ASSESSING STUDENT UNDERSTANDING OF PRICE AND OPPORTUNITY COST THROUGH A HYBRID TEST INSTRUMENT: AN EXPLORATORY STUDY

Kristen Collett-Schmitt, University of Notre Dame Ross Guest, Griffith University, Australia Peter Davies, University of Birmingham, UK

ABSTRACT

This paper reports an exploratory empirical study of students' 'understanding' of price and opportunity cost based on a hybrid test instrument which consisted of multiple choice (MC) questions that also required students to explain their answers through a constructed response (CR). Data were drawn from three cohorts of students: two from an Australian university and one from a U.S. university. The information provided by the CRs attached to the MC questions tends to support extant evidence of problems with the effectiveness of standard MC questions in assessing 'understanding,' especially of price. Additional econometric analysis suggests that MC and associated CR questions may not be as effective at capturing achievement of sophisticated conceptions or 'understanding,' compared with alternative assessment tasks such as problems and essays.

Keywords: price, opportunity cost, assessment, multiple choice, constructed response JEL codes: A20, A22, A29

Students' 'understanding' of opportunity cost and price has been widely regarded as central to economic education (e.g. Sevón & Weckström, 1989; Leiser & Halachmi, 2006; Salemi, 2005; Davies & Lundholm, 2012). However, as Davies (2011) points out, evidence from assessments of conceptions of price and opportunity cost leaves a great deal of uncertainty about what it means for a student to 'understand' price or opportunity cost.– This implies a problem for university lecturers evaluating test items on the economic concepts, and for high school teachers using similar test items to gauge students' progress and readiness to proceed to the next level. This is also a problem for researchers trying to judge the effectiveness of different forms of intervention in teaching.

In this paper, we aim to shed more light on the nature of these problems faced by economic educators and researchers, and in doing so, pilot a new approach to the design and analysis of assessment of economic 'understanding.' Our method is to examine students' answers to multiple choice questions, which are commonly used on tests to assess economic 'understanding,' alongside their accounts of why they considered a particular option to be correct. We gather evidence of students' conception of price and opportunity cost as expressed in these constructed responses. We also analyze students' multiple choice answers and the quality of their constructed responses as determinants of their overall achievement in the course.¹ . Our exploratory methodology reveals variation in students' 'understanding' of opportunity cost and price, which is missed by conventional approaches.

'UNDERSTANDING' PRICE AND OPPORTUNITY COST: CURRENT EVIDENCE

While Bloom's taxonomy suggests a distinction between 'understanding' and 'applying' a concept (Bloom, 1956), the research on students' 'understanding' of price does not (e.g. Thompson & Siegler, 2000; Meyer & Shanahan, 2002; Pang & Marton, 2003, 2005; Leiser & Halachmi, 2006). Regardless of whether this research has been undertaken in the social psychology, phenomenographic or another tradition, it has defined 'understanding' as *necessarily* including 'application' and based on experiences that are familiar to students. Using in-depth interviews with students, researchers have suggested four different conceptions of price which are evident in the way students talk or write about their experiences (Pang & Marton, 2005):

- 1. Price as a reflection of a good's intrinsic quality
- 2. Price as a reflection of cost of supply
- 3. Price as a reflection of consumers' willingness to demand
- 4. Price as a reflection of demand and supply

However, it would be premature to regard this fourfold categorization of conceptions of price as anything more than provisional. For example, Pang & Marton (2005) suggest a fifth category –price as a reflection of relative magnitudes of changes in supply and demand. Further, Davies (2011) proposes that this category be subdivided according to the context used to frame the conception (individual producer, market, or interactions between markets) and whether causation runs only from supply or demand to price or if causation from price to supply and demand is also recognized.

There has been less research on students' 'understanding' of opportunity cost. The available evidence is largely based on students' answers to multiple choice questions. Ferraro and Taylor (2005) devised a multiple choice question using the example of buying a concert ticket. By placing the question in the context of an experience which is likely to be familiar, they address the way in which the concept of opportunity cost is 'understood' in everyday experiences. However, this stance is compromised by the actual question which is, 'what is the opportunity cost?' That is, the question turns out to be asking for accuracy in a technical definition rather than seeking to expose the way that respondents understand opportunity cost in their everyday experiences. They report that only 21.6% of those attendees at an AEA conference who were asked the question opted for the correct answer. O'Donnell (2009) used the same question as Ferraro & Taylor (2005) with academics and students at his institution. He reports that the proportion of academics answering the question correctly increased when the sentence referring to opportunity cost was omitted, though his sample sizes were small. Whilst Ferraro & Taylor (2005) conclude that academic economists have a weak grasp of a fundamental concept, O'Donnell (2009) concludes that opportunity cost cannot be fundamental if such a high proportion of academics do not understand it. An alternative explanation is that this is simply a poorly-performing question, as judged by the standards usually employed in evaluating multiple choice questions. Regardless, this research implies that language such as 'better understanding' may be fit for the practice of teaching and learning theory rather than language of, 'does/does not understand.'

Greater integration of ideas has also been identified as a critical feature of 'understanding' within an academic domain (Alexander, 2000). Someone who can correctly answer multiple choice questions of the form, 'Which of these is the opportunity cost?' while

failing to use the idea of opportunity cost when asked about the movement of firms between markets, the benefits of international trade, or the impact of a tax on sales of a product, must have a trivial 'understanding' of the idea. This point has been central to the research on threshold concepts (Davies, 2012). For example, Davies & Mangan (2007) suggest that when a threshold concept of 'interaction between markets' is embedded in students' thinking, they are able to use this as an organizing idea when making sense of pricing problems. According to this proposition, it is therefore important to distinguish between students who frame their thinking about price solely in terms of individual producer decisions, students who frame their thinking about price solely in terms of behavior within an individual market, and students who frame their thinking about price in terms of interactions between markets. Likewise, it is important to distinguish between students who 'understand' opportunity cost as a way of thinking about individual decisions and students who 'understand' opportunity cost as a property of an economic system which underpins differences in value (as originally suggested by Von Wieser in 1891).

