

Can Multiple-Choice Questions Replace Constructed Response Test as an Exam Form in Business Courses? Evidence from a Business School

*By Leiv Opstad**

The discussion of whether multiple-choice questions can replace the traditional exam with essays and constructed questions in introductory courses has just started in Norway. There is not an easy answer. The findings depend on the pattern of the questions. Therefore, one must be careful in drawing conclusions. In this research, one will explore a selected business course where 30 percent of the test is comprised of multiple-choice items. There obviously are some similarities between the two test methods. Students who perform well on writing essays tend also to achieve good results when answering multiple-choice questions. The result reveals a gender gap where multiple-choice based exam seems to favor the male students. There are some challenges in how to measure the different dimensions of knowledge. This study confirms this. Hence, it is too early to conclude that a multiple-choice score is a good predictor of the outcome of an essay exam. This paper will provide a beneficial contribution to the debate in Norway, but it needs to be followed up with more research.

Keywords: multiple choice test, constructed response questions, business school, gender, regression model.

Introduction

There are some studies related to applying the multiple-choice format in exam tests. Due to the complexity of using multiple-choice, one cannot simply transfer results from other studies. Therefore, it is of interest to explore the experience from a Norwegian business school. One needs more research to get a more complete view of how different kinds of exams work (Krieg & Uyar, 2001).

Contrary to Northern America, in Norway, it is normal to have a traditional 4-hour written school exam in business courses. This constructed-response (CR) questions format places test-takers under pressure. Many want a system with multiple-choice (MC) questions such as is heavily used in the US, especially in large-scale testing (Livingston, 2009). Becker and Watts (2001) suggested that MC-test make up 45 percent of the exams in economics. Some advantages of the MC-based exam are low administrative costs, the test is easy to complete, and one avoids subjective assessments. It is indisputable how many scores each student has achieved (Walstad, 1998).

The Business School of the Norwegian University of Science and Technology (NTNU) has considered increasing the use of MC-based exams. The reasons are as follows. Firstly, there will be a common and identical exam at three geographically separate campuses from 2020 onwards. There are different instructors and the enrollment qualifications may vary depending on university grounds. Altogether, there might be from 500 to 800 students in the same compulsory course. With

*Professor, Norwegian University of Science and Technology, Norway.

several examiners involved, there can be inequality in judging of the students' performance. Two students with the identical result do not necessarily get the same grade. Ensuring equal treatment of the students requires substantial resources. With MC-based questions, there will be no bias or different assessment among the business classes. Secondly, there are financial pressures to reduce the costs associated with the exam. Traditionally, there have been two evaluators to determine a student's letter grade. Today, it is a pressure to have only one examiner, even in courses with many students. A third objective is to reduce the numbers of students asking for validation of the given grade and the numbers of students who complain about the grade decision request a new exam commission. Undergraduates in Norway have strong rights. It is easy and cost-free to ask for justification of the letter grade and demand new evaluation. The proportion of students doing this has increased. In many courses, the amount has surpassed 30 percent. One can assume that the number of undergraduates doing this will increase after establishing a joint exam among the campuses/colleges. Such reviews take up a lot of time for the instructors and the administrative staff.

The purpose of this article is to see if the MC-based exam can substitute CR-based exams in business courses. This will be done by exploring data from the macroeconomic course where one used both CR-based and MC-based tests over the last 10 years. Experiences from this course will hopefully be relevant to other subjects. This paper does not investigate whether replacing of CR with MC will affect the way undergraduates' study and acquire knowledge. This study is limited to looking at if the choice of test method affects the grade level and the ranking of students and will be achieved by comparing data from 2012 to 2016.

An important question is whether one measures the same competencies by the two methods. To answer this, it may be useful to start with Bloom's taxonomy of educational objectives (Bloom, 1956):

1. Knowledge
2. Comprehension
3. Application
4. Analysis
5. Synthesis
6. Evaluation

Knowledge is about memorizing fact while *comprehension* requires that one understands the meaning of knowledge using one's own words. *Application* measures if the student understands how to use the knowledge. *Analysis* means the student uses the principles learned to derive new knowledge. *Synthesis* implies the student is able to solve problems by combining the given information. *Evaluation* requires that the student can discuss the usefulness of the knowledge and theories.

