economics Education*, vol. 8(2), pp. 6–22.

Dow, S. (2009). 'History of thought and methodology in pluralist economics education', *International Review of Economics Education*, vol. 8(2), pp. 41–57.

Dynan, L. and Cate, T. (2009). 'The impact of writing assignments on student learning: Should writing assignments be structured or unstructured?', *International Review of Economics Education*, vol. 8(1), pp. 64–86.

Elzinga, K. G. (2001). 'Fifteen theses on classroom teaching', *Southern Economic Journal*, vol. 68, pp. 249–57.

Fleming, N. and Mills, C. (1992). 'Not another inventory, rather a catalyst for reflection', *To Improve the Academy*, vol. 11, pp. 137–55.

Flores, N. and Savage, S. J. (2007). 'Student demand for streaming lecture video: Empirical evidence from undergraduate economics classes', *International Review of Economics Education*, vol. 6(2), pp. 57–78.

Forsythe, F. (2002). 'Problem-based learning', in (Davies, P., ed.), *The Handbook for Economics Lecturers: Teaching*, Bristol: Economics LTSN, http://www.economicsnetwork.ac.uk/handbook/ (last accessed: 2 May 2010).

Frank, R. H., Gilovich, T. D. and Regan, D. T. (1996). 'Do economists make bad citizens?', *Journal of Economic Perspectives*, vol. 10, pp. 187–92.

Frey, B. S. and Meier, S. (2003). 'Are political economists selfish and indoctrinated? Evidence from a natural experiment', *Economic Inquiry*, vol. 41, pp. 448–62.

Frey, B. S., Pommerehene, W. W. and Gygi, B. (1993). 'Economics indoctrination or selection? Some empirical results', *Journal of Economic Education*, vol. 24, pp. 271–81.

Guest, R. (2005). 'Will flexible learning raise student achievement?', *Education Economics*, vol. 13, pp. 287–97.

Guest, R. and Vecchio, N. (2003). 'Are there learning spillovers in introductory macroeconomics?', *International Review of Economics Education*, vol. 1(1), pp. 36–60.

Hawtrey, K. (2007). 'Using experiential learning techniques', *Journal of Economic Education*, vol. 38, pp. 143–52.

Hilborn, R. and Walters, C. J. (2001). *Quantitative Fisheries Stock Assessment*, Boston: Kluwer Academic Publishers.

Jensen, E. J. and Owen, A. L. (2003). 'Appealing to good students in introductory economics', *Journal of Economic Education*, vol. 34, pp. 299–325.

Lage, M. J., Platt, G. J. and Treglia, M. (2000). 'Inverting the classroom: A gateway to creating an inclusive learning environment', *Journal of Economic Education*, vol. 31, pp. 30–43.

Leet, D. and Houser, S. (2003). 'Economics goes to Hollywood: Using classic films and documentaries to create an undergraduate economics course', *Journal of Economic Education*, vol. 34, pp. 326–32.

Lock, K. and Leslie, S. (2007). 'New Zealand's quota management system: A history of the first 20 years', Motu Working Paper 07-02, Wellington: Motu Economic and Public Policy Research, available at www.motu.org.nz/ (last accessed: 2 May 2010).

McGraw, T. K. (2007). *Prophet of Innovation: Joseph Schumpeter and Creative Destruction*, Cambridge MA: Belknap Press of Harvard University Press.

Owen, A. L. (2007). 'Integrating computer applications into economics electives', *International Review of Economics Education*, vol. 6(1), pp. 77–92.

Prosser, M. and K. Trigwell (1999). *Understanding Learning and Teaching: The Experience in Higher Education*. Birmingham: Open University Press for the Society for Research into Higher Education.

Reimann, N. (2004). 'First-year teaching-learning environments in economics', *International Review of Economics Education*, vol. 3(1), pp. 9–38.

Rubinstein, A. (2006). 'A sceptic's comment on the study of economics', *Economic Journal*, vol. 116, pp. C1–C9.

Savin-Baden, M. and Major, C. (2004). *Foundations of Problem-Based Learning*, Maindenhead: Open University Press and Society for Research into Higher Education.

Schumpeter, J. A. (1949). 'Science and ideology', *American Economic Review*, vol. 39, pp. 345–59.

Siegfried, J. J., Bartlett, R. L., Hansen, W. L., Kelly, A. C., McCloskey, D. N. and Tietenberg, T. (1991). 'The status and prospects of the economics major', *Journal of Economic Education*, vol. 22, pp. 197–224.

Soper, J. C. and Walstad, W. B. (1983). 'On measuring economic attitudes', *Journal of Economic Education*, vol. 14, pp. 4–17.

Ziegert, A. L. (2000). 'The role of personality temperament and student learning in principles of economics: Further evidence', *Journal of Economic Education*, vol. 31, pp. 307–22.

Zuber-Skerritt, O. (1993). 'Improving learning and teaching through action learning and action research', *Higher Education Research and Development*, vol. 12, pp. 45–58.

Author Biography

Professor Paul Dalziel has twenty-five years experience of teaching entry-level economics classes. He received a Special Achievement Award for Excellence in Teaching from Lincoln University in 2006 and a Green Award for Excellence and Achievement in Teaching from the Lincoln University Students' Association in 2010. He was nominated by Lincoln University for New Zealand's Ako Aotearoa Tertiary Teaching Excellence Awards in 2007 and again in 2009.

Contact details

Paul Dalziel
Professor of Economics
AERU, Lincoln University
New Zealand
Tel: +64 3 325 2811
Email: dalzielp@lincoln.ac.nz

An Integrated Approach to Teaching Price Discrimination

Ann Marsden and Hugh Sibly

Abstract

Textbooks present the three 'degrees' of price discrimination as a sequence of independent pricing methods and consequently provide inadequate insight as to when a firm might adopt a particular pricing strategy. The paper describes a taxonomy of the various mechanisms of price discrimination, which can be used to teach monopolistic price discrimination in an integrated way. The pricing strategy adopted by firms is based on (i) the information on consumer demand available to it and
(ii) whether the firm has the ability to conduct non-linear pricing. The paper proposes a method for ranking profit and efficiency levels under different price discrimination strategies. The proposed taxonomy is compared to the existing textbook approach.

JEL classification: A20, A22, L11, L12

1. Introduction

Students observe price discrimination in their daily life. They could, for example, purchase bus tickets at a discounted 'student' rate. Such a purchase is contingent on providing evidence that they are a student (usually by producing their student cards). They will often also observe that both adult and student travellers must choose between various size bundles of bus tickets. Larger bundles mean cheaper prices per ticket. Although this is one transaction for the student they are in fact facing two different price discrimination mechanisms. Textbooks do not provide an integrated explanation for this pricing behaviour of firms. Rather they often still provide independent explanations of the three types of price discrimination (first, second and third degree price discrimination) first proposed by Pigou (1920). The technical treatments of each type of price discrimination are usually incompatible with one another, since first degree considers linear and non-linear prices, second degree non-linear prices, and third degree linear prices. Students are left to ponder how, if at all, these types of price discrimination might be related.

In this paper we propose an integrated method to teaching price discrimination. We present a taxonomy based on two characteristics of the market the firm faces: (i) whether there exists a costless exogenous signal of customer type (such as a student card) and (ii) whether it is possible for the firm to conduct non-linear pricing. This taxonomy allows students to identify the circumstances under which the firm would undertake a particular type of price discrimination.

Recent literature takes the view that the price discrimination strategy adopted by the firm is associated with the information on consumer type (demand) available to it. Varian (1987) provides an earlier overview of this literature, in particular of the self-selection mechanisms implicit in second degree price discrimination. More recently Stole (2007) provides an overview of price discrimination in the context of competitive markets. The approach in this recent literature makes the distinction between direct and indirect price discrimination. Direct price discrimination is based on observable consumer characteristics, for example whether a customer has a student card. Indirect price discrimination is when firms cannot observe consumer characteristics and so must use some pricing strategies to identify

the customer type. This corresponds to the bus company offering students various bundle sizes to choose from.

The second market characteristic in our taxonomy relates to the ability of the firm to use non-linear pricing. There may be a variety of reasons why firms use linear pricing (e.g. government regulation, high cost of bundling), but the easiest one to motivate students in a classroom is that the goods can be resold by customers. In this paper we draw the distinction between markets in which intra-type arbitrage is possible or those in which it is not possible (e.g. whether students can costlessly re-sell a bus ticket they bought to another student). The firm utilises linear pricing in the former case and non-linear pricing in the latter.

There are a number of papers related to the teaching of price discrimination. Carroll and Coates (1999) explore the teaching of price discrimination and consider the efficiency and profit gains under first, second and third degree price discrimination. They note that assessing whether a two-part tariff enables the firm to capture more consumer surplus than if it used third degree price discrimination is not generally determinable. A student considering the two types of price discrimination would therefore be unable to link the two conceptually from the analysis. Using the approach proposed in this paper enables the student to consider second and third degree price discrimination as part of a conceptually unified approach to price discrimination. In the comment and addendum to the Carroll and Coates (1999) paper Jeitschko (2001) has concerns about what is (or should) be taken away from the classroom after a discussion of price discrimination, especially regarding implications of price discrimination on economic efficiency. The approach we take allows the efficiency implications to be assessed for a particular price discrimination strategy. This means that efficiency can be compared across strategies.

Gotlibovski and Kahana (2009) note that the typical intermediate microeconomics textbook discusses only first and third degree price discrimination, because the presentation of second degree price discrimination requires the use of a more sophisticated mathematical technique. They note that Varian (2006) uses a simple diagram to explain second degree price discrimination. Gotlibovski and Kahana (2009) build on this diagram to explain differences between price quantity packages and two-part pricing. This paper uses a similar approach in proposing an information based analysis of direct and indirect price discrimination.

Our paper proceeds as follows. First we consider the existing textbook approach and how our approach differs from this approach. Then, we outline our assumptions regarding consumer demand. We then consider markets in which intra-type arbitrage is not possible, going on to identify the optimal non-linear pricing strategy given the market's information structure. Then the profitability and efficiency of each pricing strategy is ranked. This ranking is related to the information structure. We consider markets in which intra-type arbitrage is possible and consequently linear pricing must be used by firms. We relate the market outcome under linear pricing to that under non-linear pricing. Finally we bring together the analysis of the previous sections to present an integrated approach to teaching price discrimination, before concluding the paper.

2. Existing textbook approaches

A survey of five principles of microeconomics texts, Hubbard *et al.* (2009), McTaggart *et al.* (2010), Perloff (2009), Swann and McEachern (2003) and Taylor and Frost (2006), demonstrates that existing textbooks each use a mix of linear and non-linear pricing to demonstrate the three types of price discrimination with, in most cases, no attempt to link the different types of price discrimination. In the bus ticket example, the textbook explanation for the student discount is presented separately from the quantity discount for the bundle of tickets. The former is an example of third degree price discrimination and the latter an example of second degree price discrimination. Separating the analysis

in texts offers little opportunity for the student to see these strategies as linked. The existing textbook treatments consequently provide little insight as to when a firm might adopt third degree price discrimination as opposed to second degree price discrimination or when a firm might use a mix of both types of price discrimination. Observation of firms indicates that in many cases they do not use either third or second degree price discrimination but approach their pricing using a mix of the two types of price discrimination. This is the case for the bus company discussed above. Similarly cinemas offer both student discounts and discounts for quantity. Accommodation providers offer both corporate and leisure rates, as well as discounts for extended stays. The framework we propose using non-linear pricing strategies readily models this behaviour.

In the textbooks, first degree price discrimination may be discussed in terms of customers buying either one unit or more than one unit. Where customers buy only one unit first degree price discrimination is presented using a linear market demand curve with each buyer having a different marginal benefit along the demand curve. The firm maximises profit (and incidentally social surplus) by charging each customer a price which exactly matches the marginal benefit of that customer. Where customers purchase more than one unit, first degree price discrimination is presented using non-linear pricing. It is common to see an analysis involving two consumers whose demand curves for the good differ. The firm maximises profit (and also social surplus again) by charging a tariff to each customer type equal to total benefit for the efficient quantity for that customer type (which occurs where the type's marginal benefit equals marginal cost). We use non-linear pricing to show that first degree price discrimination is a form of direct price discrimination where the firm has full information about each of its customers' characteristics.

In the textbooks third degree price discrimination is usually discussed in the context of linear pricing using the traditional Pigouvian approach. Textbooks generally consider two groups of customers whose market demand curves have different elasticities at all price levels. The firm is able to identify the group using a signal which provides information about the characteristics of that group. The firm sets the price to each group at the point where the marginal revenue of the group is equal to marginal cost. Pricing is linear since the firm sets a uniform price for each group. Customers may purchase only one unit or more than one unit but pay the same price per unit irrespective of the amount they purchase. The textbook treatment assumes that the firm sets a linear price because they have insufficient information to further separate customers within the groups. This issue is explored further later. We show that third degree price discrimination is a form of direct price discrimination where firms have some information about customer characteristics of groups and can be approached using non-linear pricing. We then incorporate the traditional Pigouvian textbook approach into our analysis.

The modern textbook treatment of second-degree price discrimination as shown in Tirole (1988) and Carlton and Perloff (2004) differs from that identified by Pigou (1920). There is some confusion regarding Pigou's original definition of second-degree price discrimination. Many recent writers include self-selection via non-linear pricing as a form of second degree price discrimination. Stole (2007) notes that Pigou (1920) did not consider second degree price discrimination as a selection mechanism, but rather thought of it as an approximation of first degree using a step function below the consumer's demand curve. As such, Pigou regarded both first and second degree price discrimination as 'scarcely ever practicable' and 'of academic interest only'.

