International Journal of Anatomy and Research,
Int J Anat Res 2018, Vol 6(2.2):5250-55. ISSN 2321-4287
DOI: https://dx.doi.org/10.16965/ijar.2018.171

Original Research Article

A STUDY OF TEST CONSTRUCT AND GENDER DIFFERENCES IN MEDICAL STUDENT ANATOMY EXAMINATION PERFORMANCE

Cheryl Melovitz-Vasan [1], Matthew Gentile [2], Susan Huff [3],Nagaswami Vasan *[4].

[1] Assistant Professor, Department of Biomedical Sciences, Cooper Medical School of Rowan University, Camden, New Jersey, USA.

[2] Director of Assessment, Office of Medical Education, Cooper Medical School of Rowan University, Camden, NJ, USA.

[3] Assessment Coordinator, Office of Medical Education, Cooper Medical School of Rowan University, Camden, NJ, USA.

[*4] Professor of Anatomy, Department of Biomedical Sciences, Cooper Medical School of Rowan University, Camden, NJ, USA.

ABSTRACT

Background: A number of studies compare cognitive abilities of male and female students from elementary school through high school employing various test constructs and reported presence of gender-related differences having to do with the mode of assessment. The purpose of the present study was to examine whether there is a difference in the performances of male and female medical students. We employed two types of test constructs viz., constructed response (CR) and selected response (SR) examinations that use the same question stem.

Materials and Methods: Two types of test questions, CR and SR, were studied. Each CR and SR question used the same question stem to assure all questions were matched. Study participants were incoming first year medical students enrolled in a six-week summer enrichment anatomy course prior to the start of the school year. Group 1 included 16 students (8 male and 8 female) and Group 2 19 students (7 male and 12 female). The course focused on study of the thorax and abdomen and the student performances were analyzed.

Results and Discussions: Mean scores and statistical analysis showed comparable performance between male and female students. The independent sample t-test showed that, statistically, there were no significant differences in performances of male and female students in the CR or SR examinations, except in the Group 1 male and female abdomen SR examination [t (14) = 1.934, p< 05 1 tailed]. Collectively, these results show that male and female students in both the groups performed better in the SR test than the CR test.

Conclusions: Gender poses no limitation to medical student performance, irrespective of the type of examination format. It is possible to adopt CR examination as a formative evaluation tool for students to identify their deficits and strategize effective learning.

KEY WORDS: Selected Response Question Examination, Constructed-Response Examination, Gender Performance In Examination Format, Medical Education.

Address for Correspondence: Prof. Dr. Nagaswami Vasan, Professor, Department of Biomedical Sciences, Cooper Medical School of Rowan University, 401 South Broadway, Camden, NJ 08103 USA. Phone: +1 (856) 361-2890 **E-Mail:** vasan@rowan.edu

Access this Article online	Journal Information
Quick Response code	**International Journal of Anatomy and Research**
	ICV for 2016 90.30 — ISSN (E) 2321-4287 | ISSN (P) 2321-8967
	https://www.ijmhr.org/ijar.htm
	DOI-Prefix: https://dx.doi.org/10.16965/ijar
DOI: 10.16965/ijar.2018.171	**Article Information**
	Received: 02 Mar 2018 — Accepted: 05 Apr 2018
	Peer Review: 02 Mar 2018 — Published (O): 05 May 2018
	Revised: None — Published (P): 05 May 2018

INTRODUCTION

Differences in cognition and achievement do not always favor one gender with a number of studies suggesting that males outperform females

on tests of visuospatial ability and mathematical reasoning, whereas females do better on tests involving memory and language use [1]. Numerous studies compare cognitive abilities of male and female students from elementary school through high school [2-4]. In this population, research comparing CR and SR tests indicates the presence of gender-related differences having to do with the mode of assessment [5-8]. In contrast, Becker and Johnston [9] and Dufresene et al [10] suggest that different aspects of knowledge are measured by MCQ and CR tests. Rodriguez [11] examined 67 empirical studies; this meta-analysis corroborates this inconsistency of findings. Few past studies, however, have evaluated performance of college level students [12-15].