The key question is to what extent is one able to measure these six levels using an MC-based test or a CR-based assessment? It is easy to create MC-based questions from the two first level of Bloom's taxonomy (Hickson & Reed, 2011). According to Buckles and Siegfried (2016), Scully (2017), and Zheng, Lawthorn, Lumley, and Freeman (2008), it is considerably more difficult but still possible to

design questions that also capture the two next dimensions. Aiken (1982) suggests also that it is challenging for one to produce an MC-based test that involves higher-order thinking. It is easier to test higher levels of learning by presenting CR-based questions (Hickson, Reed, & Sander, 2012). As a part of the assignment, it is important to promote reflective judgement, analysis, and evaluation (Dubas & Toledo, 2016; Dwyer, Hogan, & Stewart, 2014). One likely can achieve this better by applying constructed questions and essays.

If the questions are connected to the lower level of Bloom's taxonomy, one can get one result, but it may differ if one uses the higher level (Bacon, 2003). A lot of MC answers are based on partial understanding of knowledge and one might obtain the wrong expression of the students' academic abilities (Simkin & Kuechler, 2005).

Literature Review

The literature is mixed in the analysis of comparability of MC-based and CR-based tests. Important contributors to this discussion are Becker and Johnston (1999), Walstad (1998), and Walstad & Becker (1994).

An advantage of MC-based tests, compared to CR-based tests, is that it is easy for examiner to determine if the correct answer has been provided. On the other hand, they are difficult to construct. There can be a substantial variation in the quality of the questions; this affects the results. Bush (2001) showed how one could achieve high score just by guessing. If there is a question with five possible answers, one can expect 20 percent correctly answered. Assume that, of half the questions, the student finds that only two of the options could be true. This increases the likelihood that one can guess the right answer to 35 percent. This guessing can lead to a lack of validity (Mbonigaba & Oumar, 2017).

The literature shows that the effect of the various types of exam depends on the form of the questions. One can produce a MC version that is difficult or easy for students to answer. The MC-based test can be too simple and hence will not be a good indicator of students' understanding (Simkin & Kuechler, 2005). By changing the design and form of the MC questions, the students will perform differently (Chan & Kennedy, 2002).

Many studies suggest that MC and CR formats do not measure the same understanding or performance. One can miss some basic understanding by using an MC-based test. It is hard to achieve the advanced level of knowledge by using the MC format (Simkin & Kuechler, 2005). Mbonigaba and Oumar (2017), basing their investigation on Bloom's taxonomy, found that students have different scores using MC or CR, depending on the level. Therefore, the MC-test ranking is inconsistent with the level of cognitive ability. Using CR-based test develops better cognitive and writing skills (Welsh & Saunders, 1998). With essays and CR-questions, the students can write with their own words, analyze a topic, and demonstrate originality. Even though there are differences between CR and MC, there is a link between them. For instance, the reported correlation coefficient between MC and CR is quite high, but this relationship depends heavily on the

chosen questions (Chan & Kennedy, 2002). Walstad and Becker (1994) suggest the correlation coefficient is 0.69 in microeconomics and 0.65 in macroeconomics. Some authors report a substantially lower value (Hickson & Reed, 2011). The score is higher in more quantitative courses (chemistry, calculus, etc.) and less in non-quantitative courses (history, language, etc.). They ask if it is of enough economic value to include an essay component, since it does not give much more information. The connection between two CR-based tests in within the same course is similar to those one observes between MC and CR. Therefore, Hickson, Reed, and Sander (2012) suggest that the consequences will be rather small by switching to MC-test only. Furthermore, the two highest levels of academic skill (Bloom's taxonomy) are the most appropriate for master's- and PhD-level students. Only to a small extent are the students tested on these topics in the introductory courses. In addition, many students are attending introductory courses. This argues for using MC questions for undergraduates. This might explain why multiple-choice and essay scores are substantially significant for explaining achievement on a less advanced level (Becker & Johnston, 1999). However, if one wants to measure sophisticated dimensions of knowledge, there is a substantial difference between MC and CR.