The modern treatment of second degree price discrimination, beginning with Spence (1977) and Maskin and Riley (1984), utilises modern advances in information economics to explicitly model the information asymmetry between a firm and its customers. Whereas first or third degree price discrimination is used when observable customer characteristics are common knowledge, second degree price discrimination is used when the customer characteristics are private information (known only to the customer him or herself). Non-linear pricing schedules can then be used to provide customers with an incentive to self-identify. The modern textbook treatment usually follows this approach and we also use the same

approach in this paper. These treatments generally consider two customer types, one of whose demand curve lies uniformly above the other. Non-linear prices are used to provide an incentive for customers to reveal their types. Examples of such non-linear pricing are abundant, so Pigou was incorrect in asserting that second-degree price discrimination is 'of academic interest only'. The student only has to observe prices per unit in the supermarket to see examples of non-linear prices. Many supermarket items are sold in different sized packages with the price per unit falling as the packet size increases. Similarly many coffee shops offer a free coffee after the buyer has purchased a minimum number of cups of coffee thus offering a lower price per unit for the purchaser of larger quantities. The bus ticket discount for purchases of larger quantities is also an example of second degree price discrimination.

Teaching and learning the three types of price discrimination using the mix of linear and non-linear pricing methods proposed in the (inconsistent) Pigouvian taxonomy can be confusing. Textbooks do not generally link the linear and non-linear analytical approaches and may in fact avoid using the non-linear analysis of second degree price discrimination as noted by Gotlibovski and Kahana (2009) because of the mathematical difficulty. In contrast, by using an approach that systematically modifies the information available to the firm regarding the distribution of customer demands, and using linear and non-linear pricing in a systematic way, the optimal (discriminatory) pricing strategy available to firms can be identified. Thus the incentives for price discrimination are clearly identified in an integrated manner that allows students to view the firm's profit level as a function of the optimal mix of direct and indirect price discrimination strategies. Using this approach also allows the (Pigovian) third degree price discrimination linear analysis to be examined in the context of non-linear pricing. In fact we show that a firm has an incentive to use non-linear pricing rather than linear pricing where possible.

3. Customer demand

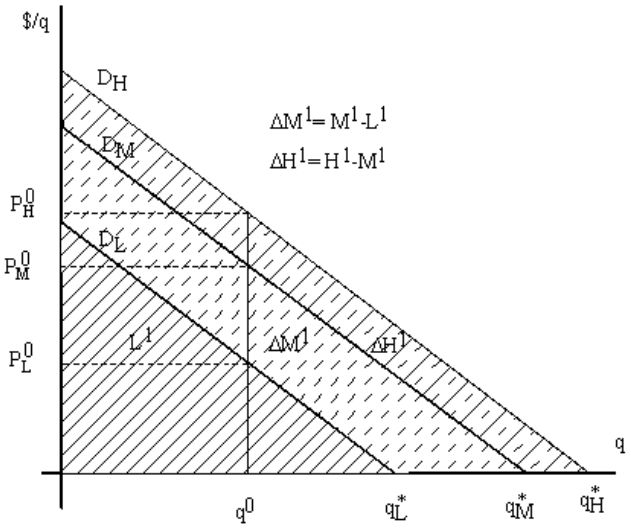
We start with the assumption that the firm, which is a monopoly, is aware of the distribution of customers' demand curves (and thus can calculate market demand), but it may not be able to costlessly associate a demand curve with a particular customer. It is also useful to restrict consideration to those cases in which demand curves of different customer types do not cross. This provides an unambiguous ranking of customers in terms of their willingness to pay, thus particular customer types can be identified as having a higher or lower demand than other customer types. This condition is known variously as 'uniform ordering', the 'sorting condition' the 'single crossing condition' or the 'Spence-Mirlees' condition' (Spence, 1977). The same demand curves can then be used throughout the analysis of direct and indirect price discrimination thereby providing students with a framework that is analytically self-contained.

We further restrict attention to the case in which there are three customer types. Discussions of price discrimination in the textbooks often consider only two customer types. However, to consider the examples of price discrimination discussed in the introduction, and to develop a conceptually encompassing information-based taxonomy, requires a minimum of three customer types. This approach is in the tradition set out in Goldman *et al.* (1984) and used by Dolan and Simon (1996). The analysis could be extended to more customer types but this yields little additional economic insight. Thus for ease of presentation we restrict consideration to three customer types.

Figure 1 illustrates the condition for three types of customers: Customers H (high demand), M (medium demand) and L (low demand). We will use the same demand curves throughout the analysis in the following sections. Note that under this condition type H customer's demand curve lies above type M customer's demand curve, which itself lies above type L customer's demand curve. Thus for any level of output, q_0, as shown in Figure 1, the marginal valuation of type H customers (P^0_H) is greater than the marginal valuation of type M customers (P^0_M) which is greater than the marginal valuation to type L customers (P^0_L). Consequently type H customers have a higher total valuation than type M customers, who have a higher total valuation than type L customers. We illustrate our analysis throughout this

section using the bus ticket example introduced earlier. Using this example an instructor could assume that high demand adult customers using the bus (e.g. commuters) are type H customers, lower demand adult customers (non-commuters) are type M customers and student customers the type L customers. We also provide a numerical example later and have annotated this example with the notation used in this section. This means that an instructor could use either the algebraic or numerical model (or both) to demonstrate this integrated approach to teaching price discrimination.

Figure 1 Demand curves satisfying uniform ordering and tariffs under first degree price discrimination



4. Non-linear pricing: no intra-type arbitrage

In this section it is assumed that it is not possible for customers of a given type to trade (i.e. resell) the firm's output amongst themselves.[1] In this case the firm can bundle its output and each must consume one of the bundles offered by the firm (or otherwise consume nothing of the good). The pricing strategy adopted by the firm depends on how readily it can identify customers as belonging to the different types. Below we consider the cases in which, (i) the firm can costlessly identify each customer's type, (ii) can costlessly identify only one type of customer, (iii) cannot identify any customer's type. We will assume for simplicity that marginal cost for the firm is equal to zero.

Pure direct non-linear pricing: all customer types costlessly identified

We will start by considering the case where a given customer's type is common knowledge, and thus firms can costlessly identify and separate the three customer types. This is a form of direct price discrimination and corresponds to Pigouvian first degree price discrimination in the textbooks. The firm can capture all the consumer surplus of each customer type by offering each customer a block tariff.

The optimal pricing structure in this case is shown in Figure 1. Type L customers are offered schedule $<q_L^*, L^1>$, which consists of a bundle of q_L^* units for tariff equal to L^1. This schedule leaves the consumer with zero consumer surplus, so the customer is indifferent between purchasing the bundle or not

[1] Inter-type trade is ruled out throughout this paper. An example of inter-type trade would be a student re-selling a bus ticket to a non-student for a profit. An example of intra-type arbitrage would be a student unbundling a bundle of 10 tickets and selling single tickets to other students for a profit. If intra-group arbitrage is possible but not inter-type arbitrage then the firm must offer a linear price to each group. If both types of arbitrage are possible then the firm must offer a common linear price to all groups.

purchasing it. For ease of analysis assume that the customer purchases the bundle when indifferent. Similarly type M customers are offered the schedule $<q_M^*, M^1>$ where $M^1=L^1+\Delta M^1$ and type H are offered the schedule $<q_H^*, H^1>$ where $H^1=M^1+\Delta H^1$. The number of type H customers is N_H, the number of type M customers is N_M and the number of type L customers is N_L. Profit, using the notation Π, is equal to:

$$\Pi^1 = N_L L^1 + N_M M^1 + N_H H^1 \quad (1)$$

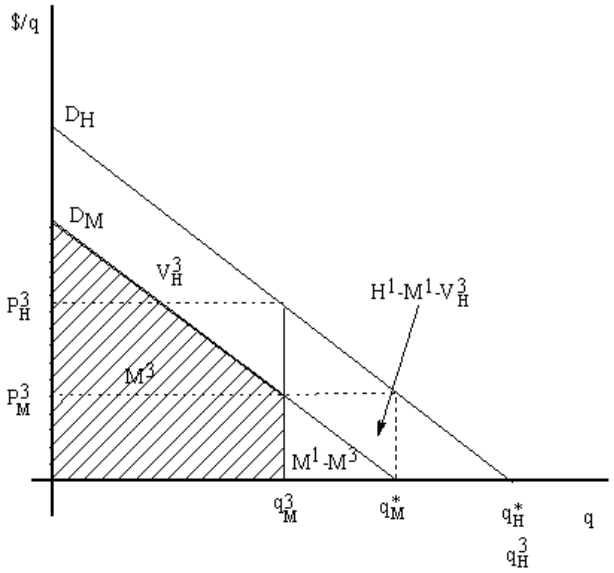
Note that each customer purchases the efficient quantity. Using the bus ticket example this means that both types of adult fare customers and the student fare customers each pay a price that exactly matches the total benefit they receive from purchasing their bundle of tickets. The adult commuter customers purchase the largest bundle, the adult non-commuter customers the next largest bundle and the student customers the smallest bundle.

Partial direct non-linear pricing: one customer type costlessly identified

Next assume the firm can costlessly identify (and thus separate) type L customers. However it cannot costlessly distinguish between type M customers and type H customers. The profit maximising pricing strategy requires the firm to separate customers according to the freely available information. In particular each identifiable group of customers potentially contains within it customers with heterogenous demands, e.g. type M and type H.

The firm has to set a schedule that ensures type M and type H customers self-select the appropriate bundle. Figure 2 shows how the non-linear pricing can be used by the firm to profitably separate the type M and type H customers.

Figure 2 Partial direct non-linear pricing



The firm offers two schedules $<q_M^3, M^3>$ and $<q_H^3, H^3>$. The former schedule is directed at type M customers and the latter type H customers. It is profit maximising for the firm to extract the entire consumer surplus from type M customers. Self-selection requires that the type H customers not purchase the schedule $<q_M^3, M^3>$. This means that the high demand customers must be guaranteed a consumer surplus V_H^3. The maximum consumer surplus the firm can extract from type H customers given this self-selection constraint occurs when $q_H^3 = q_H^*$ and $H^3=H^1-V_H^3$.

The firm's problem is then to choose the profit maximising level of q_M^3. Note that as the firm reduces q_M^3 by one unit the revenue from the tariff paid by type M customers reduces by $N_M P_M^3$, as P_M^3 is the marginal valuation of type M customers. At the same time the tariff paid by type H customers can be increased by $N_H(P_H^3-P_M^3)$ and still satisfy self selection. The profit maximising level of q_M^3 satisfies $N_M P_M^3 = N_H(P_H^3-P_M^3)$. Firm profit is thus given by:

$$\Pi^3 = N_L L^1 + N_M M^3 + N_H(H^1-V_H^3) \quad (2)$$

where $M^3 < H^1-V_H^3$. The deadweight loss is given by $N_M(M^1- M^3)$.

Note that if $N_M P_M^3 < N_H(P_H^3-P_M^3)$ for all q, then it is profit maximising to set $q_M^3 = 0$. This can happen in two ways:

(i) the ratio of P_H^3 to P_M^3 may be sufficiently high;
(ii) the ratio of N_H to N_M may be sufficiently high.

In this case the firm offers only one schedule to type M and H customers: $<q_H^*,H^1>$. This bundle is purchased only by type H customers. If, on the other hand, $N_M P_M^3 > N_H(P_H^3-P_M^3)$ for q=0 it must be the case that $q_M^3>0$. In this case the firm offers two schedules as described above.

Two variants of the information structure assumed above can be readily analysed: (i) the firm can costlessly separate type M customers from type L and type H customers, but cannot distinguish between type L and type H customers and (ii) the firm can costlessly separate type H from type L and type M customers, but cannot distinguish between type L and type M customers. The analysis of these information structures is analogous to the analysis above.

The methodology used in this subsection can be used when the firm has incomplete information on customer types. In these cases the firm maximises profit by firstly using costlessly available information to separate its customers into groups (direct price discrimination) and then further separating these groups into sub-groups of uniform type using the non-linear pricing schedules as a screening method (indirect price discrimination). Within the groups the customers with the highest demand receive a positive consumer surplus and lower demand customers buy a bundle with inefficient quantity. By offering the lower demand customers an inefficient quantity the lower demand bundles becomes less of a substitute for the highest demand bundle.

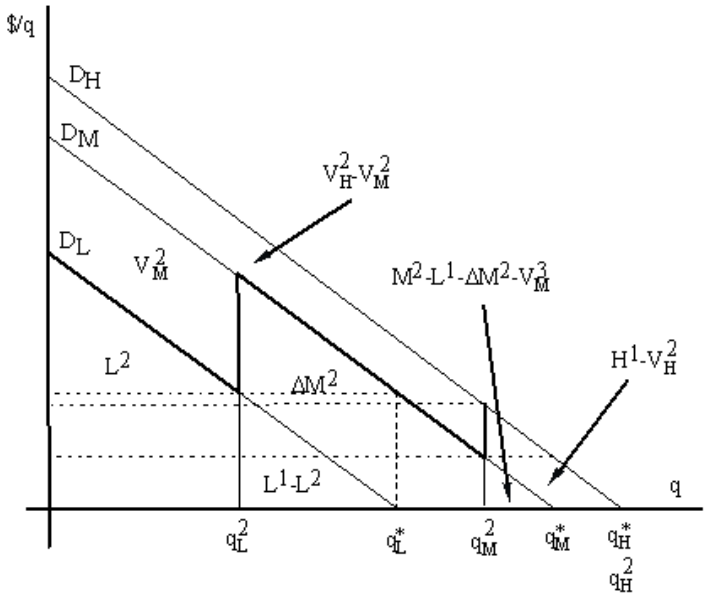
In the bus ticket example adult commuter customers purchase the same number of tickets as they did when the company could identify their customer type whereas adult non-commuter customers purchase a smaller number of tickets than they did when the company could identify their customer type. The adult commuter customers now pay a lower price per ticket and the adult non-commuter customers a higher price per ticket. The student customers pay the same price per ticket and purchase the same number of tickets as before since the company can still costlessly identify this customer type.

Pure indirect non-linear pricing: no customer types costlessly identified

Finally consider the case in which the firm cannot costlessly identify and separate any of the three customer types. In this case the firm does not have the option of using exogenously provided information (such as a student card) to separate customers, but must devise a pricing strategy that identifies a customer type through self-selection. The optimal pricing strategy does this by using pricing schedules in the same way that it separated type M and type H customers above.

The determination of the optimal pricing structure is shown in Figure 3.