Only limited studies have compared male and female student's performance in medical school. Men were reported to score better on the National Board of Medical Examiners (NBME) examination. Women performed better than men on clinical clerkships and also on clinical performance examinations [16, 17].

Summer enrichment programs: There are number of enrichment programs for matriculating medical or dental students sponsored by medical or dental schools and supported by government and private agencies. These programs aim to improve academic preparation and performance by first year medical or dental students. The summer program offered by the University of Medicine and Dentistry of New Jersey - New Jersey Medical School (UMDNJ-NJMS) provided a skill- and competence-oriented foundation in the anatomy curriculum. The students' learning and engagement with the materials were monitored and students received as-needed guidance from the faculty and education specialist to improve their study approach and time management.

Summer anatomy program set-up: The course was separated into 2 3-week units in which students dissected and studied, first, anatomy of the thorax and, second, of the abdomen. In addition to anatomy, students had physiology and biochemistry lectures, as well as exposure to clinical medicine. The UMDNJ-NJMS summer program concentrated primarily on anatomy, since medical students found it a difficult subject to master and most of them never had human anatomy in their undergraduate education. All the dissection sessions started with a 15-20 minute presentation that highlighted the clinical importance of the structures to be dissected and studied. A team of 4 students was assigned to dissect a cadaver. Overall, 12 hours was assigned for cadaver dissection per week and two teaching assistants (2nd year medical students) assisted the anatomy instructors.

This study is part of ongoing research conducted to examine test formats and their influence on medical student preparation for examination and performance on those examinations. We examined if there is a difference in the performances of male and female medical students on CR questions, compared to SR questions with the same stem, using responses from the students enrolled in the UMDNJ-NJMS summer anatomy program.

MATERIALS AND METHODS

Study population: Participants included incoming first-year medical students who were enrolled in a 6-week preparatory enrichment anatomy course during the 2 consecutive summers prior to the start of school year. Group 1 consisted of 16 students (8 male and 8 female) and Group 2 19 students (7 male and 12 female). Student ages ranged between 22 and 28 years. All students had a science background; 5 had postgraduate education and 1 had studied musculoskeletal anatomy during postgraduate education. This study received expedited IRB approval from UMDNJ.

Study design: We used a cohort study design for this study and, since the cohorts chosen are from separate, but similar populations, standardization of criteria and outcomes is possible. This study uses SR and CR testing schemes and is based on data collected on the above-described groups of students. Student's performance in the cadaver-oriented practical examination was also recorded.

This study was based on 2 types of testing: a CR test and an SR test and the examination questions included clinical vignettes. Similar question stems ensured all questions were the same (See appendix for sample items). Students

were made aware of the 2 test types and how to study for them. Two same-stem tests, each containing 30 questions that adequately represented the module content, were used. In the first test (CR), students read clinical vignettes and wrote short answers. The CR answer sheet was hand-graded using a previously-made answer key. The same-stem SR test questions were answered on a Scantron answer sheet that was computer-graded. The students had 45 minutes to complete each 30-question examination and were given a 5-minute break between examinations. They were not permitted to review any resources or consult each other during this break. For both modules, the students took the CR examination before the SR examination to avoid the 'cueing' effect. The questions were obtained from the departmental question bank and the degree of difficulty was the same for both modules in both years. The test questions for the 2 groups were different to avoid the teaching assistant remembering the past questions to help the class.

Data analysis: Paired t-tests were used to examine differences, so that individual shifts in score, if any, between the two testing modalities contributed to the comparisons. Thus, the resulting P values are for the comparison of open-ended and multiple-choice question examinations within each group and for the overall combined averages. All analyses were performed using SPSS Version 24 statistical package (IBM®).

RESULTS

Statistical analysis of individual groups and combined scores are summarized in Tables 1-3. Graphically, median scores with ranges are shown in Figure 1. The scatter plot (Figures 2A and 2B) shows the relationship between male and female student performances in CR and SR examination of thorax and abdomen modules.