Many researchers indicate there might be a gender difference in the performance depending on the exam form. The CR-based test seems to favor females, while the opposite is true for MC-based tests (Becker & Johnston, 1999; Livingston & Rupp, 2004; Smith & Edwards, 2007). The reason why females get relatively lower scores on multiple choice questions can be related to other factors than academic understanding, such as differences in reasoning, socialization, and instructional practice (Walstad & Robson, 1997). Females might have advantages at writing essays. This can be another factor explaining why they are falling behind when evaluated by MC-based questions in economic courses (Becker & Johnston, 1999).

Data, Model and Findings

The Sample

The data were taken from the school's database and from the instructor's note related to the determination of the grades. The data derive from a period of 5 years (2012–2016). The students had the same instructor and identical textbook during this interval. The exam form was based on the same principle or every year in the chosen macroeconomic course. It is compulsory and runs the second year for undergraduate students. The paper exam (4 hours) consists of three parts:

1. Apply economic theory and analysis using a familiar mathematical model.
2. Answering current issues by applying the theory.
3. A conventional multiple-choice test with five-answer questions and no negative marking.

Numbers 1 and 2 are structured/essay questions (CR). The first assignment makes up for 50 percent, the second 20 percent and the last section, 30 percent. The students were informed of this. There were 32 questions in the MC-based test. As a result, the score was determined by subtracting “2” from the number of correct answers. Since the CR- and MC-based tests were taken simultaneously for the same group of students, this gives a good opportunity to compare the outcome. The assignments reflect different levels of Bloom’s taxonomy. The first segment is somewhat predictable, so the students can learn the process in advance, but part of the questions are challenging and require good academically skill. Exercise 2 is demanding. Here, the students are inspired to apply the theory to more or less new and unknown issues. The CR-based questions (part 1 and 2) primarily test the application and analysis levels of skills, some fragment of Exercise 1 includes the Comprehension level, while the advanced Synthesis level might be touched in Exercise 2. On the MC-based test, there are questions from the whole textbook. A suggestion is that about one-third of the questions is testing the Knowledge level, another third the Comprehension level, and the final third cover the Application or Analysis level.

Table 1. The Data

	Min	Max	Mean	St. Dev.	N
MC-score (scale 0 to 100)	13.3	100	60.5	15.4	1190
CR-score (scale 0 to 100)	4.3	100	49.9	19.2	1190
Macroeconomics (grade)	0	5	3.10	1.19	1190
Gender (0:F, 1:M)	0	1	0.41	0.49	932
GPA	44.0	65.4	51.9	3.39	613
Microeconomics (grade)	0	5	3.24	1.20	887
Business mathematics (grade)	0	5	2.85	1.62	887
Management (grade)	0	5	2.88	1.26	927
Compulsory midterm MC-based test (macro-economics), 32 questions	8	31	19.6	4.3	911

Note: 0:F, 1:E, 2:D, 3:C, 4:B, 5:A.

Table 1 presents the data where gender, GPA (Grade Point Average) from high school and performance in some of the compulsory courses first year are included. For making the scores from MC and CR sections directly comparable, the scale is measured in percentages. Data are missing for some of the variables. The females were in the majority (57%). Using the same MC-method as for the final exam, the score of a compulsory midterm test was 55.

The Correlation between MC and CR

The correlation coefficients give an indicator to which degree multiple-choice questions can substitute constructed response questions (see Table 2).

Table 2a. Correlation Coefficients for the Whole

	MC	CR	MC-midterm
MC		0.62	0.43
CR	0.62		0.36
MC-midterm	0.43	0.36	

Table 2b. Correlation Coefficient between CR and MC for Each Year

	2012	2013	2014	2015	2016
Correlation coefficient	0.50	0.54	0.71	0.66	0.67

The values of all correlation coefficients were strongly significant at 1% level. The score is about 0.6 between MC and CR, but notice (Table 2b) that there is a considerable variation from year to year (0.50 in 2012 and 0.71 in 2014). Furthermore, notice that the link between the MC-based midterm test and final MC-based test is rather low. The correlation coefficient is 0.43.