THE MC-CR DEBATE AND BEYOND

Our methodology in this paper bears heavily not just upon the aforementioned literature on student 'understanding, but also on the - literature on the relative efficacy of multiple choice (MC) and constructed response (CR) questions in assessing learning outcomes. MC questions have some clear advantages in that they are free from bias in grading and enable a wide range of content to be examined for a given test duration (Walstad, 1987; Saunders & Walstad, 1990, as cited in Becker & Johnston, 1999). The argument for CR questions has been largely framed within Bloom's taxonomy of educational objectives. For example, Biggs (1999) argues that MC questions are ill-suited to measuring what Bloom's taxonomy calls 'higher cognitive skills' that involve synthesis and evaluation. This criticism has encouraged researchers to compare students' scores on MC and CR (particularly essay) questions to identify differences in how the two types of questions assess student 'understanding.'

A number of early studies from the educational measurement literature found a high positive correlation between scores on MC and CR questions, leading some to argue that it does not matter whether MC and CR questions are used (Bennett, Rock & Wang, 1991; Bridgeman & Rock, 1993; Thissen, Wainer & Wang, 1994; Lukhele et al., 1994). Becker and Johnston (1999) also found a high correlation between the two test scores in their sample. However, they argue that this result does not mean that MC and CR questions are testing the same dimensions of learning but that an unmeasured factor, such as ability, strongly determines performance on both types of questions. More recent evidence adds support for differences in MC and CR questions.

In the context of economics education, Hickson and Reid (2011) provide exhaustive econometric evidence that CR questions add information about learning not included in MC questions. Buckles and Siegfried (2006) find that MC questions are not as reliable as CR questions in assessing synthesis and evaluation in the field of economics. They argue that the chain of reasoning required for this level of learning is hard to assess in a single or even sequence of MC questions. Kennedy and Walstad (1997) found that CR questions on the 1991 AP tests for micro and macroeconomics provided additional value over MC questions in discriminating students.

This literature on conceptual 'understanding' in economics implies other problems with MC questions. If we want to distinguish between students according to the sophistication of their conception (say) of price in a given context, we could create a MC question in which each

possible response reflects a conception which researchers have identified. This, however, poses a huge challenge for item construction. For example, consider the following question:

Question stem: Which of the following is the best explanation of a rise in price given situation *X*?

- 1. A change in the intrinsic quality of the good being sold
- 2. An increase in demand but no reference to movement along the supply curve
- 3. A reduction in supply with no reference to movement along the demand curve
- 4. A reference to both a shift and movement along demand and supply curves.

This question could never be considered to be an effective test of learning outcomes. One problem with this question is that it is difficult to avoid options that students might reason to be correct. For example, a student might select option A. as the correct answer because this outcome increases demand and ultimately price. Additionally, a student might logically deduce option D. to be correct because it subsumes other options.

Another challenge of using MC questions is that they require questions to be framed such that there is only one correct answer. Economics in the real world is rarely that straightforward. Since assessment has powerful 'wash-back' effects on what students think they have to learn and on what lecturers feel obliged to teach (Biggs, 2002), multiple choice questions inevitably encourage the belief that economic problems are unambiguous and have absolute solutions.

Our study goes beyond this MC-CR debate in the literature by uncovering students' varied conceptions of opportunity cost and price using a hybrid test instrument that asks students to provide a CR to explain their answer to a MC question. We examine the text of the CRs in detail in order to infer the nature of students' 'understanding,' and then relate performance on both CR and MC components to their final achievement in the course .We find neither the MC nor CR questions are as reliable an indicator of final achievement for high GPA students as for low GPA students. This we argue is because high GPA students tend to achieve more sophisticated conceptions, which are better captured by other assessment tasks.

THE TEST INSTRUMENT AND DATA

Our exploratory test instrument consists of four multiple choice (MC) questions, which are given in the Appendix. Questions 2 and 3 were on opportunity cost (OC) and questions 1 and 4 were on market price as the outcome of supply and demand.² Students were also asked to provide a brief written explanation of the reasoning behind their answer to each question. We will refer to these accounts as the students' constructed responses (CRs), and to the selection of answers A, B, C or D as the students' MC answers. The instrument is therefore a hybrid of MC and CR questions, the unique feature of which is that the CR questions are linked to the MC questions.

In order to assess the change in student 'understanding,' students were asked to complete a pre and posttest. They were advised that performance on the test instrument would not impact their final mark in the course. The pretest was administered before any substantive teaching had occurred, and the posttest was administered after the two concepts of OC and market price (with basic supply and demand analysis) had been formally studied. Students were not provided with any answers to the test after the pretest version, and the posttest was exactly the same as the pretest. Importantly for the econometric analysis reported below, the test marks were not included in students' final grade for the course. The test was taken by two cohorts of students with the same instructor at one Australian university in 2011 and 2012 and one cohort of students at one university in the United States in 2012. After removing students who were absent from either of the tests, chose not to answer at least one question (including providing a CR), and did not have a GPA, the posttest sample size was 159 for the Australian cohorts (combined) and 188 for the U.S. cohort, while the pretest sample size was 309 for the Australian cohorts and 189 for the U.S.³

METHOD

Students' written explanations of their answer to the multiple choice questions were first categorised through qualitative analysis. In conducting this analysis, our aim was to avoid imposing a set of pre-ordained categories, although we are conscious that our familiarity with the research literature on conceptions of price is likely to have influenced our judgments regardless of our intentions. This process clearly identified the wide range of conceptions that students have of price and OC, with varying depths of 'understanding' of price in particular, which is consistent with previously-cited literature. We provide examples of our categorizations in the Appendix.