Figure 3 Pure indirect non-linear pricing



The firm offers three schedules $<q_L^2,L^2>$, $<q_M^2,M^2>$ and $<q_H^2,H^2>$, directed at type L, type M and type H customers respectively. It is profit maximising for the firm to extract the entire consumer surplus from type L customers. Self-selection requires that the type M (and H) customers not purchase the schedule $<q_L^2,L^2>$. This means that the type M customers must be guaranteed a consumer surplus V_M^2. The profit maximising level of q_L^2 satisfies $N_L P_L^2 = N_M(P_M^2 - P_L^2)$. At this point the revenue lost by reducing q_L^2 from type L customers ($N_L P_L^2$) is just offset by the gain in revenue from type M customers ($N_M(P_M^2 - P_L^2)$).

Self-selection requires that type H customers do not purchase the schedule $<q_M^2,M^2>$. This requires that type H customers be guaranteed a consumer surplus of V_H^2. The profit maximising level of q_M^2 satisfies $N_M P_M^2 = N_H(P_H^2 - P_M^2)$. At this point the revenue lost by reducing q_M^2 from type M customers ($N_M P_M^2$) is just offset by the gain in revenue from type H customers ($N_H(P_H^2 - P_M^2)$). Hence $q_M^2 = q_M^3$ and $V_H^2 = V_M^2 + V_H^3$. Note that if type H customers do not purchase $<q_M^2,M^2>$, they would not purchase $<q_L^2,L^2>$ as it provides a lower consumer surplus.

As in the analysis in the previous section, it is optimal for the firm to set $q_H^2 = q_H^*$ and to set a fee, $H^2 = H^1 - V_H^2$, which is just low enough to deter type H customers from switching to other bundles.

In this case type L customers buy an inefficient quantity and keep no consumer surplus, type M customers buy an inefficient quantity and retain some consumer surplus and type H customers buy an efficient quantity and also retain some consumer surplus. Profit is now:

$$\Pi^2 = N_L L^2 + N_M(M^3 - V_M^2) + N_H(H^1 - V_M^2 - V_H^3) \quad (3)$$

Type H customers buy the efficient quantity q_H^*, type M customers buy the inefficient quantity q_M^2 and type L customers buy the inefficient quantity, q_L^2. The deadweight loss is given by $N_M(M^1 - M^3) + N_L(L^1 - L^2)$.

The above analysis was based on the assumption that $q_M^2 > q_L^2$ where q_M^2 satisfies $N_M P_M^2 = N_H(P_H^2 - P_M^2)$ and q_L^2 satisfies $N_L P_L^2 = N_M(P_M^2 - P_L^2)$. Note that it is possible that these expressions yield values of q_M^2 and q_L^2 such that $q_M^2 < q_L^2$. This would be possible if N_M is relatively small compared with N_L and N_H. Clearly this outcome is not consistent with self-selection. In this case 'bunching' occurs. There is no separate bundle offered to type M, and both type L and type M customers purchase $<q_L^2, L^2>$. In this case q_L^2 is determined by $(N_L+N_M)P_L^2 = N_H(P_H^2 - P_M^2)$.

In the bus ticket example this means the company can no longer identify any of the customer types. The adult commuter customers purchase the same number of tickets but pay a lower price per ticket than when the company could identify the student (and thus adult) customer type. The adult non-commuter customers purchase the same number of tickets but pay a lower price per ticket than when the company could identify the student (and thus adult) customer type. Finally the student customers purchase a smaller number of tickets and pay a higher price per ticket than when the company could identify them.

Profit and efficiency ranking with non-linear pricing

Since profit varies with the level of information that a firm has about its customers we can now show how to rank profit levels according to the information available to the firm. The firm makes the maximum possible profit when it can costlessly identify and separate each customer and offer them a non-linear price that captures their entire consumer surplus. Specifically profit is:

$$\Pi^1 = N_L L^1 + N_M M^1 + N_H H^1 \quad (4)$$

Profit is lower than this maximum when only one customer type rather than all customer types can be costlessly identified and separated. If only type L customers can be costlessly identified and separated, profit is lower than Π^1 because (i) there is a lower fee paid by type M customers because they purchase only q_M^3 (fee M^3) rather than q_M^* (fee M^1) and (ii) type H customers pay a lower fee of $H^1 - V_H^3$ rather than H^1. Mathematically the difference in profit is:

$$\Pi^1 - \Pi^3 = N_M(M^1 - M^3) + N_H V_H^3 \quad (5)$$

The difference in profit is the information cost of separating type M customers. The deadweight loss increases by $N_M(M^1 - M^3)$.

Profit is even lower when no customer types rather than one customer type can be costlessly identified and separated for three reasons: (i) there is a lower fee paid by type L customer because they purchase only q_L^2 (fee L^2) rather than q_L^* (fee L^1); (ii) there is a lower fee paid by type M customers as they pay a fee of $M^3 - V_M^2$ rather than M^3; and (iii) type H customers pay a lower fee of $H^1 - V_M^2 - V_H^3$ rather than $H^1 - V_H^3$. Mathematically the difference in profit is:

$$\Pi^3 - \Pi^2 = N_L(L^1 - L^2) + N_M V_M^2 + N_H V_M^2 \quad (6)$$

The difference in profit is the information cost of identifying and separating type L customers. The deadweight loss increases by $N_L(L^1 - L^2)$.

In summary, the less capable the firm in identifying a customer's type (and the more it must rely on self-selection), the lower is its profit. Thus the information on customer type can be viewed as a valuable commodity. The more the firm must rely on pricing strategies to reveal a customer's type the greater is the deadweight loss.

5. Price discrimination with linear prices

Having introduced students to direct and indirect price discrimination using non-linear pricing the student may well ask how this relates to the ubiquitous analysis of firms who use linear pricing? This question should be addressed. The answer is that firms use linear pricing when it is not possible to use non-linear pricing strategies. There are two important circumstances in which the firm would not use non-linear pricing. The first is where all customer types have unit demand, and the firm cannot identify an individual customer's type (i.e. their willingness to pay): for example, if all bus passengers had demand for only one bus ride, but all varied in their willingness to pay for that ride. In such instances the elasticity of demand depends on the distribution of customers' willingness to pay. (Third degree price discrimination could be practised if the firm could observe group types, and the groups differ in their distribution of willingness pay.) However in those cases in which non-linear pricing can be used, customers vary in the quantity they demand. To relate the analysis of non-linear pricing to linear pricing, it is useful to consider the second circumstance in which the firm will use linear pricing. That is, in the case in which the good can be (costlessly) resold. If the firm cannot prevent resale it must offer a linear price (i.e. a common price per unit) to all customers in order to prevent arbitrage between customers. If the firm did attempt to utilise non-linear pricing, bundles could be unpacked and sold at the average price of the good in the bundle. This price would become the linear market price. Thus, when the firm cannot prevent resale, it will not be able to avoid setting a linear price. For instance, if bus tickets were sold in a bundle of 10 for $10, anyone could obtain a single ticket for $1 if (costless) resale were possible. This would effectively be the linear market price the firm is setting.

As in the case of non-linear pricing the firm using linear pricing may have full, partial or no information on individual customer's type. Pigouvain third degree price discrimination occurs when there is both linear pricing (because of intra-group arbitrage) and when the firm observes some exogenous information as to a customer's type.

Figure 4: Price discrimination with linear prices

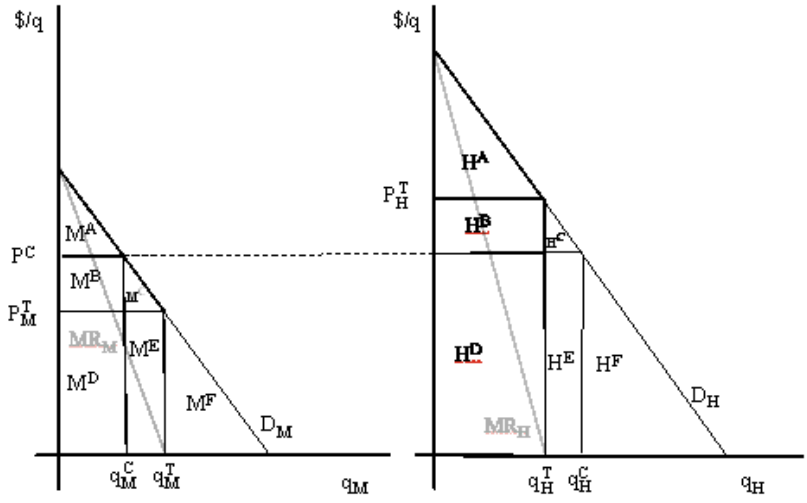


Figure 4 casts textbook (Pigouvian) third degree price discrimination in a manner that facilitates comparison with the above analysis of non-linear pricing. For brevity we present the detailed analysis with only two customer types (the minimum number of types needed to consider the implications), which are labelled M and H.

If the firm cannot identify a particular customer's type, and intra-group arbitrage is possible, it must set a common linear price. This is the case of a simple monopoly. The profit maximising, common linear price is shown as P^C, and the firm maximum profit when the firm cannot identify customer type is thus:

$$\Pi^C = N_M(M^B+M^D) + N_H(H^D + H^E) \quad (7)$$

The deadweight loss is: $N_M(M^C + M^E + M^F) + N_H H^F$.

Now consider the case in which the firm can costlessly identify customers by their type, and intra-group (but not inter-group) arbitrage is possible. The usual textbook examples of student discount on cinema tickets and geographical separation satisfy this requirement. The linear price P_M^T is charged to type M customers, who purchase quantity q_M^T, and the linear price P_H^T is charged to type H customers, who purchase quantity q_H^T. The firm's profit is:

$$\Pi^T = N_M(M^D + M^E)+ N_H(H^B + H^D) \quad (8)$$

The deadweight loss is: $N_M M^F + N_H(H^C + H^E + H^F)$.

In the bus ticket example the commuter customer can purchase any number of tickets but pays the same price per ticket irrespective of the number of tickets purchased. Likewise the non-commuter customer can purchase any number of tickets but also pays the same price per ticket irrespective of the number of tickets purchased. The commuter however pays a higher price per ticket than the non-commuter if the former has a lower elasticity of demand (which we might expect to be the case for those with 'high' demand).

The impact of the move from common linear pricing to third degree (linear) price discrimination changes profit by:

$$\Delta\Pi = N_M(M^E - M^B)+ N_H(H^B - H^E) \quad (9)$$

as type H has inelastic demand and type M has elastic demand $\Delta\Pi>0$. Thus the firm is able to utilise the information identifying customer type to increase its profit. This result is consistent with the analysis above.[2]

However, in contrast to the above analysis, deadweight loss does not reduce as firms are provided with additional information. Specifically, if the firm moves from common linear pricing to third degree (linear) price discrimination the deadweight loss increases by:

$$N_H(H^C + H^E) - N_M(M^C + M^E) \quad (10)$$

This expression is negative when demand curves are linear and output is unchanged but may not be negative in other cases. Robinson (1933) showed that the movement toward third degree discriminating prices alters the distribution of output but does not change total output when demand

[2] Note that the firm would prefer to use a non-linear price rather than a linear price. This is demonstrated by showing that the firm can increase profit by switching to a non-linear price from a linear price. If the firm could prevent resale (and thus profitably conduct non-linear pricing), it could sell a bundle consisting of q_M^C to type M customers for a fee of $M^A+M^B+M^D$, and sell a bundle of q_H^C to type H customers for a fee of $M^A+H^D+H^E$. This strategy increases the profit per customer by M^A and additionally satisfies self-selection. Hence using non-linear pricing increases profit relative to linear pricing. Of course the firm can maximise profits by adopting the pricing described above (Figure 2). We thus assume that the firm is exogenously forced to set a linear price so that we can compare textbook treatments of third degree price discrimination with the above analysis.

curves are linear. Schmalensee (1981) showed that deadweight loss increases unless output increases. Thus, when demand curves are linear, the implementation of third degree price discrimination increases deadweight loss. Ikeda and Toshimitsu (2010), using a model in which consumers have unit demand and quality is endogenous, show that an increase in total output is a necessary condition for welfare improvement with third degree price discrimination by a monopolist. This is apparent from Figure 4 once it is realised that with linear demand curves $N_M(q_M^T - q_M^C)$ must equal $N_H(q_H^C - q_H^T)$.

Thus, when output does not increase, the provision of information on customer type that allows the firm to implement third degree price discrimination lowers welfare. This conclusion, implicit in textbook treatments, is the opposite of the result presented above. The source of the divergence in the conclusion is due to the use (by the Pigouvian taxonomy) of linear pricing rather than non-linear pricing. This results in the firm reducing output to the type H customers in an attempt to capture their consumer surplus. The output supplied to type M customers however increases. This result contrasts to that obtained above for non-linear pricing, where output supplied to type M customers decreases due to implementation of price discrimination. By comparison Ikeda and Toshimitsu (2010) show that third degree price discrimination always enhances welfare when quality is endogenous, mainly because of the quality improvement owing to price discrimination increases consumer surplus.

Thus the conclusions derived from the textbook analysis of (Pigouvian) third degree price discrimination follows from the joint assumptions of linear pricing (i.e. intra-group arbitrage but no inter-group arbitrage) and the availability of an exogenous signal on customer types (or groups of types). The above analysis can be readily extended to the case of three customer types. In this event, as with non-linear pricing there are three possibilities:

> (i) pure direct linear pricing in which each type faces their own linear price;
> (ii) partial direct linear pricing in which the firm observes an exogenous signal about one type (which we take to be type L) and cannot distinguish between the other two types (H and M): in this case the firm charges one linear price to the type with the exogenous signal (type L) and a common price to the types that cannot be separated (H and M); and
> (iii) common linear pricing in which the firm cannot distinguish between all three customer types.

Using the above analysis it is readily demonstrated that the profit of the firm falls as the number of types it can identify falls (i.e. profits fall as we move from case (i) to (ii) and from (ii) to (iii)). Thus the firm will utilise exogenous signals on customer types whenever it is available.