The Gender Impacts

Table 3 shows the result of bilateral comparison of women and men.

Table 3. The Gender Effect - Independent Sample T-Test of Mean (Assuming Equal Variance)

	Females	Males	Diff.	St. Dev.	T-value	Sign. level
MC	52.7	64.7	-7.4	0.98	-7.5	0.000
CR	48.3	55.6	-7.3	1.22	-6.0	0.000

The male students perform better than the female students.

Regression Model

By using a linear regression model, one can find out more about the impact the independent variables simultaneously have on the chosen exam form. The explanatory variables were gender, GPA score, and performance in business mathematics, microeconomics and management. The GPA score measures academic ability. There was also a strong positive link between GPA scores and performance in macroeconomics (Jones, Kouliavtsev, & Ethridge Jr, 2013; Raimondo, Esposito, & Gershenberg, 1990). Mathematics competency is a key factor for success in economics (Ballard & Johnson, 2004; Opstad, 2018). It is an indicator of analytical and quantitative skills. Microeconomics and macroeconomics are closely linked together (Perumal, 2012). The data in this study reveal a correlation coefficient between performance in microeconomics and macroeconomics of 0.672. The achievement in microeconomics with a traditional 4-hour CR-based test can provide a picture of the students' skills in economics. The ability to write and present the content in a proficient manner has an impact on the students' performance. This attribute is important for the grade in a subject such as management. Therefore, this course is included as an independent variable.

The chosen linear regression production function for comparing MC- and CR-score is:

$$Y_i = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_5 + \varepsilon$$

where: Y_i : Score in macroeconomics using MC- or CR-test.

α_0 : Constant

X_1 : Gender (0: F, 1: M)

X_2 : GPA score

X_3 : Performance in Business mathematics (0: F, 1: E, 2: D, 3: C, 4: B, 5: A).

X_4 : Performance in Microeconomics (0: F, 1: E, 2: D, 3: C, 4: B, 5: A)

X_5 : Performance in an introduction course in management (0: F, 1: E, 2: D, 3: C, 4: B, 5: A)

ε : stochastic error

Due to missing data, the valid numbers in the regression were substantial smaller than the whole sample presented in Table 1. However, it looks like they are random, since they did not affect the mean values of the variables. Results are presented in Table 4.

Independent of chosen exam test, the GPA had no significant impact on the student's performance. The gender effect in favour of males was more powerful and stronger using the MC-based test. Business mathematics influenced the score, but the differences between the two assessment formats are minor. Performances in microeconomics and macroeconomics were tightly related to each other. However, the impact is higher for CR-based questions (The B-coefficient is 8.01 for the CR model and 5.66 for MC model). Unlike the other test method, there is a substantial significant link between performance in management and the CR-score.

Discussion

The findings in this research do not give an unambiguous answer to whether this is a good idea or not to replace CR-based questions with MC- based questions. There is a near relationship between the two alternative test methods. This is similar to results of prior studies. However, the correlation coefficient depends heavily on the design of the questions. This can explain why this rate varies between 0.5 and 0.7 (Table 2b). On the other hand, this is substantially higher than the link between the similar test methods (MC) used within the same course at different times during the semester (see Table 2a). This is a point that Walstad and Becker (1994) and Hickson, Reed, and Sander (2012) highlight. Even with an identical test method, there will be a considerable variation in the students' achievement. From this point of view, the correlation rate between MC and CR is rather high. Hence, one needs to be careful in drawing conclusions by just comparing MC- and CR-score. Regardless of the chosen identical method, the

correlation coefficient will probably be substantial lower than 1.0 for having two or more assignments for the same target group.