One question we wish to address is whether students' answers to these multiple choice questions provide reliable evidence about underlying conceptions. To answer this question, we cross-tabulated students' answers to the multiple choice questions with the categorizations of conceptions suggested by students' CR responses. The coding into categories, the score for each category, some examples of responses, and the number of student responses in each category are given in the Appendix. The number of student responses includes both the pretest and posttest results. Although the response categories are not, in principle, mutually exclusive, no response was assigned to more than one category in order to facilitate analysis.

After the cross-tabulation of MC and CR scores, we analyzed the following relationships:

- 1. The correlation between the MC and CR scores for price and the same for OC.
- 2. The correlation between the two MC scores for price and the same for OC; and the correlation between the two CR scores for price and the same for OC.
- 3. The significance of the MC scores and CR scores on both the pretest and posttest as determinants of overall achievement in the course, controlling for the students' grade point average (GPA) at the start of the course.

DATA ANALYSIS AND RESULTS

If the MC questions are good indicators of the level of 'understanding,' we would expect the MC scores on the two price questions to be highly correlated and the MC scores on the two OC questions to be highly correlated. The same should apply to the CR scores. Since both types of questions evaluate the same broad concept, we might expect the MC scores and the CR scores to be highly correlated for both the price and the OC questions (Lukhele et al., 1994; Bridgeman & Rock, 1993). The results of these correlations are given in Table 1. The results in Table 1 apply only to the posttest sample in order to investigate the effects of 'understanding' after the concepts of price and opportunity cost were addressed in the course. Table 1 indicates that, while the correlation coefficients are positive (and statistically significant), the strength of the correlation for the price questions is much lower than for the OC questions. In particular, the

TABLE 1: Correlation coefficients between scores on test items (posttest correct answers)								
Test items	Correlation coefficient (phi)	Probability on χ^2 test						
The two MC scores on price	0.20	0.01						
The two MC scores on OC	0.54	0.00						
The two CR scores on price	0.33	0.00						
The two CR scores on OC	0.49	0.00						
The sum of MC scores and sum of CR	0.68	0.00						
scores on price								
The sum of MC scores and sum of CR	1.02^{*}	0.00						
scores on OC								
*Notes:								
(i) Unlike the Pearson correlation coefficient, the phi coefficient is generally not bound from -1 to +1.								
(ii) In order to ensure a suitable sample	size, NA is treated as a separ	rate category for all correlation						
coefficients.								

correlation between the sum of the two MC scores and the sum of the two CR scores for the price questions is much lower than the corresponding scores for the OC questions.

Given that the MC and CR scores are positively correlated, albeit more for the price questions, we then explored whether CR scores add value to MC scores – that is, whether CR scores are capturing learning outcomes beyond those captured by MC scores. As a preliminary exercise, we run a simple (OLS) regression, for all four questions combined on the post test, of CR scores on MC scores, GPA, and dummy variables to capture gender and institutional effects:

$$CRscore_{i} = \alpha_{0} + \alpha_{1}MCscore_{i} + \alpha_{2}GPA_{i} + \alpha_{3}DFA_{i} + \alpha_{4}DFUS_{i} + \alpha_{5}DMUS_{i} + \varepsilon_{i}$$
(1)
0.03 2.80 0.20 -0.07 0.11 -0.13 ,
(0.08) (15.06) (3.21) (-0.46) (0.63) (-0.78)

where *CRscore* and *MCscore* are the student's average scores on the two CR and MC items respectively, DFA is a dummy taking the value of 1 for female Australian students and zero otherwise, DFUS is a dummy taking the value of 1 for female U.S. students and zero otherwise, and DMUS is a dummy taking the value of 1 for male U.S. students and zero otherwise (hence the omitted category is male Australian students). The numbers in parentheses are t statistics. The inclusion of GPA in this regression is based on the assumption that higher GPA students tend to achieve higher-order learning outcomes; hence the GPA variable is meant to capture any higher-order learning outcomes in CR scores not captured by MC scores. We find that GPA is statistically significant and positive, suggesting that there may be information about the degree of sophistication in student conceptions of price and OC that is reflected in the CR scores but not captured by the MC scores. Gender and institutional effects were not statistically significant.

More detailed analysis of the CR and MC responses sheds further light on the type of information captured by CR scores. Consider the two questions on price (Questions 1 and 4). In Question 1 for example, of the students who chose the correct MC response (B), 23% mentioned demand without mentioning supply and 25% mentioned supply without mentioning demand, even though the correct MC response implicitly refers to the importance of both supply and demand. For the other question on price (Question 4), of the students who chose the correct MC response the correct MC response to the importance of both supply and demand. For the other question on price (Question 4), of the students who chose the correct MC response (C), 30% had entirely fallacious reasoning. This is higher than the 25% of students

expected to choose the right MC response simply by guessing. What is more concerning is the fact that 13% of students who chose the wrong MC response for Question 4 gave a high quality CR, which was in the same category as the explanations given by the 70% of students who chose the correct MC response. These results suggest that the MC options are not properly discriminating the quality of 'understanding' of the concept of price. The OC questions (Questions 2 and 3) also reveal more reliability than the questions on price. For Question 2, 86% of students who chose the correct MC response also gave the highest quality CR. For Question 3, the corresponding figure was 75%. In both OC questions, less than 5% of students who chose the wrong MC response gave the best CR.