Combined group 1 and 2 performance scores of male and female in CR and SR on thorax and abdomen examinations with range are described in the thorax CR test, male students performed a mean of 59.7 points (± 11.2, median 60.2, range 46-79); the female students performed a mean of 57.5 (± 12.0, median 56.5, range 43-73). In the SR test, male students scored a mean of 69.7 (± 10.3, median 68.2, range 63-88); the female students scored a mean of 70.7 (± 12.2, median 68.0, range 48-88). In the abdomen module CR test, male students performed a mean of 73.0 points (± 14.2, median 72.0, range 36-96); female students performed a mean of 71.3 (± 14.4, median 73.0, range 53-88). In the SR test, male students scored a mean of 81.6 (± 9.0, median 84.0, range 68-100); female students scored a mean of 80.5 (± 11.1, median 76.0, range 53-92).

The independent sample t-test (Table 2) shows no statistically significant differences in performances in either the thorax or abdomen modules by male and female students in the CR or SR examinations, except for the Group 1 male and female abdomen SR examination [t (14) = 1.934, p< 05 1 tailed].

Correlational analyses were used to examine the relationship between male and female student examination performance on CR and SR question examination types (Table 3). For examinations of the thorax module, results indicated a strong positive correlation between scores on CR examinations and SR examinations for all students, r (34) = .58, p <.001.; a strong positive correlation in female students on CR and SR examinations, r (19) =.66, p<.001; and no significant difference was found when examining the relationship between male student performance on CR and SR examination.

Table 1: Male and Female Student performance in the examinations.

Module	Thorax		Abdomen	
Test format	CR	SR	CR	SR
Group 1	Scores	Scores	Scores	Scores
Male (n=8)	57.4 ± 10.1	70.9 ± 11.0	71.0 ± 11.7	79.0 ± 11.0
Female (n=8)	55.0 ± 11.9	70.4 ± 11.8	68.6 ± 11.9	71.6 ± 9.1
Group 2				
Male (n=7)	62.4 ± 8.2	69.8 ± 8.4	75.0 ± 18.2	85.2 ± 11.6
Female (n=12)	59.1 ± 11.6	73.0 ± 11.8	73.0 ± 15.1	86.3 ± 8.7
Combined				
Male (n=15)	59.7 ± 11.2	69.7 ± 10.3	73.0 ± 14.2	81.6 ± 9.0
Female (n=20)	57.5 ± 12.0	70.7 ± 12.2	71.3 ± 14.4	80.5 ± 11.1

Table 2: Independent Sample t-test.

	Thorax examination			Abdomen examination		
	CR	MCQ	Difference score	CR	MCQ	Difference score
Overall male and female	ns	ns	ns	Ns	ns	ns
Group 1 male and female	ns	ns	ns	Ns	T (14) = 1.934, p<05 (1 tailed)	ns
Group 2 male and female	ns	ns	ns	Ns	ns	ns

Table 3: Correlation matrix.

	Thorax	Abdomen	CR examination	MCQ examination
	CR examination and MCQ examination	CR examination and MCQ examination	Thorax and abdomen	Thorax and abdomen
Overall	r = .58 (p<.001)- Strong positive correlation	r = .63 (p<.001)-Strong positive correlation	ns	ns
Males	ns	r = .73 (p<.001)-Very strong positive correlation	ns	ns
Females	r = .66 (p<.001)- Strong positive correlation	r = .57 (p<.001)-Strong positive correlation	ns	ns

Fig. 1: The graph showing the median scores and range of male and female student performances.

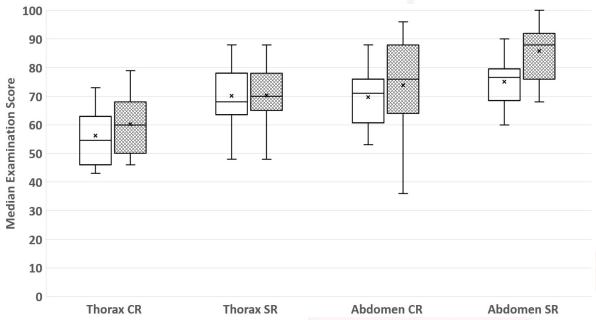


Fig. 2A: Scatterplot Results Thorax Exam.