Table 4. Finding, Regression Model

	MC model			CR model		
	Coefficient (B) (Unstandardised)	T-value	Sig. level	Coefficient (B) (Unstandardised)	T-value	Sig. level
Constant	10.9 (8.52)			6.6 (9.73)		
GPA	0.48 (0,11)	1.27	0.20	0.17 (0.19)	0.90	0.366
Gender	5.01 1.12)	2.89	0.004 (***)	2.49 (1.28)	194	0.053 (*)
Business Math	1.45 (0.42)	4.46	0.000 (***)	1.24 (0.48)	2.58	0.010 (**)
Microeconomics	5.66 (0.61)	3.44	0.000 (***)	8.01 (0.70)	11.40	0.000 (***)
Management	0.13 (0.46)	0.27	0.787	2.22 (0.53)	4.20	0.000 (***)
	Adj. $R^2=0.376$ $N=473$			Adj. $R^2=0.442$ $N=473$		

Note: Standard error in parenthesis, *, ** and *** denote significance at the 10%, 5%, and 1% level, respectively. All VIP (Variable Importance of Projection) values are between 1 and 2.

Similarities between MC and CR

The regression model shows many similarities between the two assignments formats (CR and MC). The input variables explain nearly the same proportion of the variance for dependent variable (the value of adjusted R -square). The *GPA* has no significant impact on the performance on neither MC-based exam nor CR-based exam. The influence of *business mathematics* is almost the same for the two discussed test methods. It is not surprising, since a substantial proportion of MC-based questions require good skills in calculation.

Differences between MC and CR

The regression model also documents differences between the two exam forms. Adjusting for other variables, it turns out that there is a significant *gender gap*. Men will gain if one decides to select MC questions as an exam design. This is similar to many prior studies. There is strong gender equality in Norway but still there are substantial gender differences. This is a paradox. We will try to find some explanations (Ahlstrom & Asarta, 2019). The females struggle with quantitative courses (Naqvi & Naqvi, 2017; Opstad, 2020). This might be a reason why females prefer non-quantitative majors. Men select more quantitative subjects like

finance. This result might be related to that the women choose less advanced mathematics at high school (Opstad, 2018). Therefore, they have poorer attitudes towards mathematics (Opstad & Årethun, 2019). This can explain why females perform poorer than males in quantitative courses like macroeconomics. This study confirms this. However, this factor does not explain why males outperform females with MC-based test compared to CR-based test. Data are not available in this research to explain this difference. One reason might be that women in general are more risk-averse than men are (Pekkarinen, 2015). The time and competitive pressure are probably stronger with MC-questions compared to CR-questions. Therefore, the female undergraduates are less confident in answering multiple-choice questions. Hence, they perform relatively better in CR-based exam. Another explanation can be sociocultural differences (Johnson, Robson, & Taengnoi, 2014) and student-specific characteristics (Opstad & Fallan, 2010; Johnson, Robson, & Taengnoi, 2014). However, Riener & Wagner (2017) emphasize that the gender difference in MC-designed exam depends on the difficulty of the questions.

There is a considerable connection between the *performance in management* and for CR score, but this link does not seem to exist for the MC score. The explanation is probably that writing and presentation ability is catch up to a greater extent with CR questions than MC questions. This confirms results from some prior research. *Performance in microeconomics* is a good indicator for success in macroeconomics, but we notice the influence is stronger for CR-test. One reason may be that this exam method more closely captured the desired skills.

Another explanation might be that the exam in microeconomics is CR-based. May be the effect would have been differently if there was a MC-format exam in microeconomics.

Following Bloom's taxonomy, the two test alternatives will probably emphasize the distinct dimensions differently. This problem is not easy to solve. Some aspects are harder to measure using MC-test. This paper shows that there are some differences between CR- and MC-based tests. The question is whether the two tests complement or substitute for each other. Probably, more investigation is needed to make such a determination. To do so, one must follow two steps. Firstly, one must increase the percentage of MC questions of the final exam from, for instance, 30 to 50 percent of the final exam. Secondly, one should put more effort into designing the MC questions as well as increase the proportion of testing at the application/analysis level.

Limitations and Further Research

This study has compared the two exam formats with data from only one business school. Since the design of questions is critical for the result, one must be careful with respect to drawing general conclusions. More independent variables in the regression model would have been desirable, but available information from the school's database was limited. An important issue that is not touched in this

paper is whether a change of exam format from CR to MC will alter the way students acquire their knowledge.