A comparison of pre and posttest responses also indicates a problem with MC questions as an indicator of learning for the concept of price. For Question 1, 8% more students chose the correct MC response on the posttest than on the pretest. For Question 4, the corresponding figure was 7%. This is a low improvement rate. The improvement rate is better for the CRs, especially for Question 1. As evidence, for each of Questions 1 and 4, only half the proportion of students gave an irrelevant or nonsensical CR answer on the posttest, compared with the pretest (5% compared with 10%). For Question 1, 13% more students gave the highest quality CR on the posttest test than on the pretest; for Question 4, however, the corresponding figure was only 2%. This suggests that assessing learning of the concept of price is complicated and difficult even using CR questions.

The improvement rate was much better for the OC questions. For Question 2, 19% more students chose the correct MC response on the posttest than on the pretest. For Question 3, the corresponding figure was 37%. And for both Questions 2 and 3, of the students who chose the correct MC answer, roughly the same proportion provided the best CR on the posttest as on the pretest. This means that the relatively strong improvement in performance on the OC questions was reflected in both the CRs and MC answers.

These results suggest that these MC questions do not pick up the wide range of 'understandings' of the concept of price that are held by all students. The MC questions are also not picking up the change in 'understanding' from the pretest to the posttest. Further, assessment of 'understanding' appears to be more difficult for the concept of price than for opportunity cost, regardless of the type of question used. However, since this is an exploratory study with a small number of items, we cannot discount the possibility that the results are peculiar to these particular questions.

Next we investigate the significance of the MC scores and CR scores as determinants of overall achievement in the course, for both the pretest and posttest. Here, we do not distinguish between the price and OC questions. Rather, we average the CR scores for all four questions and do the same for the MC scores. A student's final mark in the course is the indicator of overall achievement. For both the U.S. and Australian students approximately 50% of the final mark was determined from MC questions (including the mid-term and final exams). The remaining 50% of the final mark was determined by problem- solving tasks and short essays (and, in the case of the U.S. students, a 5% class participation mark). Since a student's final mark is not dominated by the one assessment type, there is no obvious bias in the ability of either CR or MC questions on the test instrument to determine the final mark. Again, note that scores on the test instrument were not included in the final mark.

We start with a simple regression model for each question type:

$$Final_mark_i = \alpha_0 + \alpha_1 CRscore_i + \alpha_2 DFA_i + \alpha_3 DFUS_i + \alpha_4 DMUS_i + \varepsilon_i$$
(2)

$$Final_mark_{i} = \alpha_{0} + \alpha_{1}MCscore_{i} + \alpha_{2}DFA_{i} + \alpha_{3}DFUS_{i} + \alpha_{4}DMUS_{i} + \varepsilon_{i}$$
(3),

where *Final_mark* is the student's final mark⁴ for the course as a measure of overall achievement. There is a potential endogeneity problem due to measurement error in the right hand variables *CRscore* and *MCscore*, since these are measured at a time period prior to the measurement of the dependent variable *Final_mark* (Becker and Salemi 1977). However, we reject the null hypothesis of endogeneity applying a Hausman test to each equation, using the square of *CRscore* and *MCscore*, respectively, as instruments. This was also the case for the models that follow.

Table 2 presents the results for the posttest data and Table 3 includes results for the pretest data for the three cohorts combined. Both the CR and MC scores are highly significant (and positive) determinants of the final mark for both the posttest (Table 2, Models (2) and (3)) and pretest (Table 3, Models (2) and (3)) regressions. The institutional effect was significant in all models and the gender effect was significant in most models. It therefore turned out to be important to control for these effects but we do not discuss them further as they are outside the focus of this paper.

TABLE 2: Regression Results for Posttest Data							
	The depe	endent variable	is the final man	rk for the course	е.		
	(2)	(3)	(4)	(5)	(6)	(7)	
CRscore	2.98**		2.33**		9.56**		
	(6.18)		(5.14)		(4.10)		
MCscore		7.12**		5.67**		40.63**	
		(3.14)		(2.74)		(4.09)	
GPA			5.42**	5.92**	8.92**	9.55**	
			(7.83)	(8.44)	(6.85)	(7.81)	
CRscore*GPA					-1.32**		
					(-3.16)		
MCscore*GPA						-6.38**	
						(-3.59)	
Female, Aust	-3.26*	-3.21*	-5.07**	-5.21**	-4.71**	-5.36**	
	(-1.90)	(-1.80)	(-3.17)	(-3.16)	(-2.97)	(-3.31)	
Male, U.S.	20.12**	20.90**	12.84**	12.70**	13.18**	13.09**	
	(11.96)	(11.74)	(7.10)	(6.72)	(7.37)	(7.04)	
Female, U.S.	19.26**	20.70**	12.20**	12.59**	12.74**	12.88**	
	(10.98)	(11.40)	(6.59)	(6.59)	(6.94)	(6.85)	
Notes: All equations	were estimated	l using Ordinar	y Least Square	s, with t-statisti	ics given in par	rentheses. The	
sample size was 347 after adjustments. Significance at 5% and 10% is denoted by ** and * respectively.							

122

TABLE 3: Regression Results for Pretest Data							
	The dep	endent variable	is the final man	k for the course			
	(2)	(3)	(4)	(5)	(6)	(7)	
CRscore	3.20**		2.67**		3.14		
	(5.56)		(5.34)		(1.55)		
MCscore		7.42**		5.12**		0.55	
		(2.98)		(2.36)		(0.06)	
GPA			6.80**	6.92**	6.98**	6.58**	
			(12.84)	(12.78)	(7.39)	(7.82)	
CRscore*GPA					-0.09		
					(-0.24)		
MCscore*GPA						0.85	
						(0.52)	
Female, Aus.	-1.31	-1.41	-3.72**	-3.93**	-3.73**	-3.97**	
	(-0.83)	(-0.86)	(-2.68)	(-2.76)	(-2.69)	(-2.78)	
Male, U.S.	21.31**	22.36**	10.61**	11.57**	10.65**	11.39**	
	(11.43)	(11.55)	(5.84)	(6.15)	(5.83)	(5.95)	
Female, U.S.	21.57**	22.75**	10.89**	11.83**	10.92**	11.74**	
	(11.25)	(11.67)	(5.87)	(6.25)	(5.86)	(6.17)	
Notes: All equations	were estimated	using Ordinar	y Least Square	s, with t-statisti	ics given in par	entheses. The	

sample size was 498 after adjustments. Significance at 5% and 10% is denoted by ** and * respectively.