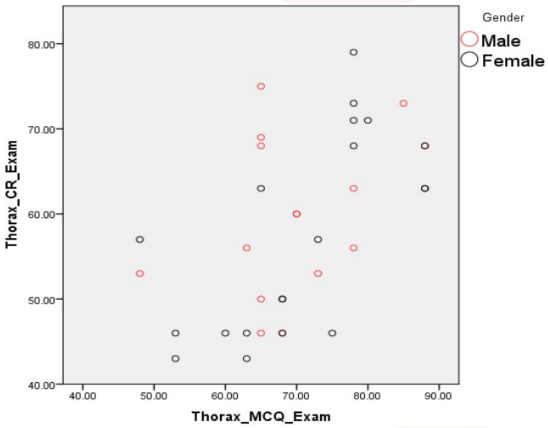
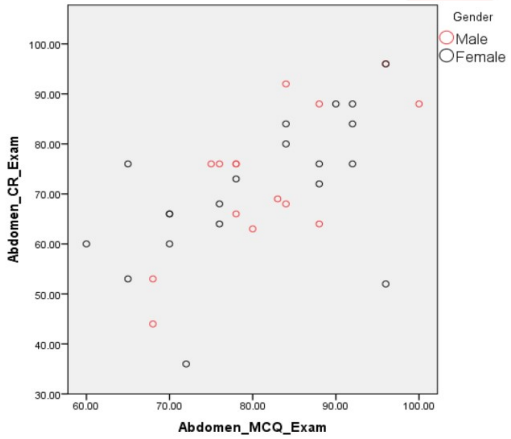


Fig. 2B: Scatterplot Results Abdomen Exam.



DISCUSSION

The result of the present study shows that the format of examination reflects how the student prepare for the test, supporting earlier observations [5, 12-14, 18, 19]. In the CR examination, there is no opportunity to guess the correct answer, since students had to create the answers [20]. This requires deeper understanding of the material. In a study involving third and fourth year medical students, it was found that, on same-stem SR and CR (un-cued) tests, students scored higher on the SR, compared to CR examinations [21]. This suggests that un-cued, same-stem CR examinations test students' real knowledge and give an additional discernment of the academically borderline students.

The results from the abdomen module (Table 1) support the idea that the students prepare based on the type of testing that they anticipate. Collectively, the results (Table 1) show that male and female students in both the groups performed better in the SR test, when compared to CR test, and both male and female students from groups 1 and 2 performed better in the abdomen module in both CR and SR examinations. The assessment data further suggests that female students who perform better on CR examinations also perform better on SR examinations related to the thorax. The correlational analysis depicted in the scatterplots (Figures 2A and 2B) clearly indicates the reasonableness of assuming a linear association between the variables.

Cognitive abilities between male and female students have been a focus of number of studies [2-4]. Many of these studies compared SR and CR tests and reported the presence of gender-related differences ascribed to the mode of assessment [5-8]. In contrast, Becker and Johnston [9] and Dufresene, et al [10] claim that different aspects of knowledge are measured with SR and CR tests. From the analysis of examination results, Weaver and Raptis [13] concluded that an examination of introductory atmospheric or oceanic science curricula, which

is made up of 60% multiple choice questions and 40% constructed response questions, did not skew to favor any particular gender. An analysis of mental test scores from 6 studies showed that, although average gender differences have been generally small and stable over time, the test scores of males consistently have larger variance. Except in tests of reading comprehension, perceptual speed, and associative memory, males typically outnumber females substantially among high-scoring individuals [12]. It is important to emphasize that the past studies were primarily directed at the performances of grade school students and few were conducted on the college level students [12-14].

The raw score analysis (Table 1) showed no appreciable difference either in the thorax or abdomen modules between the male and female students in the CR or SR examination. As expected, the results further showed that the SR scores of both the thorax and abdomen modules for both male and female student were higher, compared to CR examination (Table 1). Additionally, the CR score in the abdomen module of both male and female students was higher, compared to the thorax module. In the thorax module, Group 2 female students scored higher in the SR, compared to CR examination, while such an increase was only modest when the scores of male students were compared. There is some vague indication that male students have an advantage with SR, whereas others have found an advantage for males with CR questions [6, 18, 22]. Other gender-related claims include differences in "testwiseness" [23] and that male students are more willing to guess than female students [24].