To further research, it would be interesting to find out more of why there is a gender difference between the two test methods. One should also investigate in how the test methods influence the learning style and the academic qualification of the undergraduates.

Conclusion

In Norway, it is normal to use structured/essay questions as an exam format. There is a limited tradition of using MC questions. Due to identical exams from several colleges, it is challenging to avoid judgment bias; it also incurs considerable administrative costs. Therefore, there is pressure to increase the use of MC-based tests, especially in courses with more than 500 students. The findings of this study suggest that those tests can partly replace each other and partly supplement each other. The result indicates that MC-based test favor male students. The decision maker must consider this. It is difficult, however, to measure some dimensions of understanding by using MC.

The discoveries derived from this research can give constructive input into the discussion about more MC-based tests in introductory courses at business schools in Norway. However, one should experiment more with the design of the questions. There are many arguments against excluding essay questions, and only use MC-based questions in the final exam. For the time being, it might be valuable to follow the advice of Becker and Johnston (1999), who suggest using both forms of testing.

References

- Ahlstrom, L.J., & Asarta, C.J. (2019). Navigating the economics major: the effect of gender on students' degree pathways. In J. Hall & K. Lawson (eds.), *Teaching Economics* (pp. 115-136). Cham: Springer.
- Aiken, L.R. (1982). Writing multiple-choice items to measure higher-order educational objectives. *Educational and Psychological Measurement*, 42(3), 803-806.
- Bacon, D. (2003). Assessing learning outcomes: a comparison of multiple-choice and short-answer questions in a marketing context, *Journal of Marketing Education*, 25(1), 31-36.
- Ballard, C.L., & Johnson, M.F. (2004). Basic math skills and performance in an introductory economics class. *The Journal of Economic Education*, 35(1), 3-23.
- Becker, W.E., & Johnston, C. (1999). The relationship between multiple choice and essay response questions in assessing economics understanding. *Economic Record*, 75(4), 348-357.
- Becker, W.E., & Watts, M. (2001). Teaching economics at the start of the 21st century: still chalk-and-talk. *American Economic Review*, 91(2), 446-451.
- Bloom, B.S. (1956). *Taxonomy of educational objectives, the classification of educational goals, handbook I: cognitive domain*. Edited by Benjamin S. Bloom. New York: Longmans, Green and Company.