6. An integrated view

The analysis above can be brought together to provide an integrated view of the firm's pricing strategy. This is summarised in Table 1.

Table 1: Matrix of price discrimination options available for the firm

		Intra-group arbitrage	
		Not possible	Possible
Information on customer type	Perfect	Pure Direct, non-linear	Direct, linear
	Partial	Partial Direct, non-linear	Partial Direct, linear
	None	Pure Indirect, nonlinear	Common linear price

The firm faces two exogenous factors in the market environment when setting pricing. The first is whether it can costlessly utilise some exogenous signal to identify customers by their type. The second is whether customers can undertake intra-group arbitrage. (Note if inter-group arbitrage is also possible then the firm must use a common linear price.) The table indicates the pricing strategies adopted in the six market environments. The analysis above indicates that the firm's profit falls as we move either down or across the table. Thus firms always have an incentive to bundle output when possible and utilise available signals of customer types. Note that, as pointed out by Carroll and Coates (1999), it is not possible to tell whether profits are larger under pure indirect non-linear pricing (second degree price discrimination in textbooks) and direct linear pricing (third degree price discrimination in textbooks). However making this comparison is misleading for students. If the firm could conduct non-linear pricing and could identify customers by type, it would be better off using direct non-linear pricing.

Many students will benefit from seeing a numerical example of the model described above. The results of a numerical simulation are presented in Table 2. In this simulation it is assumed that $V_i(q)=q(a_i-q/2)$, where $a_H=20$, $a_M=18$ and $a_L=16$, and $N_H=N_M=N_L=10$. The table allows a comparison across pricing strategies of the quantity sold to each type (q_i), the price per unit paid by each type ($\bar{P}_i$), e profit generated by each type (Π_i), total profit (Π), consumer surplus obtained by each type (CS_i), total consumer surplus (CS), deadweight loss associated with each type (DWL_i) and total deadweight loss (DWL). There are numerous effects illustrated by this example, and the instructor can emphasise particular effects as appropriate. However in this table one can clearly see the benefit to the firm from an exogenous signal about customer type, and from the ability to conduct non-linear pricing. The impact of the presence of exogenous customer information on consumer surplus and deadweight loss is also illustrated.

Table 2: Numerical simulation of price discrimination strategies

	Quantity			**Price per unit**			**Profit**			
	q_H	q_M	q_L	$\bar{P}_H$	$\bar{P}_M$	$\bar{P}_L$	Π_H	Π_M	Π_L	Π
Pure direct, non-linear	20	18	16	10	9	8	200 (=H^1)	162 (=M^1)	128 (=L^1)	490
Partial direct, non-linear	20	16	16	8.4	10	8	168 (=H^3)	1600 (=M^3)	1280 (=L^1)	456
Pure indirect, non-linear	20	16	12	7.2	8.5	10	144 (=H^2)	136 (=M^2)	120 (=L^2)	400
Direct, linear	10	9	8	10	9	8	100	81	64	245
Partial direct, linear	10.5	8.5	8	9.5	9.5	8	99.75	80.75	64	244.5
Common linear price	11	9	7	9	9	9	99	81	63	243

Table 2 (cont.): Numerical simulation of price discrimination strategies

	Consumer surplus				Deadweight loss			
	CS_H	CS_M	CS_L	CS	DWL_H	DWL_M	DWL_L	DWL
Pure direct, non-linear	0	0	0	0	0	0	0	0
Partial direct, non-linear	32 ($=V_H^3$)	0	0	32	0	2	0	2
Pure indirect, non-linear	56 ($=V_H^2$)	24 ($=V_M^2$)	0	80	0	2	8	10
Direct, linear	50	40.5	32	122.5	50	40.5	32	122.5
Partial direct, linear	55.125	36.125	32	123.25	45.125	45.125	32	122.25
Common linear price	60.5	40.5	24.5	125.5	40.5	40.5	40.5	121.5

7. Conclusions

When a market consists of more than one customer type a firm can potentially maximise its profits by charging different types different prices. We present a taxonomy that can make it clear to students that the ability of the firm to realise these prospective profits depends on (i) the information it has on customers and (ii) the ability of the firm to use non-linear pricing (specifically the absence of intra-group arbitrage). This approach demonstrates to students that information on customer types is valuable to firms. When a given customer's type is private information the firm can use non-linear pricing schedules to provide customers with an incentive to reveal their type. However, our approach demonstrates how extracting this information comes at a cost to the firm: profit is lower than would be the case if customers' types were common information. In addition we provide a method of demonstrating to students how the optimal screening method distorts the quantity available to low demand customers and thus generates a deadweight loss. When the firm cannot use non-linear pricing (i.e. when intra-group arbitrage is present) the firm cannot use a pricing mechanism to identify a given customer's type, and its profit consequently suffers.

We have presented an integrated approach to teaching the theory of price discrimination. This approach allows for a more coherent understanding of the different strategies adopted by firms. It emphasises that firms use price discrimination strategies as a means of maximising profit given the constraints imposed by their market's characteristics. This approach enables a straightforward explanation of the pricing strategies used by firms in many common real world examples.

References

Carlton, D. and Perloff, J. (2004). *Modern Industrial Organization,* Sydney: The Addison-Wesley Series in Economics.

Carroll, K. and Coates, D. (1999). 'Teaching price discrimination: Some clarification', *Southern Economic Journal*, vol. 66(2), pp. 466–80.

Dolan, R. J. and Simon, H. (1996). *Power Pricing: How Managing Price Transforms the Bottom Line*, New York: The Free Press.

Goldman, M.B., Leland H.E. and Sibley, D.S. (1984). 'Optimal nonuniform pricing', *The Review of Economic Studies,* vol. 51(2), pp. 305–19.

Gotlibowski, C. and Kahana, N. (2009). 'Second-degree price discrimination: A graphical and mathematical approach', *Journal of Economic Education*, vol. 40(1), pp. 68–79.

Hubbard, R.G. *et al.* (2009). *Microeconomics,* Frenchs Forest NSW: Pearson Education Australia.

Ikeda, T. and Toshimitsu, T. (2010). 'Third degree price discrimination, quality choice, and welfare', *Economic Letters*, 106, pp. 54–6.

Jeitschko, T.D. (2001). 'Issues in price discrimination: A comment on and addendum to "Teaching price discrimination: Some clarification" by Carroll and Coates', *Southern Economic Journal*, vol. 68(1), pp. 178–86.

McTaggart, D. *et al.* (2010). *Microeconomics*, Frenchs Forest NSW: Pearson Education Australia.

Maskin, E. and Riley, J. (1984). 'Monopoly with incomplete information', *RAND Journal of Economics,* vol. 15(2), pp. 171–96.

Perloff, J.M. (2009). *Microeconomics,* Boston Massachusetts: Pearson Education, Inc.

Pigou, A. C. (1920). *The Economics of Welfare,* London: Macmillan.

Robinson, J. (1933). *The Economics Of Imperfect Competition,* London: Macmillan.

Schmalensee, R. (1981). 'Output and welfare implications of monopolistic third-degree price discrimination', *American Economic Review,* vol. 71(1), pp. 242–7.

Spence, A. M. (1977). 'Nonlinear prices and welfare', *Journal of Public Economics,* vol. 8, pp. 1–18.

Stole, L. (2007). 'Price Discrimination and Competition', in (M. Armstrong and R. Porter, eds), *Handbook of Industrial Organization, Volume 3,* North-Holland: Elsevier.

Swann, M. and W.A. McEachern. (2003). *Microeconomics. A Contemporary Introduction*, Southbank Victoria: Thomson.

Taylor, J.B. and Frost, L. (2006). *Microeconomics,* Milton Queensland: John Wiley and Sons.

Tirole, J. (1988). *The Theory of Industrial Organization*, Cambridge: The MIT Press.

Varian, H. (1987). 'Price Discrimination', in (R. Schmalensee and R. Willig, eds), *Handbook of Industrial Organization, Volume 2,* North-Holland: Elsevier.

Varian, H. (2006). *Intermediate Microeconomics: A Modern Approach*, New York and London: W.W. Norton & Company.

Author Biography

Ann Marsden is currently lecturer with the Greenwich School of Management. Ann previously held the position of lecturer with the School of Economics and Finance, University of Tasmania. Her research and teaching interests are in the industrial economics area with a particular focus on pricing. She has also worked as an economist with London Underground and holds an MSc (Economics) from Birkbeck College. She has recently submitted her Doctoral thesis from the University of Tasmania.

Hugh Sibly is a senior lecturer in the School of Economics and Finance, University of Tasmania. He teaches in the areas of industrial organisation, game theory and macroeconomics. His areas of research interest include pricing in macroeconomics, utility pricing, industrial organisation and behavioural economics. Hugh is also a member of the Australian Copyright Tribunal. He received his PhD from the University of Melbourne.

Contact details

Ann Marsden
Greenwich School of Management
Royal Hill
London
SE10 8RD
United Kingdom
Tel: +44 (0)20 7254 8214

Hugh Sibly
University of Tasmania
Churchill Avenue
Sandy Bay
TAS 7005
Australia
Tel: +61 3 6226 2825
Email: Hugh.Sibly@utas.edu.au

Liquidity Trap in an Inflation-targeting Framework: A Graphical Analysis[1]

Pavel Kapinos

Abstract

This paper presents a simple New Keynesian model with alternative assumptions regarding the conduct of monetary policy. The central bank is assumed to either follow a Taylor rule or minimise a social welfare loss function. The model can be tractably described by means of a straightforward graphical apparatus, which, so far, has not been extended to include the treatment of the liquidity trap. The paper presents an analysis of the zero nominal interest rate bound using this apparatus and discusses the implications of pre-emptive monetary easing when the macroeconomic conditions suggest that the bound may restrict future monetary policy effectiveness.

JEL classification: A22, E32, E52

1. Introduction

On 16 December 2008, the Federal Open Markets Committee established 'a target range for the federal funds rate of 0 to 1/4 percent'. For several weeks prior to this announcement, the effective federal funds rate and the yield on Treasury bills had hovered in that range, making the FOMC announcement a mere recognition of existing reality and raising the practical relevance of the zero bound on the nominal interest rate for the conduct of monetary policy. The United States has thus joined Japan whose own experience with the liquidity trap in the late 1990s provided the impetus for the recent efforts to study monetary policy in this environment.

The recent theoretical literature on this subject extends the now standard New Keynesian models (see Clarida, Gali and Gertler (1999) and Woodford (2003) for a comprehensive treatment) into the setting where the nominal interest rate is bound by zero.[2] This strand of literature suggests that monetary policy should be more accommodative towards inflation than in the models where this bound does not feature. Eggertson and Woodford (2003) study the effect of this constraint in the context of a forward-looking model with perfect foresight and show that the effects of the liquidity trap can be reduced by the central bank's credible commitment to keep the nominal interest rates low even after the effects of a negative demand shock have passed. This allows forward-looking agents to act on the promise of an inflationary environment with negative real interest rates in the present and hence lifts the economy

[1] The author thanks the editor and an anonymous referee for valuable suggestions. All remaining errors are the author's.

[2] Svensson (2003a) and Jeanne and Svensson (2007) discuss mechanisms for escaping the liquidity trap in the context of an open economy, such as Japan. This paper focuses on the treatment of issues that arise in the closed-economy setting.

out of the liquidity trap sooner. Adam and Billi (2006, 2007) study the conduct of optimal monetary policy under commitment and discretion without perfect foresight, in a stochastic environment, and find that the zero nominal interest rate bound makes it optimal for the central bank to pursue the reduction of nominal interest rates more aggressively and pre-emptively in its response to negative demand shocks than in the case where liquidity trap is not a possibility. Kato and Nishiyama (2005) analytically derive the optimal monetary policy reaction function when the zero nominal interest constraint binds in the context of a backward-looking deterministic model and find that it is more aggressive and expansionary than the rule that is optimal in the absence of that bound.

The main contribution of the present paper is to develop the graphical apparatus for discussing these results that should be accessible to an audience of undergraduate students and non-specialists. Starting with the work of Romer (2000) and Taylor (2000), modern monetary theory has been rendered more accessible to wider audiences. Carlin and Soskice (2005) survey several alternatives and propose a tractable model with backward-looking dynamics. Bofinger, Mayer and Wollmershäuser (2006) build on their work to present extensive graphical analysis of the conduct of monetary policy in the inflation-targeting framework. Guest (2003) and Turner (2006) present alternative specifications of inflation-targeting models that facilitate their exposition in the classroom. Furthermore, several undergraduate textbooks, such as Carlin and Soskice (2006) and Jones (2007), have used a New Keynesian model with adaptive inflationary expectations as the framework for analysing short-run fluctuations. Weise (2007) shows that a pedagogical version of the New Keynesian model can be extended to incorporate additional considerations, such as the term structure of interest rates. Kapinos (2010) provides a description of the several versions of the model using an Excel workbook. This paper considers a setup that is similar to the ones studied in this body of work and extends it to examine the effect of the zero nominal interest rate bound.

The present paper also studies the consequences of uncertainty surrounding the possibility that a large negative demand shock – the standard source of the liquidity trap – may (or may not) materialise in the future. A central bank's ability to anticipate a large negative demand shock and hence engage in pre-emptive monetary easing that is shown to be optimal in the context of this paper's model may dramatically reduce social welfare losses relative to the scenario where no pre-emptive action is undertaken. However, pre-emptive easing will generate social welfare losses if the shock fails to materialise. This paper provides a framework for studying the tradeoffs involved in this setup.

The rest of the paper is organised as follows. The following section describes the changes in the standard analytical apparatus that result from the zero nominal interest rate bound for two versions of a model of short-run dynamics under alternative assumptions regarding the conduct of monetary policy. We then argue in favour of a pre-emptive monetary easing when the future macroeconomic conditions make the zero nominal interest rate bound a distinct possibility using the graphical apparatus developed in the previous section. Finally, the article concludes.