Next we control for GPA in Models (4) and (5):

$$Final_mark_i = \alpha_0 + \alpha_1 CRscore_i + \alpha_2 DFA_i + \alpha_3 DFUS_i + \alpha_4 DMUS_i + \varepsilon_i$$
(4)

$$Final_mark_{i} = \alpha_{0} + \alpha_{1}MCscore_{i} + \alpha_{2}DFA_{i} + \alpha_{3}DFUS_{i} + \alpha_{4}DMUS_{i} + \varepsilon_{i}$$
(5),

where GPA^5 is the student's grade point average at the start of the course⁶ and *i* is the student subscript. GPA is highly significant, while the MC and CR scores remain significant in both the pretest and posttest regressions. This suggests that both assessment types are effective indicators of final achievement.

Perhaps the more interesting question is whether the strength of MC and CR scores as determinants of the final mark depends on the academic ability of the students. The effect could go either way a priori. Take the potential effect of GPA in mediating the effect of CR scores on the final mark. It may be that the performance on CR questions is a better indicator of final achievement for lower-ability students than for higher-ability students, because the former are more likely to be confused by the language of a MC question and therefore benefit more from the opportunity to explain their 'understanding' through a CR. On the other hand, we might observe the opposite effect if the higher-ability students are able to demonstrate their more sophisticated 'understanding' on CR questions rather than through a MC box-ticking exercise. Assuming GPA is a measure of academic ability, we introduce the interaction terms *MCscore***GPA* and *CRscore***GPA*:

$$Final_mark_{i} = \alpha_{0} + \alpha_{1}CRscore_{i} + \alpha_{2}GPA_{i} + \alpha_{3}CRscore_{i} * GPA_{i} + \alpha_{4}DFA_{i} + \alpha_{5}DFUS_{i} + \alpha_{6}DMUS_{i} + \varepsilon_{i}$$
(6)

$$Final_mark_{i} = \alpha_{0} + \alpha_{4}MCscore_{i} + \alpha_{2}GPA_{i} + \alpha_{3}MCscore_{i} * GPA_{i} + \alpha_{4}DFA_{i} + \alpha_{5}DFUS_{i} + \alpha_{6}DMUS_{i} + \varepsilon_{i}$$
(7)

$$mal_mark_i = \alpha_0 + \alpha_1 MCscore_i + \alpha_2 GPA_i + \alpha_3 MCscore_i * GPA_i + \alpha_4 DFA_i + \alpha_5 DFUS_i + \alpha_6 DMUS_i + \varepsilon_i$$
(7)

The CR scores, MC scores and interaction terms are significant in the posttest but not pretest regressions, indicating that the effects of MC and CR scores are due to the learning process. In Model (6) of the posttest regression, the CR score by itself is significantly positive while the interaction variable, CRscore*GPA, is significantly negative. The same is true in Model (7) of the posttest regression for the MC score. This indicates that GPA is mediating the impact of the two types of scores on final marks. In other words, students with higher CR scores (on the posttest) achieved a higher final mark, but this effect was weaker for students of higher ability. The same is true for MC scores. There must therefore be some learning outcomes captured in final achievement that are measured less accurately by MC and the associated CR questions for high-GPA students than for low-GPA students.⁷ Those learning outcomes are measured by the assessment tasks included in final achievement that are not MC or CR questions, such as the problem-solving tasks and short essays that make up 50% of final achievement. These types of tasks may be more effective in discerning more sophisticated conceptions, compared with MC and CR questions. On the assumption that high-GPA students achieve more sophisticated conceptions than low-GPA students, the results suggest that MC and CR questions are not as effective in discerning these more sophisticated conceptions as are the other tasks included in the final mark.

CONCLUSION

The exploratory hybrid test instrument analyzed here consisted of multiple choice (MC) questions that required students to construct short paragraph responses (CR) to justify their choices. The analysis of the CRs indicated a wide range of conceptions of price and opportunity cost, in particular varying depths of 'understanding' of price, which is consistent with the literature. The additional information provided by the CRs revealed shortcomings in the effectiveness of standard MC questions in assessing 'understanding' of economic concepts, especially price.

The regression analysis indicated that both MC scores and CR scores on the posttest were alone significant determinants of the final mark, but their effect also depends on the student's GPA – the effect of both MC and CR scores was weaker for higher-GPA students. Given that approximately 50% of the final mark was made up from other types of assessment tasks consisting of problems and short essays, the mediating effect of GPA suggests that performance on these alternative tasks varies with GPA. We posit the reason may be that higher GPA students achieve more sophisticated conceptions, which are more reliably captured by assessment tasks such as problems and short essays. In that case, the MC questions and associated CRs would not be as effective in capturing the sophistication of student conceptions.

These results, although exploratory due to the small sample of MC questions, suggest implications for assessment of fundamental concepts such as price and OC in the wider context of international progress towards assessment of learning standards in economics. The OECD's Assessment of Higher Education Learning Outcomes (AHELO)⁸ Feasibility Study, which was completed at the end of 2012, has developed a test instrument consisting of 45 multiple choice (MC) questions and one constructed response (CR) question (not linked to any of the MC questions). It is therefore an instrument heavily reliant on MC questions. The evidence in this paper suggests caution in relying on such instruments. It provides an argument for weighting the balance further towards CR questions relative to MC questions, and supplementing with other

types of assessment tasks such as problems and essays. Of course, such decisions have to take account of the higher costs of the grading and moderating process for questions other than MC.

