

A Quantitative Study Exploring the Relationship of Gender to the Development of Clinical Reasoning Skills Among Medical Students

Khin-Htun, S.^{1*} & Than, H.²

¹Medical Education Fellow and Honorary Assistant Professor, University of Nottingham, UK

²Trust Grade Doctor, Queen's Medical Centre, Nottingham University Hospitals, Nottingham, UK

***Correspondence to:** Dr. Swe Khin-Htun, Medical Education Fellow and Honorary Assistant Professor, University of Nottingham, UK.

Copyright

© 2019 Dr. Swe Khin-Htun, *et al.* This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: 15 January 2019

Published: 22 February 2019

Keywords: *Curriculum Model; Clinical Reasoning; Medical Education*

Abstract

The development of clinical reasoning (CR) skills in the medical students can be influenced by many factors. One of the attributes which is suggested to play an important role in CR skills is gender.

To ascertain further insight into this theory, the marks of the CR questions in the summative knowledge exam papers of third year, i.e., clinical phase 1 (CP 1), and final year, i.e., Clinical Phase 3 (CP 3) medical students from the University of Nottingham (UoN) were reviewed.

Data on how well the male and female students performed on CR questions were collected and analyzed across three categories of school years (2012, 2013 and 2014) for CP 1 and (2014, 2015 and 2016) for CP 3.

In summary, the results of CR scores of summative written exams for both genders showed no significant difference in both CP 1 and CP 3 datasets for three cohorts.

In conclusion, the gender has no impact on the development of CR in medical students.

Abbreviations (if used)

CR: Clinical reasoning

CP1: Clinical phase 1, the first clinical phase

CP3: Clinical phase 3, the third and final clinical phase

UoN: The University of Nottingham

Introduction

The Clinical Reasoning (CR) is defined as a recursive, multidimensional, and complex process which involves informal and formal strategies for the analysis of patients' information and evaluation of its information [1]. Medical errors as an effect of faulty reasoning lead to mortality and morbidity of the patients and care users [2]. CR skills are commonly acquired through experiential learning of Undergraduate (UG) students within the course of undergoing the clinical curriculum. However, CR development has not been formally taught in the past.

The development of Clinical Reasoning (CR) can be influenced by many factors. A research conducted by Karwowski, Gralowski, and Szumski (2015) [3] revealed that CR is different with respect to the gender of students. The male students have been found having better CR skills than their female student counterparts have [4]. In a similar study, gender also played a role in rationality of decision-making, wherein male students are more rational decision-makers as compared to female counterparts [5]. Although in some circumstances the female students outdo their male counterparts, the male students have a better understanding of the perspectives of CR. The female students show more interest in learning the ideas, concepts, theories, and contents of clinical curriculum, but the males are observed to have more potential to implement all the concepts and potentials in their real and professional life [3]. Baron-Cohen *et al.* (2015) [4] argued that perception of the nursing curriculum, contents, practices, theories, and concepts differs with respect to the gender of students. Sometimes, the male students show more compatibility and potential in developing the CR than the females. Moreover, Siani and Assaraf (2016) [6] claimed that psychological characteristics of male and female students differ from each other with respect to learning and development of CR skills. These findings are supported by the study of Doane, Kelley and Pearson (2016) [7].

CR Problems (CRP) and Diagnostics Thinking Inventory (DTI) were used to assess the gender effect on CR ability of medical students in a study conducted by Groves, O'rourke and Alexander (2003) [8]. The result of the study showed that female students were found to have better ability only for CRP, but not for DTI although they were a positive predictor of CR ability. This is likely because of their more careful and thorough approach in diagnosis, which results in deliberate identification of all critical features from case

presentation [8]. Not only female medical students, but also female nursing students showed better ability in resolving problems and facing the critical situation [9].

On the other hand, in medical and dental literature, it is showed that gender has no influence in the ability of CR in Undergraduate (UG) students [10-15].

In a lot of research, CR is compared between male and female students using experimental and research assessment methods. However, the unique aspect of our study is the use of multipurpose standardized assessment tools (in this case, we use summative written exam papers). They enable assessment across a predetermined set of problems and can sample a broad set of content domains within limited testing time.

Materials and Methods

The research question is “Does gender influence outcomes as measure by summative written exams?”

There are different types of assessments of CR such as, experimental and research methods, standardized assessments and workplace-based assessments. Standardized assessment tools are divided into multipurpose assessment tools and tools with specific aims to assess CR. Multipurpose assessment tools include standardized examinations such as written papers or objective structured clinical examinations (OSCE). They enable assessment across a predetermined set of problems, but take place in an artificial environment and do not assess all aspects of CR.