- Buckles, S., & Siegfried, J.J. (2006). Using multiple-choice questions to evaluate in-depth learning of economics. *Journal of Economic Education*, 37(1), 38-47.
- Bush, M. (2001). A multiple choice test that rewards partial knowledge. *Journal of Further and Higher Education*, 25(2), 157-163.
- Chan, N., & Kennedy, P.E. (2002). Are multiple-choice exams easier for economics students? A comparison of multiple-choice and “equivalent” constructed-response exam questions. *Southern Economic Journal*, 68(4), 957-971.
- Dubas, J.M., & Toledo, S.A. (2016). Taking higher order thinking seriously: using Marzano’s taxonomy in the economics classroom. *International Review of Economics Education*, 21(Jan), 12-20.
- Dwyer, C.P., Hogan, M.J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking Skills and Creativity*, 12(Jun), 43-52.
- Hickson, S., & Reed, B. (2011). More evidence on the use of constructed-response questions in principles of economics classes. *International Review of Economics Education*, 10(2), 28-49.
- Hickson, S., Reed, W.R., & Sander, N. (2012). Estimating the effect on grades of using multiple-choice versus constructive-response questions: data from the classroom. *Educational Assessment*, 17(4), 200-213.
- Johnson, M., Robson, D., & Taengnoi, S. (2014). A meta-analysis of the gender gap in performance in collegiate economics courses. *Review of Social Economy*, 72(4), 436-459.
- Jones, C.T., Kouliavtsev, M.S., & Ethridge Jr, J.R. (2013). Lower level prerequisites and student performance in intermediate business courses: Does it matter where students take their principles courses? *Journal of Education for Business*, 88(4), 238-245.
- Krieg, R.G., & Uyar, B. (2001). Student performance in business and economics statistics: does exam structure matter? *Journal of Economics and Finance*, 25(2), 229-241.
- Livingston, S.A. (2009). *Constructed-response test questions: why we use them; how we score them*. R&D Connections, No. 11. Educational Testing Service.
- Livingston, S.A., & Rupp, S.L. (2004). *Performance of men and women on multiple-choice and constructed-response tests for beginning teachers*. Princeton, NJ: Educational Testing Service.
- Mbonigaba, J., & Oumar, S.B. (2017). Multiple-choice questions and written questions matched according to levels of cognitive ability in an applied course: evidence and practical implications. *Africa Education Review*, 14(1), 139-154.
- Naqvi, A., & Naqvi, F. (2017). A study on learning styles, gender and academic performance of post graduate management students in India. *International Journal of Economics & Management Science*, 6(398), 2-6.
- Opstad, L. (2018). Success in business studies and mathematical background: the case of Norway. *Journal of Applied Research in Higher Education*, 10(3), 399-408.
- Opstad, L. (2020). Why are there different grading practices based on students’ choice of business major. *Educational Process: International Journal*, 9(1), 43-57.
- Opstad, L., & Fallan, L. (2010). Student performance in principles of macroeconomics: the importance of gender and personality type. *International Review of Economic Education*, 9(1), 76-92.
- Opstad, L., & Årethun, T. (2019). *Attitude towards mathematics among economics and business students in Norway. Is there any gender difference?* In Proceedings of the 16th International Conference on Enterprises, Systems, Accounting, Logistics and Management (16th ICESALM 2019), 93-109.
- Pekkarinen, T. (2015). Gender differences in behaviour under competitive pressure: evidence on omission patterns in university entrance examinations. *Journal of Economic Behavior & Organization*, 115(Jul), 94-110.

- Perumal, M. (2012). Should we teach microeconomics before macroeconomics? Evidence from an Australian university. *Australian Journal of Economics Education*, 9(2), 1-14.
- Raimondo, H.J., Esposito, L., & Gershenberg, I. (1990). Introductory class size and student performance in intermediate theory courses. *The Journal of Economic Education*, 21(4), 369-382.
- Riener, G., & Wagner, V. (2017). Shying away from demanding tasks? Experimental evidence on gender differences in answering multiple-choice questions. *Economics of Education Review*, 59(Aug), 43-62.
- Scully, D. (2017). Constructing multiple-choice items to measure higher-order thinking. *Practical Assessment, Research, and Evaluation*, 22(1), 4.
- Simkin, M.G., & Kuechler, W.L. (2005). Multiple-choice tests and student understanding: what is the connection? *Decision Sciences Journal of Innovative Education*, 3(1), 73-98.
- Smith, L., & Edwards, L. (2007). A multivariate evaluation of mainstream and academic development courses in first-year microeconomics. *South African Journal of Economics*, 75(1), 99-117.
- Walstad, W. (1998). Multiple choice tests for the economics course. In W. Walstad & P. Saunders (eds.), *Teaching Undergraduate Economics: A Handbook for Instructors* (pp. 287-230). New York, NY: Irwin/McGraw-Hill.
- Walstad, W.B., & Becker, W.E. (1994). Achievement differences on multiple-choice and essay tests in economics. *The American Economic Review*, 84(2), 193-196.
- Walstad, W.B., & Robson, D. (1997). Differential item functioning and male-female differences on multiple-choice tests in economics. *The Journal of Economic Education*, 28(2), 155-171.
- Welsh, A.L., & Saunders, P. (1998). *Essay questions and tests*. In W.B. Walstad & P. Saunders (eds.), *Teaching Undergraduate Economics: A Handbook for Instructors* (pp. 305-318). New York: McGraw-Hill.
- Zheng, A.Y., Lawthorn, J.K., Lumley, T., & Freeman, S. (2008). Application of Bloom's taxonomy debunks the "MCAT Myth". *Science*, 319(5862), 414-415.