ENDNOTES

- 1. The test used to provide this evidence was not part of overall assessment in the course. The "course" refers to the 13 week program of study, which might otherwise be called a "unit" or "subject".
- 2. Question Q1 came from the Economic Literacy Quiz produced by the National Council of Economic Education, at http://www.councilforeconed.org/news-information/economic-literacy-quiz/; and Questions 2 to 4 came from the UK Economics Network Test Bank, at http://www.economicsnetwork.ac.uk.
- 3. For the data from the Australian university, a potential bias is acknowledged in that the smaller sample on the posttest test reflected the smaller number of students who happened to attend class on the date that the test was administered, since it was administered in a normal lecture (without prior notice). A number of students had decided not to regularly attend class since the time of the pretest, which was on day one of the course.
- 4. Final marks were determined on a 100-point scale for all three cohorts.
- 5. Students at the U.S. university were enrolled as first-semester students, and thus did not have a GPA at the start of the course. For these students, self-reported scores on the ACT or SAT standardized U,S, college admissions exams were used as a proxy for prior learning and ability. First, reported ACT or SAT scores were divided by the maximum possible score on each exam. Scores were then multiplied by 7, in order to keep them on the same 7-point scale as GPA at Griffith University. If students completed both the SAT and ACT, an average of the two scores was used instead. Econometric models do not distinguish between SAT and ACT scores.
- 6. Since GPA was obtained at the start of the course, it does not include the final mark that is being represented by the dependent variable in this regression.
- 7. These results may reflect a ceiling effect, in that higher ability students (who score higher grades) have less scope to demonstrate the effect of 'understanding' on their final grades simply because they are closer to the maximum marks that can be awarded for the course. However, there were a small proportion of scores in the 80's and 90's among all cohorts, which suggests that most of the high ability students had as much scope to improve their final scores as the low ability students.
- 8. AHELO stands for Assessment of Higher Education Learning Outcomes. The website is http://www.oecd.org/document/22/0,3746,en_2649_35961291_40624662_1_1_1_1_0.0.html

REFERENCES

- Alexander, P.A. (2000). Toward a model of academic development: Schooling and the acquisition of knowledge: The sequel. *Educational Researcher*, **29**: 28–33.
- Becker, W.E. & M. Salemi. (1977). The learning and cost effectiveness of AVT supplemented instruction: specification of learning models. *Journal of Economic Education*, 8, 77-92.
- Becker, W. & C. Johnston. (1999). The relationship between multiple choice and essay response questions in assessing economics understanding. *The Economic Record*, 75(231), 348-357.
- Bennett, R.E., D. A. Rock & M. Wang. (1991). 'Equivalence of free-response and multiple-choice items.' *Journal of Educational Measurement*, 28(1), 77–92.
- Biggs, J. (1999). *Teaching for quality learning at university*. Buckingham: The Society for Research into Higher Education and Open University Press.
- Biggs J.B. (2002). Aligning teaching and assessment to curriculum objectives. LTSN Imaginative Curriculum Guide IC022.
- Bloom, B.S. (Ed.), M.D. Engelhart, E J. Furst, W.H. Hill & D.R. Krathwohl. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain. New York: David McKay.
- Bridgeman, B. & D. A. Rock. (1993). Relationships among multiple choice and open ended analytical questions. *Journal of Educational Measurement*, 30(4), 313-29.
- Buckles, S. & J.J. Siegfried. (2006). Using multiple-choice questions to evaluate in-depth learning of economics. *Journal of Economic Education*, 37, 48–57.
- Davies, P. (2011). Students' conceptions of price, value and opportunity cost: some implications for future research. *Citizenship, Social and Economic Education, 10*(2-3), 101-110.
- Davies, P. (2012). Threshold concepts in economics. In G. Hoyt and K.M. McGoldrick (Eds.), *The International Handbook on Teaching and Learning Economics (pp. 250-256)*. Aldershot: Edward Elgar Press.
- Davies, P. & C. Lundholm. (2012). Young people's understanding of socio-economic phenomena: conceptions about the free provision of goods and services. *Journal of Economic Psychology*, 33, 79-89.
- Davies, P. & J. Mangan. (2007). Threshold concepts and the integration of understanding in economics. Studies in Higher Education, 32, 711-726.
- Ferraro, P.J. & L. O. Taylor. (2005). Do economists recognize an opportunity cost when they see one? A dismal performance from the dismal science. *Contributions to Economic Analysis and Policy*, *4*, 1-12.
- Harris. R.B. & W.C. Kerby. (1997). Performance sssessment as a complement to multiple-choice testing: the evidence from California's golden state examinations in economics. *Journal of Economic Education*, 28(2), 122-34.
- Heath, J.A. (1989). An econometric model of the role of gender in economic education. *American Economic Review*, 79(2), 226-230.
- Hickson, S. & B. Reed. (2011). More evidence on the use of constructed-response questions in principles of economics classes. *International Review of Economics Education*, 10(2), 28-48.
- Kennedy, P. & W. Walstad. (1997). Combining multiple-choice and constructed-response test scores: an economist's view. *Applied Measurement in Education*, *10* (4), 359-375.
- Leiser, D. & R.B. Halachmi. (2006). Children's understanding of market forces. *Journal of Economic Psychology*, 27, 6-19.
- Lukhele, R., D. Thissen & H. Wainer. (1994). On the relative value of multiple-choice, constructed-response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement*, *31*, 234–250.
- Marton, F. & M. F. Pang. 2008. The idea of phenomenography and the pedagogy of conceptual change. In S. Vosniadou (Ed.), *International Handbook of Conceptual Change* (pp. 533-559). New York: Routledge.
- Meyer, J.H.F. & M. Shanahan. (2002). On variation in conceptions of 'price' in economics. *Higher Education*, 43, 203-225.
- Moss, P.A. (1992). Shifting conceptions of validity in educational measurement: implications for performance assessment, *Review of Educational Research*, 62, 229-258.
- O'Donnell, R.(2009). The concept of opportunity cost: is it simple, fundamental or necessary? *Australasian Journal of Economics*, 6 (1). Retrieved July 31, 2013 from

http://www.uq.edu.au/economics/AJEE/docs/index.php?dir=Volume+6%2C+Number+1%2C+2009%2F.