Some factors possibly contributed to the results observed in the present study, when a comparison is made with high school and undergraduate college students. Medical students have four years of undergraduate education, have completed more advanced level science courses, and, in general, are highly proficient in an organized approach to learning. Their admission to medical school is based on high scholastic accomplishment. Hence, one would conceivably expect both male and female medical students to perform at an equal level, irrespective of the examination format, and that there may not be gender differences in test performance.

Limitations of the study: The small sample size is one of the limitations of this study that precludes a generalizability of the results. Additionally, this is the first study to evaluate the role of test format and student gender in examination performance in a medical school course that is more visuospatial related. This demonstrates a need for similar studies on a larger scale. Considering the advantages of the SR examination, while at the same time understanding the value of CR examination, the issue becomes how a medical school with larger class sizes can implement the latter approach. Lastly, this study was conducted on a select group of students in an enrichment program and the generalizability of the results needs additional evaluation, using a general population.

Future direction for medical educators: The study illustrates ways for medical educators to 1) adopt CR examination for formative evaluation early on in medical education for the students to identify their deficits and strategize effective learning; 2) use CR test format for identifying students who might benefit from academic advising and assistance from faculty and educational specialists early on in their medical education to avoid potential failure; and 3) with the availability of 'ExamSoft' software (ExamSoft, Dallas, TX) for creation, administration, and assessment, combined with Excel (Microsoft Corp., Redmond, WA), it is possible now to administer CR examinations for larger classes for summative assessment [25].

CONCLUSION

The present study demonstrates that, without changes in the curricular content or pedagogy changes in assessment, approaches can alter student learning behavior and appreciably increase specific knowledge, improving academic performances. This study also points to the fact that the gender poses no limitation to medical student examination performance, irrespective of the type of examination format.

ABBREVIATIONS

CR- Constructed Response

SR- Selected Response

Conflicts of Interests: None

REFERENCES

[1]. Downing K, Chan SW, Downing WK, Kwong T, Lam TF. "Measuring gender differences in cognitive functioning", Multicultural Education & Technology Journal 2008; 2: 4-18,

[2]. Maccoby, EE, Jacklin CN. The Psychology of Sex Differences. Stanford, Calif.: Stanford University Press. 1974.

[3]. Benbow, CP. "Sex Differences in Mathematical Reasoning Ability in Intellectually Talented Preadolescents: Their Nature, Effects, and Possible Causes." Behavioral and Brain Sciences 1988; 11: 169-232.

[4]. Mazzeo J, Schmitt AP, Bleistein CA. Sex-Related Performance Differences on Constructed Response and Multiple- Choice Sections of Advanced Placement Examinations Bulletin College Board Report No. 92-7, ETS RR No. 93-5, College Entrance Examination Board, New York. 1993.

[5]. Traub RE, MacRury K. Multiple-Choice vs Free-Response in the Testing of Scholastic Achievement. 6th Ed. Toronto, Ontario, Canada: Ontario Institute for Studies in Education. 1990.

[6]. Bolger, N., and T. Kellaghan. "Method of Measurement and Gender Differences in Scholastic Achievement." Journal of Educational Measurement. 1990; 27:165-74.

[7]. Murphy, RJL. "Sex Differences in GCE Examination Entry Statistics and Success Rates." Education Studies. 1980; 6: 169-78.

[8]. Murphy, R.J.L. "Sex Differences in Objective Test Performance." British Journal of Educational Psychology. 1982; 52:213-19.

[9]. Becker, W., & Johnston, C. The relationship between multiple choice and essay response questions in assessing economics understanding. Economic Record. 1999; 75, 348-357.

[10]. Dufresne RJ, Leonard WJ, Gerace W. J. Making sense of students' answers to multiple-choice questions. Phys Teach. 2002; 40:174-180.

[11]. Rodriguez, MC. Construct equivalence of multiple-choice and constructed response items: a random effects synthesis of correlations. Journal of Educational Measurement. 2003; 40:163-184.

[12]. Hedges LV, Nowell A. Sex differences in mental test scores, variability, and numbers of high-scoring individuals. Science. 1995; 269: 41-45.

[13]. Weaver AJ, Raptis H. Gender differences in introductory atmospheric and oceanic science exams: multiple choice versus constructed response questions. J Sci Educ Technol. 2001; 10:115–126.