2. Monetary policy rules in a New Keynesian model with adaptive inflationary expectations

This section presents a brief overview of the now standard three-equation model that has been used for the analysis of monetary policy. First, the household sector Euler equation describes the negative relationship between a measure of real activity, such as the output gap, and the real interest rate. Alternatively, it can be motivated by the standard Keynesian treatment of the aggregate expenditure function. Second, the firms' first-order condition with respect to their own prices in a staggered pricing framework gives rise to inflation as a positive function of output gap. Finally, the monetary authority is assumed to follow either an instrument rule where it sets the nominal interest rate in response to deviations of inflation from the long-term target and output gap or a targeting rule where it minimises a

social welfare loss function period by period.[3] The former was proposed by Taylor (1993) as a descriptor of the conduct of monetary policy in the United States and the latter can be derived as a second-order approximation of the household utility function.

Instrument version

A popular way to model a central bank's decision to set the nominal interest rate is by means of a Taylor rule, whereby the target rate is a function of inflation, output gap and possibly other variables. Clarida, Gali and Gertler (2000) provide a comprehensive study that analyses the Taylor rule in the context of the US data. This monetary rule (referred to as the MR schedule in this paper) has the pedagogical advantage of discussing the systematic conduct of monetary policy without resorting to calculus. The instrument version of the model, therefore, is defined by the standard IS, aggregate supply (AS), and MR equations:

$$x_t = \sigma r^* - \sigma(i_t - \pi_t) + \varepsilon_t^x, \tag{1}$$

$$\pi_t = \pi_{t-1} + \kappa x_t + \varepsilon_t^\pi, \tag{2}$$

and

$$i_t = i^* + \gamma_\pi(\pi_t - \pi^*) + \gamma_x x_t + \varepsilon_t^i, \tag{3}$$

where π^* is the inflation rate targeted by the central bank, r^* is the long-run real interest rate that is consistent with the output gap of zero and, as it follows, $i^*=r^*+\pi^*$; π_t is inflation, x_t is the output gap and i_t is the nominal interest rate.[4] The so-called Taylor principle states that the stabilising monetary policy requires that the central bank should respond to a percentage increase in inflation by raising the nominal interest rate by more than 1%, hence $\gamma_\pi>1$. Empirical studies of versions of the model that feature forward-looking expectations, such as Cho and Moreno (2006) and Rabanal and Rubio-Ramirez (2005), suggest that the values for the elasticity of intertemporal substitution, σ, slope of the Phillips curve, κ, and the degree of central bank's responsiveness to output gap, γ_π, all fall in the range between 0 and 1. Plugging (3) into (1) and solving for inflation, one can obtain the aggregate demand (AD) relation:

$$\pi_t = \pi^* + \frac{\varepsilon_t^x - \sigma\varepsilon_t^i}{\sigma(\gamma_\pi - 1)} - \frac{\sigma\gamma_x + 1}{\sigma(\gamma_\pi - 1)} x_t. \tag{4}$$

Together with (2) it determines the equilibrium output gap and inflation; the latter is taken as exogenous in (1) and (3). Note that the Taylor principle ensures that the slope of aggregate demand is negative.

---

[3] Svensson (2002, 2003b) introduced this 'instrument'/'targeting' nomenclature to distinguish between a monetary policy response function that sets the nominal interest rate as a linear function of inflation and output gap and the central bank's minimisation of a social welfare loss objective.

[4] Note that since the IS/MR diagrams below will have the nominal interest rate on the vertical axis, the graphical version of the IS schedule can be written as:

$$i_t = r^* + \pi_t - \frac{1}{\sigma} x_t + \frac{1}{\sigma} \varepsilon_t^x,$$

Unlike the models where the real interest rate features in the IS/MR plane, both IS and MR will shift due to changes in inflation. Introducing this minor complication, however, provides a more direct way to study the effect of the zero bound on the nominal interest rate.

As in the Keynesian tradition, the zero nominal interest rate bound on monetary policy becomes effective due to a large negative demand shock, ε_t^x<0.[5] Formally, the condition for the requisite size of the negative demand shock to generate the liquidity trap is given by the following inequality:

$$\varepsilon_t^x \le -\frac{\sigma\kappa(\gamma_\pi - 1) + \sigma\gamma_x + 1}{\kappa\gamma_\pi + \gamma_x} r^* + \frac{(\gamma_\pi - 1)(1 - \sigma\kappa)}{\kappa\gamma_\pi + \gamma_x} \pi^* + \frac{\sigma\gamma_x + \gamma_\pi}{\kappa\gamma_\pi + \gamma_x} (\pi_{t-1} + \varepsilon_t^\pi) - \frac{1 - \sigma\kappa}{\kappa\gamma_\pi + \gamma_x} \varepsilon_t^i \quad (5)$$

This inequality has an intuitive interpretation. To make the bound effective, the negative demand shock has to overcome the positive pressure on the interest rate coming from a contractionary monetary policy shock,[6] the long-run real interest, a cost-push shock and the long-run inflation target. Note that, since in the long run $\pi_{t-1}=\pi^*$, the coefficient on the long-run inflation target will be $-\frac{1 - \sigma\kappa + \gamma_\pi\sigma\kappa + \sigma\gamma_x}{\kappa\gamma_\pi + \gamma_x} < 0.$ This has important implications to be discussed more fully below. A higher choice for the long-run inflation target by the central bank makes it less likely that a negative demand shock of a given size will generate the liquidity trap by providing an inflationary cushion.[7]

Once the negative demand shock is sufficiently large, the IS and MR equations change to reflect the zero nominal interest rate bound. The latter is now simply given by i_t=0, while the IS schedule (1) and, correspondingly, AD schedule (4) become, respectively:

$$x_t = \sigma r^* + \sigma\pi_t + \varepsilon_t^x, \quad (6)$$

$$i_t = r^* + \pi_t - \frac{1}{\sigma} x_t + \frac{1}{\sigma} \varepsilon_t^x. \quad (7)$$

This suggests that the IS schedule becomes vertical and the AD schedule becomes upward sloping. The latter result is obtained because the central bank will no longer be able to engage in stabilising monetary policy and lower the nominal interest rate more than proportionately in response to a given decrease in inflation, because the nominal interest rate is at zero. Furthermore, comparing the slopes of

---

[5] In particular, solving (2) and (4) for equilibrium inflation and output and using (3) to determine the equilibrium interest rate, we have:

$$i_t = r^* - \frac{(\gamma_\pi - 1)(1 - \sigma\kappa)}{\sigma\kappa(\gamma_\pi - 1) + \sigma\gamma_x + 1} \pi^* + \frac{\sigma\gamma_x + \gamma_\pi}{\sigma\kappa(\gamma_\pi - 1) + \sigma\gamma_x + 1} (\pi_{t-1} + \varepsilon_t^\pi)$$
$$+ \frac{\kappa\gamma_\pi + \gamma_x}{\sigma\kappa(\gamma_\pi - 1) + \sigma\gamma_x + 1} \varepsilon_t^x + \frac{1 - \sigma\kappa}{\sigma\kappa(\gamma_\pi - 1) + \sigma\gamma_x + 1} \varepsilon_t^i.$$

Setting this equation equal to zero and solving for the demand shock, one can obtain the size of the shock necessary to make the zero nominal interest rate constraint binding. If the liquidity trap is induced by a negative supply shock, the economy will first experience lower inflation and a positive output gap, as the AS schedule shifts down, and then, as adaptive expectations start working, the AS will keep shifting down along the upward-sloping portion of the AD schedule until output gap is zero. Hence the economy will settle at a lower inflation rate and zero output gap, a scenario not nearly as dramatic as the model's response to large negative demand shock.

[6] Clearly, the central bank would generate large costs in terms of a negative output gap by engaging in a discretionary contraction while the demand shock is large and negative.

[7] There are additional benefits associated with higher inflation targets. Sensitivity analysis with respect to key parameters of the infinite horizon model laid out in the Appendix and impulse response functions of the variables to an anticipated future negative demand shock are available from the author upon request. These results show that when the shock materialises, the relative gains from pre-emptive easing discussed below decline sharply, suggesting that guessing whether or not a sizeable negative shock will realise will be less necessary. See Billi (2009) for a derivation of a positive optimal inflation rate under model uncertainty and occasionally bounding zero nominal interest rate and Billi and Kahn (2008) for a tractable discussion of issues driving this result.

(7) and (2), a straightforward determinacy condition must be satisfied for unique equilibrium inflation and output gap to exist:

$$\frac{1}{\sigma} > \kappa. \tag{8}$$

This is, the upward-sloping branch of AD must be steeper than AS. If this condition does not hold, a unique equilibrium will not exist in this model.

Figure 1 Transitory vs. permanent negative demand shock in the instrument model

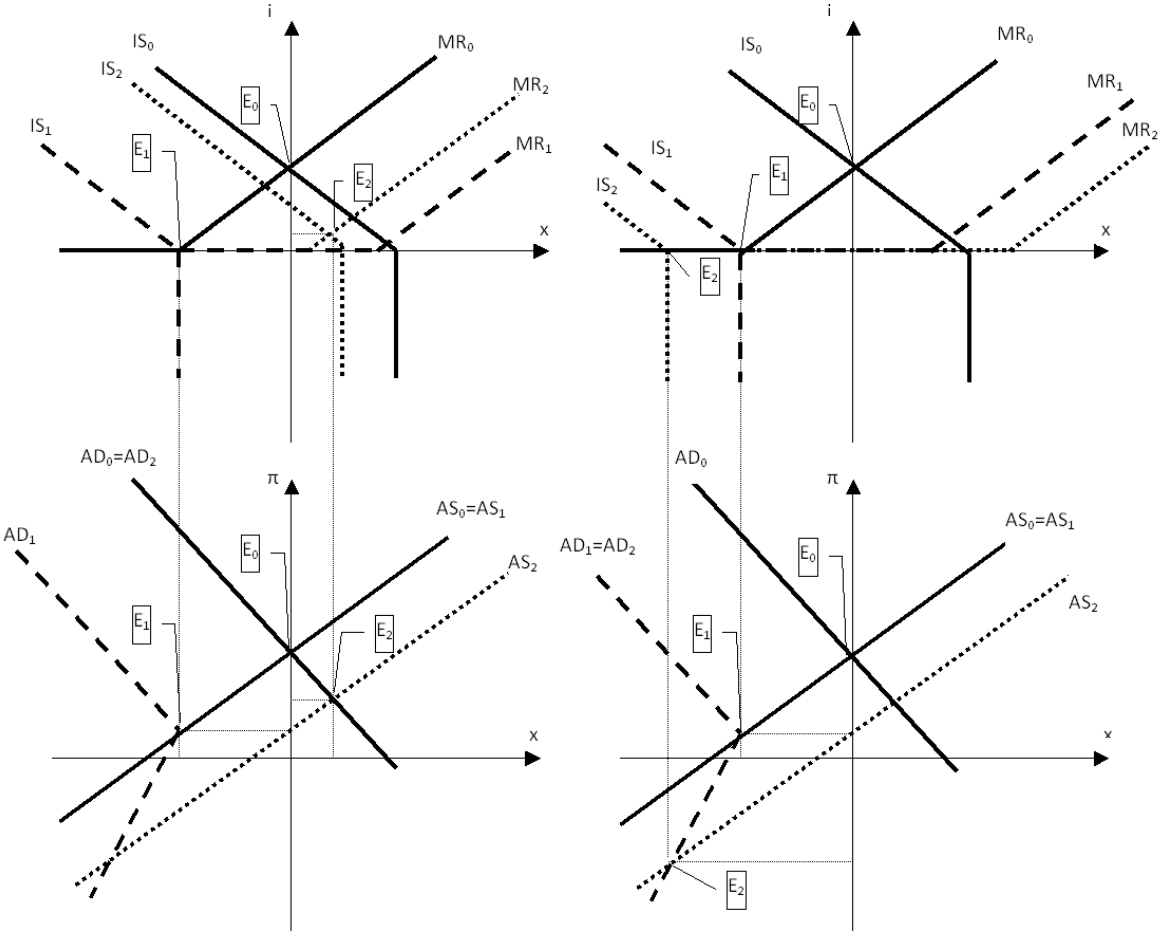


Figure 1 demonstrates the effect of a large negative demand shock using the graphical representation of the model's schedules, assuming that the model is characterised by the long-run equilibrium at time t=0. For simplicity, the figure assumes that the shock at t=1 is such that the equilibrium nominal interest rate exactly equals 0. The left panel assumes that the shock is transitory and fully disappears at t=2. The IS schedule shifts down from IS_0 to IS_1 due to the shock and the effect of lower equilibrium inflation. The MR shift from MR_0 to MR_1 is also motivated by lower equilibrium inflation, but, once the zero bound on the nominal interest rate is reached, it becomes horizontal. Since the central bank is no longer able to follow the Taylor principle and respond to falling inflation by decreasing the nominal interest rate by even larger amounts, the AD schedule becomes upward-sloping. At t=2, firms peg their inflationary expectations to the equilibrium level of π at t=1 and the AS schedule shifts down to AS_2. However, by then the effect of the transitory shock on the AD schedule has disappeared, hence AD returns to $AD_2=AD_0$. This means that $\pi_2>\pi_1$ and, as inflationary expectations adjust adaptively, the economy eventually returns to the long-run equilibrium.

The right panel of Figure 1 demonstrates that for the liquidity trap to be a salient issue in this model, the negative demand shock has to be sufficiently persistent. In this panel, the demand shock is permanent: AD_0 shifts down to AD_1=AD_2 and stays in that position. Now that lower equilibrium inflation pushes the AS schedule downwards in subsequent time periods along the upward-sloping demand schedule, the model does not have a stable equilibrium: output gap and inflation will both fall indefinitely.[8] As inflation falls, the vertical IS schedule shifts farther to the left.

Avoiding the liquidity trap, therefore, is necessary to prevent economic destabilisation in this setup. The next section will discuss how this may be done by the central bank's engaging in pre-emptive expansionary monetary policy. Alternatively, instructors may discuss the use of fiscal policy that here is modelled by the demand shock, ε_t^x. A negative demand shock that comes, for instance, from a dramatic decrease in asset prices may be offset by means of a fiscal expansion. Timing issues are also critical here: the longer the fiscal policy makers wait to implement a stimulus package, the deeper the economy will sink into a recession.