Pang, M. & F. Marton. (2003). Beyond "lesson study": comparing two ways of facilitating the grasp of some economic concepts. *Instructional Science*, *31*, 175-194.

- Pang, M. & F. Marton. (2005). Learning theory as teaching resource: enhancing students' understanding of economic concepts. *Instructional Science*, 33, 159-191.
- Pong, W-Y. (1998). Students' ideas of price and trade. Economic Awareness, 9(2), 6-10.
- Salemi, M. (2005). Teaching economic literacy: why, what how? *International Review of Economics Education*, 4, 46-57.
- Saunders. P. & W. Walstad. (1990). The principles of economics course: handbook for instructors. New York: McGraw Hill.
- Sevón, G. & S. Weckström. (1989). The development of reasoning about economic events: A study of Finnish children. *Journal of Economic Psychology*, *10*, 495-514.
- Shanahan, M., G. Foster & J.H.F. Meyer. (2006). Operationalising a threshold concept in economics: a pilot study using multiple choice questions on opportunity cost. *International Review of Economics Education*, *5*, 29-57.
- Tan, K.C.D., N.K. Goh, L.S. Chia & D.F. Treagust. (2002). Development and application of a two-tier multiple choice diagnostic instrument to assess high school students' understanding of inorganic chemistry qualitative analysis. *Journal of Research in Science Teaching*, *39*, 283-301.
- Thissen, D., H. Wainer & X. Wang. (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*, *31*, 113–23.
- Thompson, D. R., & R.S. Siegler. (2000). Buy low, sell high: The development of an informal theory of economics. *Child Development*, *71*, 660–677.
- Von Wieser, F. (1891). The Austrian school and the theory of value. *Economic Journal*, 1, 108-121.
- Voska, K.W. & H.W. Heikkenen, (2000). Identification and analysis of student conceptions used to solve chemical equilibrium problems, *Journal of Research in Science Teaching*, *37*, 160-176.
- Walstad, W.B. (1987). Measurement instruments in William E. Becker and William B. Walstad (eds). *Econometric* Modeling in Economic Education Research, Boston: Kluwer-Nisjhoff.

APPENDIX

THE HYBRID TEST INSTRUMENT

Answer all questions and underneath each question briefly explain your answer. Question 1:

If your local government sets a maximum amount that landlords can charge in rent, what is the most likely result?

A. There will be more apartments available than people want to rent.

- B. There will be fewer apartments available than people want to rent.
- C. The number of apartments available will be equal to the number of people that want to rent.
- D. Don't Know.

Please explain:

Question 2:

A ticket in a raffle in which the prize is a day trip to Sydney is purchased for \$1. The regular price for this trip is \$200, but the organisers of the raffle are offering a cash sum of \$150 to the winner as an alternative to the trip itself. What will be the real cost for the winner of the raffle, if they decide to go on the trip?

- A. \$1
- B. \$150
- C. \$199
- D. \$200

Please explain:

Question 3:

If people are offering to pay \$200 for tickets to the World Cup final on the black market, and someone gives you a ticket, what does it cost you to attend the game ?

- A. \$200 because you could have sold the ticket for \$200 on the black market.
- B. Nothing because you have not paid anything for the ticket.
- C. \$200 less the value of your time if you did not go to the game.
- D. \$200 plus the value of your time if you did not go to the game.

Please explain:

Question 4:

Suppose a massive new oil field was discovered, causing the price of petrol to fall to \$0.40 per litre for the foreseeable future. According to economic analysis, which of the following would most likely happen?

- A. The price of small, fuel-efficient vehicles would probably rise as suppliers were forced to cope with falling demand.
- B. The price of electric cars would fall, but only if they could be designed to run on petrol instead of electricity.
- C. Prices for vehicles with relatively low fuel efficiency, such as SUV's (or 4 wheel drives), would probably rise.
- D. The price of electricity would probably rise.

Please explain:

END OF TEST

QUALITATIVE DIFFERENCES BETWEEN STUDENTS' CONSTRUCTED EXPLANATIONS OF THEIR ANSWERS

These data encompass responses from all three student cohorts on both the pretest and posttests.