[14]. Stanger-Hall KF. Multiple-choice exams: An obstacle for higher-level thinking in introductory science classes. CBE Life Sci Educ. 2012; 11:294-306.

[15]. Federer MR, Nehm RH and Pearl DK. Examining Gender Differences in Written Assessment Tasks in Biology: A Case Study of Evolutionary Explanations. CBE Life Sci Educ. 2016 Spring; 15(1):

[16]. Case SM, Becker DF, Swanson DB. Performances of men and women on NBME Part I and Part II: the more things change…. *Acad Med.* 1993; *68* (10 suppl): S25-27.

[17]. Cuddy MM, Swanson DB, Clauser BE. A multilevel analysis of examinee gender and USMLE step 1 performance. Acad Med. 2008; 83(10 Suppl): S58-62.

[18]. Simkin MG, Kuechler WL. Multiple-choice tests and student understanding: What is the connection? Decis Sci Journal Innovat Educ. 2005; 3:73-97.

[19]. Martinez ME. Cognition and the question of test item format. Educ Psychol. 1999; 34:207-218.

[20]. Lukhele R, Thissen D, Wainer H. On the relative value of multiple choice, constructed response, and examinee selected items on two achievement tests. J Educ Meas. 1994; 31:234-250.

[21]. Veloski JJ, Binowitz, HK, Robeson MR. A solution to the cueing effects of multiple-choice questions: The un-Q format. Med Educ. 1993; 27:371-375.

[22]. DeMars DE. Gender differences in mathematics and science on a high school proficiency exam: the role of response format. Appl Meas Educ. 1998; 11:279-299.

[23]. Zimmerman DW, Williams RH. A new look at the influence of guessing on the reliability of multiple-choice tests. Appl Psychol Meas. 2003; 27:357-371.

[24]. Ben-Shakhar G, Sinai Y. Gender differences in multiple-choice tests: the role of differential guessing tendencies. J Educ Meas. 2005; 28: 23-35.

[25]. Sam AH, Hameed S, Harris J, Meeran K. Validity of very short answer versus single best answer questions for undergraduate assessment. BMC Med Educ. 2016; 16:266–269.

How to cite this article:

Cheryl Melovitz-Vasan, Matthew Gentile, Susan Huff, and Nagaswami Vasan. A STUDY OF TEST CONSTRUCT AND GENDER DIFFERENCES IN MEDICAL STUDENT ANATOMY EXAMINATION PERFORMANCE. Int J Anat Res 2018;6(2.2):5250-5255. **DOI:** 10.16965/ijar.2018.171

Appendix: Sample questions:

Thorax: CR question:

CASE: A patient complains of tiring easily and shortness of breath on exertion. Auscultation of the chest showed a diastolic heart murmur, heard best at the apex of the heart, and the presence of a collapsing pulse (forcible pulse that rapidly diminishes).

Q. Which heart condition best explains the 'collapsing pulse'?

Answer: Aortic valve insufficiency.

SR question:

CASE: A patient complains of tiring easily and shortness of breath on exertion. Auscultation of the chest showed a diastolic heart murmur, heard best at the apex of the heart, and the presence of a collapsing pulse (forcible pulse that rapidly diminishes).

Q. Which of the following conditions best explains the 'collapsing pulse'?

A. Tricuspid disease.

B. Bicuspid disease.

C. Aortic valve insufficiency.

D. Coarctation of the aorta.

E. SA node problem.

Answer: C

Abdomen: CR question:

CASE: A 62-year-old woman has a venogram that shows a blockage in the left renal vein where it crosses anterior to the aorta. Additional imaging studies indicate that the occlusion is caused by an aneurysm of an artery that lies immediately anterior to the vein.

Q. In which artery that the aneurysm is most likely located?

Answer: Superior mesenteric

SR question:

CASE: A 62-year-old woman has a venogram that shows a blockage in the left renal vein where it crosses anterior to the aorta. Additional imaging studies indicate that the occlusion is caused by an aneurysm of an artery that lies immediately anterior to the vein.

Q. The aneurysm is most likely located in which of the following arteries?

A. Celiac

B. Superior mesenteric

C. Left colic

D. Inferior mesenteric

E. Middle colic

Answer: B