Figure 1 also illustrates the recent ostensibly puzzling finding of Eggertsson (2008): an adverse cost-push shock, say, due to higher unionisation rates that raise firms' labour costs, will have beneficial effects on output gap and inflation and help the economy escape from the liquidity trap. This result has an intuitive interpretation. The main problem in a liquidity trap is falling inflation, which, since the nominal interest rate is at zero, raises the real interest rate and depresses the output gap that, in turn, lowers inflation even further. A cost-push shock raises inflation and hence breaks this vicious cycle. Of course, the benefits of these higher labour costs pertain only to the situation when the zero nominal interest rate bound is effective. Implementing policies that permanently increase firms' costs may be hard to reverse, once the economy recovers from the liquidity trap.

Targeting version

Svensson (2002, 2003b) has long criticised the use of Taylor rules as descriptors of monetary policy. Although they do seem to summarise a large proportion of a modern central bank's decision-making process, no central bank has formally committed to following such a rule. Instead, the conduct of monetary policy may be more adequately described by the central bank's choosing a nominal interest rate so as to minimise a discounted stream of social welfare losses given by the weighted average of squared departures of inflation from its target level and output gap from its target level of zero. Carlin and Soskice (2005, 2006) provide a graphical description of this problem in the case where the zero nominal interest rate does not place a bound on the conduct of monetary policy.[9]

In this sub-section, instead of following an instrument rule, such as (3), the central bank conducts optimal monetary policy by minimising a social welfare loss objective on a period-by-period basis

[8] Note that this does not happen with respect to a permanent negative demand shock that is small enough not to induce the liquidity trap. In that case, the economy will eventually converge to zero output gap and lower long-run level of inflation; this point is discussed in the following section. Furthermore, it should be clear that the sustained fall in output gap and inflation is due to adaptive inflationary expectations. A similar, if smaller in magnitude, result would hold if a fraction of firms anchored their expectations to the central bank's inflation target and the remaining firms continued to form their expectations adaptively. It is also possible to study the effect of a permanent demand shock with forward-looking expectations, as in Eggertson and Woodford (2003); however, that setup does not lend itself easily to graphical analysis.

[9] Wiese (2007) also moves away from the Taylor rule description of monetary policy and allows the central bank to set the interest rate exogenously. Postulating a social welfare loss function provides a criterion for the level that the central bank may want to choose.

without taking into consideration the effect of adaptive inflationary expectations on future outcomes.[10] Algebraically, the central bank solves:

$$\min_{\pi_t, x_t} \left\{ \left(\pi_t - \pi^* \right)^2 + \alpha x_t^2 \right\}, \tag{9}$$

subject to (2). The first-order condition to this problem is the aggregate demand schedule:

$$\pi_t = \pi^* - \frac{\alpha}{\kappa} x_t . \tag{10}$$

and the interest rate consistent with implementing this optimal policy can be recovered from (1).

Ordinarily, it is only a cost-push shock that represents a trade-off between stabilising inflation and output gap and generates a social welfare loss; relatively small demand shocks can be offset by adjusting the nominal interest rate, such that $\Delta i_t = \frac{1}{\sigma} \varepsilon_t^x$. However, if the negative demand shock is sufficiently large, the zero nominal interest rate bound becomes effective and the central bank will not be able to offset the shock fully.[11] The zero nominal interest rate bound, therefore, will become binding if:

$$\varepsilon_t^x \leq -\sigma r^* + \frac{\kappa(1-\sigma\kappa)}{\kappa^2 + \alpha} \pi^* - \frac{\alpha\sigma + \kappa}{\kappa^2 + \alpha} (\pi_{t-1} + \varepsilon_t^\pi), \tag{11}$$

where the intuitive interpretation is similar to (11). If this conditions holds, then the central bank will not be able to act in the optimal manner, hence aggregate demand will not be given by (10) but by (7). That is, the AD schedule will again become upward-sloping, if condition (11) is met and inflation is decreasing.

Figure 2 illustrates how the targeting model responds to a permanent negative demand shock at time *t=1* starting in the long-run equilibrium at *t=0*. The left panel shows the effect of a shock that is sufficiently small to be offset without the nominal interest rate falling to zero. The shock pushes the IS schedule to the left[12] but the central bank can pick a point on this schedule that fully offsets the effect of the shock on the position of the AD schedule.

[10] The setup where the central bank minimises the expected discounted stream of social welfare losses is dealt with algebraically in the next section.

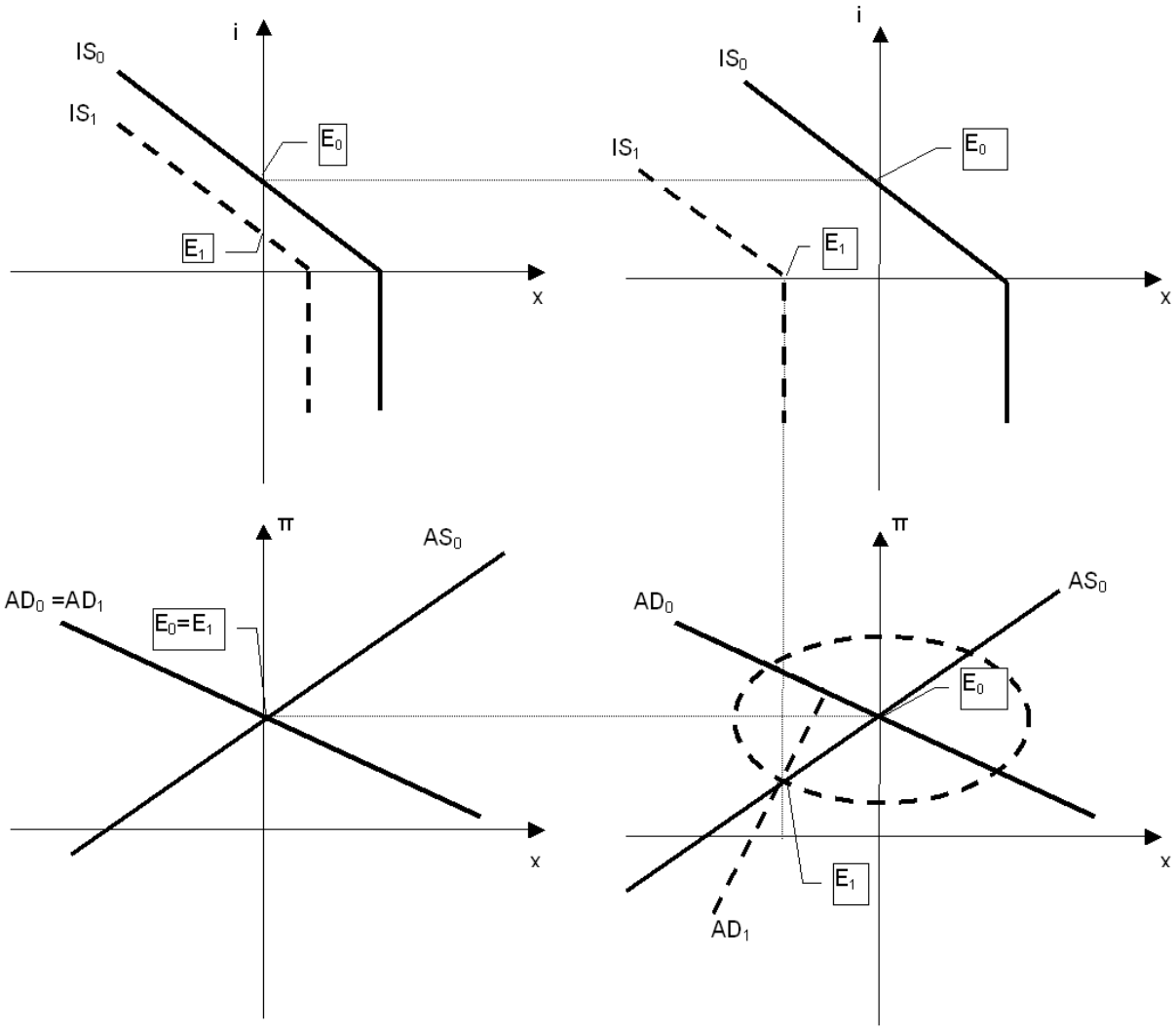
[11] Solving for the equilibrium output gap and inflation using (2) and (10) and plugging the result into (1), we can find the equilibrium interest rate as:

$$i_t = r^* - \frac{\kappa(1-\sigma\kappa)}{\sigma(\kappa^2 + \alpha)} \pi^* + \frac{\alpha\sigma + \kappa}{\sigma(\kappa^2 + \alpha)} (\pi_{t-1} + \varepsilon_t^\pi) + \frac{1}{\sigma} \varepsilon_t^x .$$

One can then set the nominal interest rate to zero and solve for the demand shock to obtain its value that would trigger the liquidity trap.

[12] The graph shows the final position of the IS schedule that also incorporates the change in equilibrium inflation.

Figure 2 Liquidity trap in the targeting model: small vs. large shock



The right panel of Figure 2 demonstrates what happens when picking the optimal point on the new IS schedule would imply an equilibrium nominal interest rate that is less than zero. Now that the central bank is bound by the zero interest rate constraint, it will operate on the upward-sloping portion of the AD schedule given by (7), as it is unable to achieve a bliss point with respect to the social welfare loss function. The ellipse centred on the long-run equilibrium point describes the social welfare loss that is thus incurred, assuming that $0<\alpha<1$, i.e. that the central bank cares more about stabilising inflation than output gap.[13] As in the right panel of Figure 1, once the equilibrium outcome happens on the upward-sloping portion of the AD schedule, the AS schedule will start shifting down, increasing the magnitude of the negative output gap, lowering inflation, and generating ever larger social welfare losses.

[13] For $\alpha=1$, the social welfare loss will be represented by a circle; for $\alpha>1$, it will be represented by an ellipse elongated along the vertical axis.

3. A case for pre-emptive monetary easing

In early 2008, the US Federal Reserve pursued an aggressive series of interest rate cuts that was unprecedented in its scope. At an emergency meeting on 22 January, the Federal Open Market Committee (FOMC) made the first 0.75% cut in the target federal funds rate (FFR), the largest single-day decrease since the early 1980s. This action came a week before the regularly scheduled meeting on 30 January, when the FOMC cut the FFR target by another 0.5%. On 18 March, the FOMC decided to further the cut by another 0.75% and followed up on 30 April, reducing the target rate by another 0.25% to 2%. The last two actions generated some disagreement within the FOMC, with Richard Fisher (President of the Dallas Fed) and Charles Plosser (President of the Philadelphia Fed) voting against the measures both times. The FOMC press releases acknowledged inflationary risks and the decisions drew heavy criticism from a number of prominent monetary economists. For instance, Rogoff (2008) described the United States as 'ground zero for global inflation'. The series of interest rate cuts undertaken by the Fed then took a pause until the fall of 2008.

Why did the Fed undertake this bold policy action? Although, given the extent of the financial crisis that unfolded in the fall of 2008 with short-term interest rates hitting zero in December, it may seem that the Fed was insufficiently aggressive it its cuts, in real time, its decision-making seemed extraordinarily expansionary. With some measures of inflation, such as the Consumer Price Index, registering annualised increases in excess of 4%, the highest level in about two decades, the Fed's focus on stimulating the real side of the economy was far from uncontroversial. This section attempts to provide a graphical apparatus that should facilitate the explanation of possible alternatives to an advanced undergraduate audience.

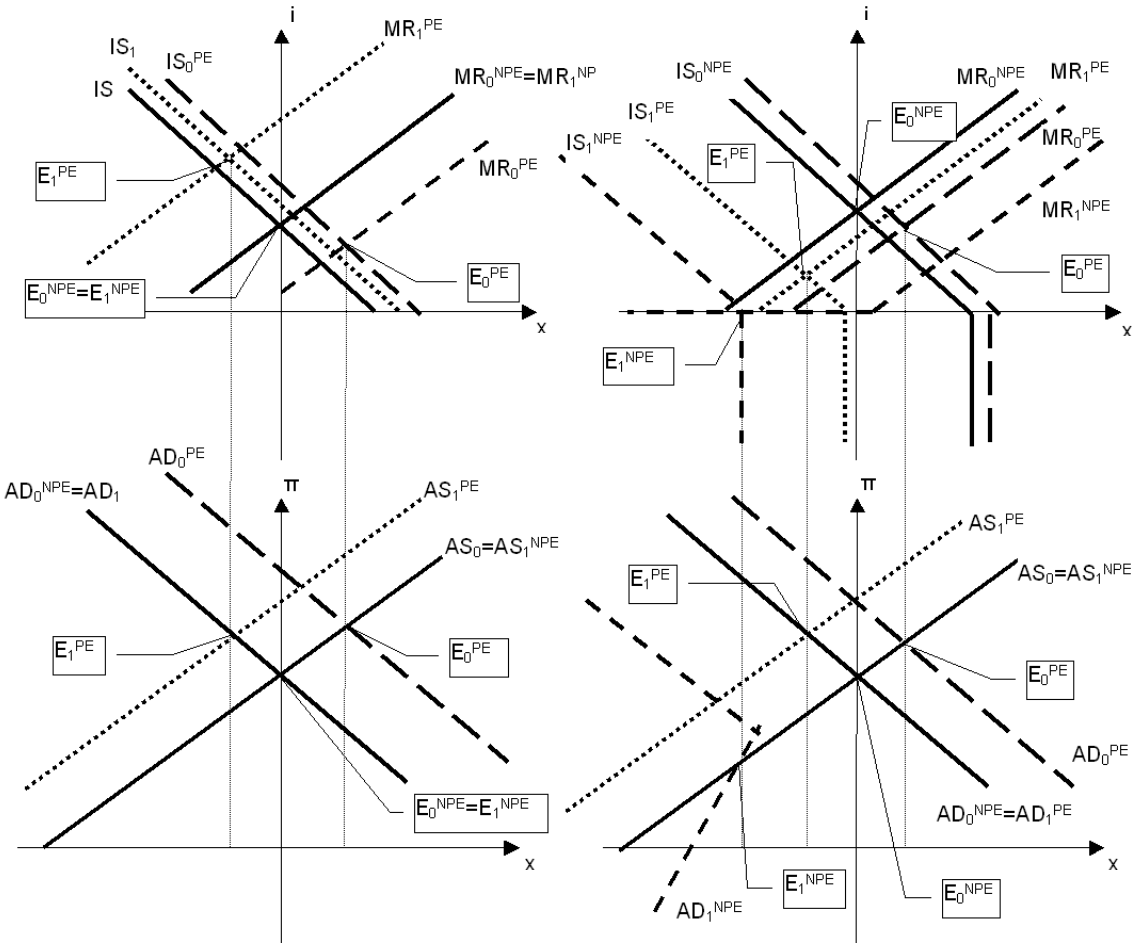
The Appendix lays out the theoretical motivation for pre-emptive monetary easing. Instead of myopically following the optimisation problem described by (9), the central bank optimises the discounted stream of social welfare losses over the infinite horizon, recognising the effect of adaptive inflationary expectations on future outcomes. Although dynamic optimisation may be accessible to some students, the technical treatment of this problem is considerably more complex than the simple framework discussed above. Graphical treatment should be more than sufficient to illustrate how engaging in monetary expansion *prior* to an anticipated negative demand shock can improve macroeconomic outcomes. As before, the following two sub-sections discuss this scenario first in the instrument version of the model and then in the targeting one.