	Multiple choice option and numbers of students. * denotes correct						
		QUESTIONT	of stude	ans.	wer		CI .
Туре	Score	Example	No response or multiple choice				
			answers	Α	B *	С	D
No response	NA		10	6	14	5	10
0. Irrelevant or nonsensical answer	0		0	56	46	20	12
1. Explanation of effect on price that is unrelated to supply and demand	0	"The government rent limitation only sets the maximum price a landlord can't set for rent, and does not necessarily effect the 'wanting to rent' effect of people."	1	30	30	20	9
2. Maximum price will have no effect	0	"Whether or not the government impose maximum amount of rent restrictions or not will not have a direct impact on any of the above answers."	0	2	4	12	1
3. Maximum rent will increase demand	1	"It will increase the popularity of renting. This is because people will become more financially able to rent making the practice more popular. More demand, same supply."	0	7	187	9	2
4. Maximum rent will reduce supply	2	"The landlords who are trying need to charge more than the max amount set by the government will no longer want to/be able to support/invest in the apartment. New potential landlords will be less inclined to enter the market due to lack of freedom in terms of rent they charge."	0	8	125	1	2
5. Maximum rent will increase demand and reduce supply	4	"If the government cap the amount that can be charged for rent I think this will make less apartment available. If the amount is capped to a reasonable price people would be able to afford a more range of apartments meaning more people will be looking for apartments."	0	2	77	0	0
6. No further consideration of alternatives	2	"The amount of information is insufficient to arrive at an answer. Ideally we would also like to know whether the amount the governments sets is higher or lower than the current occupancy rate for available houses and the current number and more info of new dwellings being built."	1	2	3	3	25

			Multiple ch	oice o	ption a	nd nun	nbers	
	(QUESTION 1	of students. * denotes correct					
			answer					
Туре	Score	Example	No					
			response					
			or					
			multiple					
			choice					
			answers	Α	B *	С	D	
7. If maximum	4	"It depends on whether this would						
rent is set above		encourage landlords to raise or lower the						
the current		price in comparison to the previous						
equilibrium then		market value. However, it would be						
there will be a		most likely that the government would						
glut (S>D) (i.e.		impose this law if the rents were too						
misunderstands		high. Therefore, the rents would be						
equilibrium		lower and there would be fewer						
process), but if		apartments available."						
max is price								
below								
equilibrium there								
will be a shortage								
(D>S)			0	5	11	2	14	
8. If maximum	5	"If there is a maximum set (assuming						
rent set below		that the maximum is lower than the						
equilibrium there		landlords preferred amount) then renting						
will be a		will become cheaper and therefore more						
shortage, but		people are more likely to rent instead of						
otherwise no		buying therefore creating fewer						
effect		apartments."	2	4	60	1	4	

QUESTION 2			Multiple cho of students. *	ice opti denote	on and	l num ect ans	bers swer
Туре	Score	Example	No				
			response or				
			multiple				
			choice				
			answers	Α	B *	С	D
No response	NA		17	24	19	16	6
0. Irrelevant or	0						
nonsensical							
answer			0	7	8	5	3
1. Uses the value	0	"Winner \$1 cost to win - $200 - 1 =$					
of the item as the		\$199"					
guide to cost			0	14	2	60	25
2. Treats 'real	1	"\$1 as they are receiving a \$200 trip for					
cost' as expenses		purchasing a \$1 ticket. \$1 is all they					
directly incurred		have spent"	0	321	3	6	0
3. Understands	3	"They put in \$1 to start but it also					
opportunity cost in		depends if they go on a day they would					
terms of foregone		normally work. Then they would lose a					
earnings, but		day's wage or a day's leave."					
ignores the cash			4	9	9	1	3
4. Understands	5	"The alternative to the trip was \$150					
opportunity cost in		cash and so the winner is effectively					
terms of foregone		spending this amount by going on the					
cash		trip and not cashing in the money"	0	31	249	0	3

			Multiple	choic	e optio	on and	1
	QUESTION 3			f stud	ents. *	deno	tes
			CO	rrect a	answer		
Туре	Score	Examples	No				
			response or				
			multiple				
			answers	Α	В	С	D*
No response	NA		11	14	45	22	27
0. Irrelevant or	0						
nonsensical							
answer			2	5	10	7	10
1. Treats 'real	1	"Because you paid nothing for the ticket					
cost' as the price		therefore no cost incurred."					
that is paid (says							
answer is 0)			1	4	157	2	4
2. Recognizes that	3	"Opportunity cost means calculating					
foregone wages		implicit costs as well. For example you					
are part of		may have had the opportunity to spend					
opportunity cost		your time earning money at a job but					
		didn't because you went to the game."	1	5	16	28	39
3. Recognizes that	3	"Opportunity $cost = 200 could have got					
a trade-on value of		from selling ticket."					
an item is part of							
the opportunity							
cost			1	77	18	11	16
4. Recognizes both	5	"I believe it should be \$200 PLUS the					
foregone earnings		value of your time if you didn't attend					
and trade on value		the game. You could have sold the ticket					
in calculating		for \$200 which means you are \$200					
opportunity cost-		worse off PLUS the value of your time.					
-		You could have been working and					
		earning money or spending time doing					
		something you enjoy more."	3	5	3	13	288

			Multiple choice	e opti	ion ar	nd num	bers	
	QU	ESTION 4	of students	s. * de	notes	correc	t	
			answer					
Туре	Score	Examples	No response					
			or multiple					
			answers	Α	В	C*	D	
No response	NA		34	44	26	52	5	
0. Irrelevant or	0							
nonsensical answer			4	24	24	36	19	
1. Suggests that	1	"Because people would prefer to go						
demand for car types		for cheaper option or just a petrol						
depends on innate		car."						
qualities of car			1	11	8	7	6	
2. Believes that a fall	2	"The demand for fuel efficient cars						
in demand with		may fall as three is less stress from						
downward sloping S		petrol prices therefore the supply						
that does not shift		would also need to be less but at a						
will raise price		higher price to achieve profits in						
		which were achieved when there was						
		a higher demand."						
			1	95	5	13	4	
3. Expects rise in	3	"People will probably demand more						
demand and price for		vehicles, meaning price will also						
all types of car		increase."	0	14	2	13	0	
4. Expects rise in	4	"When the price of petrol falls the						
demand for SUVs		fuel cost for a car owner will						
causing a rise a price		decrease. More people would like to						
(recognizes relative		buy the low fuel efficiency cars since						
price)		they will have less cost than before.						
		When the demand is larger the price		1				
		will be higher."	2	36	25	330	4	