Pre-emptive easing in the instrument model

Suppose that the central bank receives information that, in the next time period, there is a high probability that a large negative – and sufficiently persistent – demand shock will hit the economy. Given the discussion in the previous section, it should be clear that, if any monetary policy action is delayed until the shock hits, the economy will enter into a protracted recession, with no self-correcting mechanism to return it to the long-run equilibrium output gap of zero. If the central bank ordinarily sets the nominal interest rate using a Taylor rule (3), it may want to avoid this situation by pre-emptively applying a negative exogenous monetary shock, $\varepsilon_0^i < 0$, before $\varepsilon_1^x < 0$ hits the economy in the subsequent time period. That is, given the prevailing conditions at *t=0*, the central bank will lower the nominal interest rate below the level suggested by the systematic component of the Taylor rule.

Figure 3 details this process: in the left panel, $\varepsilon_1^x < 0$ does not materialise, whereas in the right panel it does. Pre-emptive easing accomplishes a downward shift of the MR schedule, from MR_0^{NPE} – the

Figure 3 Pre-emptive easing in IM: (left) shock doesn't materialise; (right) shock does materialise



benchmark of no pre-emptive easing – to MR_0^{PE}.[14] The lower interest rate stimulates aggregate demand that shifts from AD_0^{NPE} to AD_0^{PE}. The higher equilibrium inflation will ensure that at *t=1* aggregate supply shifts up due to higher adaptive inflationary expectations. Since the shock does not materialise in the left panel, the central bank will have to generate a monetary contraction at *t=1*, shifting the MR schedule to MR_1^{PE}. Relative to no pre-emptive action where the economy would be at rest in the long-run equilibrium, this will generate a recession to make inflation return to the target level. Furthermore, note that the central bank can accelerate the process of inflation returning to its long-run level by raising the nominal interest to a level exceeding that given by (10), if it wants to avoid, say, the loss of credibility associated with higher inflation.

The right panel of Figure 3 explains the difference between pre-emptive easing and strictly following (3) if $\varepsilon_1^x < 0$ does materialise. In the absence of pre-emptive easing, the economy falls into the liquidity trap and the subsequent dynamics follow as per the discussion that accompanied the right panel of Figure 1. With pre-emptive easing, however, the central bank has generated an inflationary cushion that allows for the nominal interest rate to stay positive, giving the monetary authority room for additional action. For simplicity, the figure assumes that under pre-emptive easing the central bank offsets the demand shock, so that the aggregate demand schedule returns to the long-run position.

[14] The final position of MR_0^{PE} accounts for the fact the pre-emptive easing gives rise to higher equilibrium inflation at *t=0* due to the upward-sloping AS schedule.

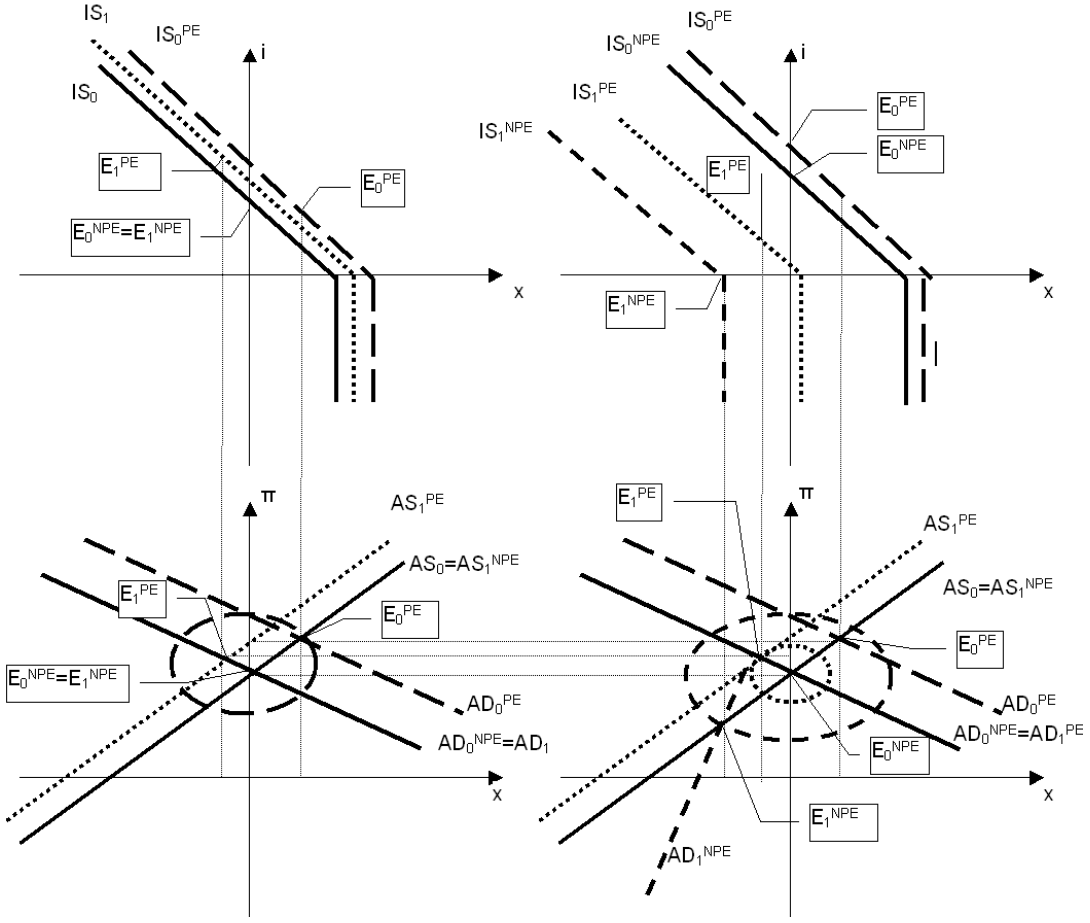
Pre-emptive easing in the targeting model

This sub-section builds on the analysis carried out above and considers the possibility of departure from the optimal leaning-against-the-wind rule (10) when the central bank anticipates a large negative demand shock in the subsequent time period. In that case, the central bank may want to choose to lower the nominal interest rate pre-emptively, pushing out aggregate demand and generating a larger inflationary cushion that will make the onset of a liquidity trap in the next time period less likely.

The left panel of Figure 4 shows the effect of this pre-emptive easing, if the negative shock fails to materialise at *t=1*. The central bank lowers the nominal interest rate at *t=0*, which generates inflation above the long-run target level and pushes the IS schedule up to IS_0^{PE}. Higher inflation at *t=0* will trigger higher inflationary expectations at *t=1*, shifting aggregate supply to AS_1^{PE}. (As before, in the absence of pre-emptive easing, the AS schedule would have remained at its original position.) Since the shock actually does not materialise, the AD schedule returns back to the original position at *t=1*. The central bank now leans against the inflationary wind and generates a negative output gap to bring inflation down to the level given by its long-run target.

The right panel of Figure 4 describes the effect of pre-emptive easing if the shock does materialise. In the absence of pre-emptive easing, the zero bound on the nominal interest rate becomes effective in *t=1*, which generates a relatively large social welfare loss in that time period and, if the shock is sufficiently persistent, more – and potentially larger – social welfare losses in the subsequent time periods. With pre-emptive easing, however, the central bank has enough leeway to use the nominal interest rate to offset the effect of the negative demand shock, possibly returning the AD schedule to its long-run position, with a much smaller social welfare loss.

Figure 4 Pre-emptive easing in TM: (left) shock doesn't materialise; (right) shock does materialise



4. Conclusion

This paper has developed the graphical apparatus for studying the effects of the zero nominal interest rate bound in the now standard model of short-run fluctuations with adaptive inflationary expectations. It has formally derived the result in the context of this model that an optimising central bank should engage in pre-emptive easing if the zero bound may be effective in the future. Furthermore, it has related its main findings to the recent policy-making challenges faced by the US Federal Reserve.

Of course, the model studied here is intentionally simplified, so that it can be captured graphically in an accessible fashion. In so doing, it has abstracted from important issues, such as the role of the forward-looking behaviour and the loss of inflation-fighting credibility that engagement in pre-emptive easing may bring. Exploring this agenda is an ongoing effort that will undoubtedly intensify in the near future.

References

Adam, K. and Billi, R. (2006). 'Optimal Monetary Policy under Commitment with a Zero Bound on Nominal Interest Rates', *Journal of Money, Credit, and Banking*, vol. 38(7), pp. 1877–1905.

Adam, K. and Billi, R. (2007). 'Discretionary Monetary Policy and the Zero Lower Bound on Nominal Interest Rates', *Journal of Monetary Economics*, vol. 54(3), pp. 728–52.

Ball, L. (1999). 'Efficient Rules for Monetary Policy', *International Finance*, vol. 2, pp. 63–83.

Billi, R. (2009). 'Optimal Inflation Rate for the US Economy', *Federal Reserve Bank of Kansas City Working Paper 07-03*.

Billi, R. and Kahn, G. (2008). 'What is the Optimal Inflation Rate?', *Federal Reserve Bank of Kansas City Economic Review*, 2nd Quarter.

Bofinger, P., Mayer, E. and Wollmershäuser, T. (2006). 'The BMW Model: A New Framework for Teaching Monetary Economics', *Journal of Economic Education*, vol. 37(1), pp. 98–117.

Carlin, W. and Soskice, D. (2005). 'The 3-Equation New Keynesian Model – A Graphical Exposition', *Contributions to Macroeconomics*, vol. 5(1), article 13.

Carlin, W. and Soskice, D. (2006). *Macroeconomics: Imperfections, Institutions, and Policies*, Oxford: Oxford University Press.

Cho, S. and Moreno, A. (2006). 'A Small-Sample Study of the New-Keynesian Macro Model', *Journal of Money, Credit, and Banking*, vol. 38, pp. 1461–81.

Clarida, R., Gali, J. and Gertler, M. (1999). 'The Science of Monetary Policy: A New Keynesian Perspective', *Journal of Economic Literature*, vol. 37(4), 1661–1707.

Clarida, R., Gali, J. and Gertler, M. (2000). 'Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory', *Quarterly Journal of Economics*, vol. 115(1), pp. 147–80.

Eggertsson, G. B. (2008). 'Was the New Deal Contractionary?', *Federal Reserve Bank of New York Working Paper*.

Eggertsson, G. and Woodford, M. (2003). 'The Zero Bound on Interest Rates and Optimal Monetary Policy', *Brookings Papers on Economic Activity*, vol. 34, 139–235.

Guest, R. (2003). 'Modifying the Taylor-Romer Model of Macroeconomic Stabilisation for Teaching Purposes', *International Review of Economics Education*, vol. 2(1), pp. 58–68.

Jeanne, O. and Svensson, L. (2007). 'Credible Commitment to Optimal Escape from a Liquidity Trap: The Role of the Balance Sheet of an Independent Central Bank', *American Economic Review*, vol 97, pp. 474–90.

Jones, C. (2007). *Macroeconomics,* New York: W.W.Norton.

Kapinos, P. (2010). "A New Keynesian Workbook," *International Review of Economics Education,* 9(1), 111-123.

Kato, R. and Nishiyama, S-I (2005). 'Optimal Monetary Policy When Interest Rates Are Bounded at Zero', *Journal of Economic Dynamics and Control*, vol. 29, 97–133.

Rabanal, P. and Rubio-Ramirez, J. (2005). 'Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach', *Journal of Monetary Economics*, vol. 52, pp. 1151–66.

Rogoff, K. (2008). 'An Inflation Reality Check', http://www.project-syndicate.org/commentary/rogoff40.

Romer, D. (2000). 'Keynesian Macroeconomics without the LM Curve', *Journal of Economic Perspectives*, vol. 14(2), 149–69.

Svensson, L. (1997). 'Inflation Forecast Targeting: Implementing and Monitoring Inflation Rates', *European Economic Review*, vol. 41, pp. 1111–46.

Svensson, L. (2002). 'Inflation Targeting: Should It Be Modeled as an Instrument Rule or a Targeting Rule?', *European Economic Review*, vol. 46, 771–80.

Svensson, L. (2003a). 'Escaping from a Liquidity Trap and Deflation: The Foolproof Way and Others', *Journal of Economic Perspectives*, vol. 17(4), pp. 145–66.

Svensson, L. (2003b). 'What Is Wrong with Taylor Rules? Using Judgment in Monetary Policy through Targeting Rules', *Journal of Economic Literature*, vol. 41, pp. 426–77.

Taylor, J. (1993). 'Discretion versus Policy Rules in Practice', *Carnegie-Rochester Conference Series on Public Policy*, vol. 39, 195–214.

Taylor, J. B. (2000). 'Teaching Modern Macroeconomics at the Principles Level', *American Economic Review Papers and Proceedings*, vol. 90(2), 90–4.

Turner, P. (2006). 'Teaching Undergraduate Macroeconomics with the Taylor-Romer Model', *International Review of Economics Education*, vol. 5(1), pp. 73–82.

Wiese, C. (2007). 'A Simple Wicksellian Macroeconomic Model', *The B. E. Journal of Macroeconomics (Topics)*, vol. 7(1), article 11.

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, Princeton, NJ.

Appendix: Theoretical motivation for pre-emptive easing

This appendix derives the theoretical motivation for pre-emptive easing along the lines of Kato and Nishiyama (2005) who also work with a backward-looking model following Ball (1999) and Svensson (1997). The model in this paper has a simpler lag structure, which allows us to arrive at the key result in a very accessible analytical fashion. Since inflationary expectations are backward-looking, this analysis is similar to the finding of Eggertsson and Woodford (2003) who show that with forward-looking inflationary expectations the liquidity trap can be avoided by promising to generate excess inflation, even after the effect of a persistent negative demand shock has passed.

Although minimising (9) provides a straightforward way to derive an aggregate demand relation that can be easily graphed, in practice, monetary authorities are concerned with the effect of their policy on future social welfare outcomes as well. Hence the more general formulation of the central bank's problem in the context of the model used in above is:

$$\min_{\pi_t, x_t, i_t} \frac{1}{2} E_t \sum_{s=t}^{\infty} \beta^{s-t} \left\{ \left(\pi_s - \pi^*\right)^2 + \alpha x_s^2 \right\}, \tag{A1}$$

subject to (2), (1), and the nominal interest non-negativity constraint:

$$i_t \geq 0. \tag{A2}$$

The Lagrangian associated with this problem is:

$$L = \frac{1}{2} E_t \sum_{s=t}^{\infty} \beta^{s-t} \left\{ \left(\pi_s - \pi^*\right)^2 + \alpha x_s^2 - \lambda_s^{AS} [\pi_{s-1} + \kappa x_s + \varepsilon_s^{\pi} - \pi_s] - \lambda_s^{IS} [\sigma r^* - \sigma(is - \pi s) + \varepsilon_s^{x} - x_s] - \lambda_s^{i} i_s \right\}$$

As Kato and Nishiyama (2005) emphasise, it is important to ensure that the Lagrangian multipliers are positive, because they will enter the central bank's optimal policy rule, signalling the direction of its adjustment relative to the case where (A2) is not binding. The first-order conditions with respect to π, x and i are:

$$\begin{aligned} &\pi_t - \pi^* + \lambda_t^{AS} - \sigma\lambda_t^{IS} - \beta E_t \lambda_{t+1}^{AS} = 0, \\ &\alpha x_t - \kappa\lambda_t^{AS} + \lambda_t^{IS} = 0, \\ &\sigma\lambda_t^{IS} - \sigma\lambda_t^{i} = 0. \end{aligned} \tag{A3}$$

If (13) does not bind, then $\lambda_t^i = \lambda_t^{IS} = 0$ and $\lambda_t^{AS} = \frac{\alpha}{\kappa} x_t$. Hence for a positive equilibrium interest rate, it is only aggregate supply that may impose a constraint on the conduct of monetary policy.

To see why pre-emptive monetary easing is optimal in this setting, suppose that initially at time period t the central bank does have room to manoeuvre and (13) does not bind but starting with $t+1$ it may. Aggregate demand at time period t becomes:

$$\pi_t = \pi^* - \frac{\alpha}{\kappa} x_t + \beta E_t \lambda_{t+1}^{AS}, \tag{A4}$$

or,

$$\pi_t = \pi^* - \frac{\alpha}{\kappa} x_t - E_t \sum_{s=t+1}^{\infty} \beta^{s-t} \left(\pi_s - \pi^*\right) + E_t \sum_{s=t+1}^{\infty} \beta^{s-t} \lambda_s^{i}. \tag{A5}$$

Equation (A4) draws an intuitive contrast with (10): the latter is static and easier to graph but does not account for the fact that today's inflationary outcome will affect future welfare through adaptive inflationary expectations. The other representation, equation (A5), emphasises the role of the zero bound on the nominal interest rate. Insofar as it can become effective in the future and the Lagrangian multipliers associated with it, λ_t^i, can be positive, it is optimal for the current inflation rate to be higher than it would have been in the absence of that possibility. This can be achieved through monetary easing in the current time period.

Similar intuition can be gleaned for the instrument model. Although Taylor rules like (3) are imposed on the model exogenously, one could combine (A5) with (1) to back out the optimal interest rate response function:

$$i_t = r^* + \pi_t + \frac{1}{\sigma}\varepsilon_t^x + \frac{\kappa}{\alpha\sigma}(\pi_t - \pi^*) + \frac{1}{\sigma}E_t\sum_{s=t+1}^{\infty}\beta^{s-t}\left(\pi_s - \pi^*\right) - \frac{1}{\sigma}E_t\sum_{s=t+1}^{\infty}\beta^{s-t}\lambda_s^i. \qquad \text{(A6)}$$

This equation has a number of intuitive implications. The Taylor principle is satisfied given (8) and $0<\alpha<1$, i.e. the optimal response function will be stabilising under these conditions. Anticipation of future inflation in excess of the target level suggests monetary tightening now, as the central bank recognises that current lower inflation will moderate its future levels through adaptive expectations. Finally, the possibility that the zero nominal interest rate bound may become effective in the future calls for pre-emptive monetary easing now.

Author Biography

Pavel Kapinos is an assistant professor of economics at Carleton College. He received his Ph.D. in economics from the University of Illinois at Urbana-Champaign in 2004. His primary area of expertise is applied and theoretical monetary economics, which he studies in the context of New Keynesian dynamic stochastic general equilibrium models.

Contact details

Pavel Kapinos
Carleton College
1 N College Street, Northfield, Minnesota 55057, USA
Tel: +1-507-222-7676
Fax: +1-507-222-4049
Email: pkapinos@carleton.edu

Teaching Innovations in Economics: Strategies and Innovations for Interactive Instruction

Edited by Michael K. Salemi and William B. Walstad. Edward Elgar Publishing Limited, 2010; 274 pages; £75
Reviewed by Charles A. Stull

Teaching Innovations in Economics is one result of a series of workshops, online training, and support for scholarly work organised by two economists prominent in the field of economic education, Michael Salemi and William Walstad. The 30 contributors to this volume were active in a National Science Foundation funded Teaching Innovations Program, either as workshop organisers or participants. The workshop programme was intended to encourage economists to investigate, adopt and disseminate new teaching techniques. The book continues this mission.

The book begins with a lengthy description of the Teaching Innovations Program, with details on its workshop curriculum, the use of an online system to maintain contact with participants as they implemented innovations in their own classes, and a description of how participants were encouraged to engage in the scholarship of teaching and learning in economics. These initial chapters are probably most useful to readers who plan to develop a similar workshop programme. The later chapters have two natural audiences: teachers of economics and economists interested in initiating research on teaching pedagogy.

The core of the book focuses on innovative teaching methods that could be incorporated into economics courses: cooperative learning, experiments, discussion, formative assessment, context-rich problems, cases and interaction for large classes. Each technique is introduced by a workshop organiser. Three instructors then provide descriptions of how they employed the technique in their courses. The examples are drawn from a variety of undergraduate classes, including principles of economics, intermediate theory and field courses, and a variety of types of learning environments such as liberal arts colleges, large universities, community colleges and distance learning courses. A number of the examples could be directly implemented by many instructors, but the book's main emphasis is broader. More than a manual of specific activities and exercises, this volume provides an agenda for developing innovative methods by any teaching economist.

Not all chapters accomplish this goal equally well. For example, the chapters on cooperative learning and context-rich problems provide particularly clear motivation for using those methods, with many examples of the techniques, and then present novel applications by the workshop alumni. On the other hand, the chapter on experiments merely asserts the value of using experiments and then focuses on the experience of the instructors in adapting existing experiments to their own classes. All the teaching methods chapters provide cases of successful use of innovations in economics classes but some chapters provide more detailed descriptions so that the reader can create new applications of these teaching methods. For example, the chapter on 'Case Use in Economics Instruction' starts by defining

the case method and identifying five characteristics of a 'star-quality' case. It then discusses how cases can encourage higher order learning in economics. The chapter describes, in detail, four cases in which students are asked to make a decision based on their analysis of information provided. These range from a relatively simple case where a hypothetical student needs to choose between a cash payment and a valuable baseball card, to a more complex case of market analysis for a prospective local coffee shop, to a substantial multi-part case on Zimbabwe's hyperinflation. The authors discuss the learning goals for each case and their experience of using it in class. Appendices provide copies of student handouts for three of the four cases. This detailed approach allows instructors to use a specific case, as is, or to use it as a model for designing their own cases.

The book may also be useful for economists interested in engaging in scholarship on teaching and learning. The introductory chapters use an inclusive definition of scholarship in this area – the concept of the 'teaching commons' which includes both formal research leading to journal publication and less-formal sharing of innovative teaching methods and materials. While the book extensively models the dissemination of instructional strategies, some chapters also include quantitative measures on learning outcomes. The volume references formal presentations and publications that resulted from the project. The references listed at the end of each chapter provide a valuable introduction to the academic literature on each teaching strategy.

The book closes with an assessment of the Teaching Innovation Program based on workshop evaluations, a retrospective survey, and participation in post-workshop training and scholarly activities. The vast majority of participants (85–95%) responded very positively to the programme, reflecting an interest in applying active learning strategies within the profession. While this book is not a substitute for attending a programme such as the Teaching Innovation Program workshops, it does illustrate the success of pedagogic innovations across a broad range of courses and gives economists inspiration to try an alternative to the traditional chalk-and-talk lecture.

Contact details

Charles Stull
Department of Economics and Business
Kalamazoo College
1200 Academy Street
Kalamazoo, MI 49006
USA

Email: cstull@kzoo.edu

Famous Figures and Diagrams in Economics

Edited by Mark Blaug and Peter Lloyd. Edward Elgar Publishing Limited, 2010; 468 pages; £120
Reviewed by Patricia Hughes

Famous Figures and Diagrams in Economics describes the role of our most important visual tools in the development of economic theory. The goal of the book is to provide an account of each of the most prominent diagrams in economic analysis along with the history of each discovery. Each topic is presented by an expert in the field, providing a variety of approaches emphasising the origins, controversies, graphical representations, applications and extensions of each subject. The coverage of topics is exhaustive including basic demand and supply analysis, welfare economics, general equilibrium analysis, open economies, macroeconomic analysis and stabilisation, growth theory, and specialty topic areas. All told, 58 topics are described in 468 pages.

'Figures are part of the basic toolbox of the modern economist' (p.5). Diagrams are used in investigation and exploration leading to discovery. Diagrams illustrate a discovery arrived at by other means: observation, intuition, logic, mathematics. This book provides the history, development, and controversies that surround the most famous diagrams in economics. The approach highlights the complexities of the research, the beauty of the graph in illustrating the idea and enlightening the student, and the extensions in the field using more sophisticated mathematical techniques. While the graphs are staples in all economic textbooks, understanding the development of the model reveals the intricacies and complexities of the ideas behind the graphs.

The book opens with the most widely used model in economics, the demand-and-supply or Marshallian cross diagram. From today's perspective, we tend to see the axes as reversed in terms of the dependent and independent variables. Originally Marshall viewed the market in terms of quantity adjustments to discrepancies between demand price and supply price, yielding the current diagram with price as the dependent variable. The Marshallian cross is used to illustrate different perspectives on market stability, and the debate between Marshall and Walras on market disequilibrium adjustment.

A perfect illustration of the complexities of theory often overlooked is contained in the chapter on long-run and short-run cost curves. The chapter introduces the notion of costs curves and the assumptions made in developing the standard set of short-run and long-run average cost curves. As an insight into the development of cost theory, the authors trace an interesting dilemma in reconciling the short-run and long-run average cost curves, and confusion about the respective equilibria. Intuitively it was thought that the LRAC curve should coincide with the minimum points on the various SRAC curves, since these are all long-run equilibrium points. As it turns out, mathematically/graphically it is impossible to draw the LRAC to connect these points and be forever below the SRAC curves. The confusion arises with the definition of long-run equilibrium and the assumptions imposed on reaching that equilibrium.

The book also illustrates the lighter side of graphical origins, as with the Laffer curve, which emerged from the interactions of 'a Californian professor of business economics (Arthur Laffer), two presidential aides, a Washington restaurant and a cocktail napkin' (p.412). While a number of economists had presented a similar proposition for various kinds of taxes, Laffer is given credit as the populariser with the first, hand-drawn diagram.

The chapters vary in their emphasis, but there is a common approach of introducing the 'famous' diagram in a historical context, summarising the applications of the tool in the context of economic debates or paradoxes, providing extensions of the basic graph, and outlining advancements made in the algebraic analysis. The chapters are brief by design, yet provide an amazing amount of relatively advanced material across multiple subjects. Such an approach does require the reader to be well schooled in economics to appreciate the progression of theory through graphs and diagrams. The text assumes a relatively high level of knowledge, and for those individuals the book provides a wealth of information as to the historical developments and seminal works in the field. For those that are unacquainted with a particular topic, the graphs may seem overly complicated and the text incomprehensible.

This book will be a wonderful reference manual for those well versed in economic theory. It can be used by instructors as a supplement to contemporary textbook material, which tends to be presented in a clear, concise, targeted fashion. In a history of economic thought course or any theory course, this text can provide a richness of material to expand and contextualise the presentation. As texts become more mainstream, they tend to exclude or minimise the original works that are then lost to the next generation of students. While it is conjectured that 'Mordecai Ezekiel's 1938 paper made "The Cobweb Theorem" and his famous diagram well-know to every student of economics' (p.184), I suspect this is no longer the case.

Contact details

Patricia Hughes,
Economics Department,
St. Cloud State University,
386 Stewart Hall,
720 Fourth Avenue South,
St. Cloud, MN 56301-4498

Email: pahughes@stcloudstate.edu