In this study, standardized knowledge written exam papers were used to compare the effect of gender. The Clinical phase 1 (CP1) students and The Clinical phase 3, CP3 students have to sit summative knowledge paper at the end of their clinical phase. CP1 has one knowledge paper but CP3 has 2 knowledge papers. The performance on the portion of these summative knowledge papers which are predominantly CR questions of CP 1 and CP3 students were compared to see the effect of gender on the development of CR skills. There are three cohort groups in each stage: CP 1 (2012, 2013 and 2014) and CP3 (2014, 2015 and 2016). As study started in 2011 and ended in 2016, the average number of students per each cohort is between 318-351 [Table 1].

Table 1: Number of Participants in CP1 and CP3 by Cohort

CP1	Students in Cohort	CP3	Students in Cohort
2012	351	2014	335
2013	344	2015	350
2014	327	2016	318

Questions Classification: Before the Exam

Firstly, the summative knowledge papers are reviewed and categorized into CR or no-CR questions during standard setting meetings, which are conducted by 15-25 experts including specialists from different specialties such as gastroenterology, respiratory medicine, general surgery as well as GPs, Director of clinical skills, clinical teaching fellows, module leads, medical educators and some junior doctors. We have invited all the doctors, educators from the different trust where University students are clinically attached to. It was led by the Director of assessment. The number of the raters varied from 25-35. Their specialities are varied, and their levels are varied from F1 to clinical lead by each occasion. Mainly, it is mandatory for very experienced clinical lead, course directors and education fellows (who have explicit education role in trust) to attend most of the meetings. We want variety so we open up for other clinicians too. If any discrepancies arise regarding questions whether they are CR or non-CR during standard setting meetings, the team not only takes Bloom’s taxonomy of learning domains into account, but also maps these questions against the three statements 8c, 8g and 14f on ‘Outcomes for Graduates from Tomorrow’s Doctors’ published by GMC to categories the questions. Then, these questions are discussed until mutually agreed. Only the questions that assess the third category (apply) to the sixth category (evaluate) of Bloom’s cognitive processes are accepted as CR questions.

For example, in the first question (Table 2) the students can answer this by using their knowledge alone but in the second set of questions (Table 3), the students need to mobilise their knowledge and apply their knowledge in the clinical context to answer the questions so we assume only these types of questions as CR questions and used to compare in this study.

Table 2: Example of non-CR questions (knowledge only questions)

The following are all arterial blood gas results taken from patients with shortness of breath. For each blood gas result, pick the correct interpretation from the drop-down list.
 Note: Normal ranges: PH 7.35-7.45; pO2 11-14kPa; pCO2 4.5-6kPa; Bicarb 22-27mmol/L; Base excess -2/+2

A. Metabolic Acidosis B. Metabolic alkalosis C. Mixed acidosis	D. Mixed alkalosis E. Respiratory acidosis F. Respiratory alkalosis
--	---

i. PH 7.25, pO2 16.3, pCO2 2.6, Bicarb 9, Base excess -15, FiO2 0.8
 ii. PH 7.52, pO2 17.3, pCO2 2.0, Bicarb 26, Base excess +1, FiO2 0.21
 iii. PH 7.21, pO2 6.9, pCO2 7.8, Bicarb 35, Base excess -2, FiO2 0.6
 iv. PH 7.21, pO2 8.36 pCO2 6.9, Bicarb 15, Base excess -8, FiO2 0.8
 (4 marks)

Table 3: Example of CR questions

All of the following patients have cirrhosis of the liver. Match the clinical vignette to the appropriate diagnosis.

<p>A. Alpha 1 anti-trypsin deficiency</p> <p>B. Autoimmune hepatitis</p> <p>C. Chronic alcoholism</p> <p>D. Gallstone disease</p> <p>E. Haemochromatosis</p> <p>F. Hepatitis A</p> <p>G. Hepatitis C</p>	<p>H. Hepatocellular carcinoma</p> <p>I. Non-alcoholic steato-hepatitis</p> <p>J. Pancreatic cancer</p> <p>K. Paracetamol overdose</p> <p>L. Primary biliary cirrhosis</p> <p>M. Primary sclerosing cholangitis</p> <p>N. Wilson’s disease</p>
--	--

i. A 48-year old man with new onset diabetes mellitus, two knee replacements, two hip replacements and a shoulder replacement who has a very high ferritin.

ii. A 32-year old reformed intravenous user of illicit drugs.

iii. A 58-year old woman positive for anti-mitochondrial antibodies.

iv. A 42-year old non-smoker who was diagnosed by his GP with chronic obstructive pulmonary disease (COPD) for 5 years ago.

v. A 63-year old obese, diabetic man of Indian ethnic origin.

(5 marks)

In the final data, CR questions can cover: being given a history and being asked to formulate the diagnosis for each case; being given physical findings and being asked to choose the most likely diagnosis; being given investigation results and being asked to find a diagnosis and provide the treatment plan; being given a diagnosis and being asked to choose the matching case vignette or history; being given a history and being asked to match the investigation findings to the interpretation of the findings.

Questions Classification: After the Exam

For each exam paper, routine psychometric analysis is carried out to ensure a high-quality assessment tool. Classical Test Theory (CTT) and Item Response Theory (IRT) are used to conduct the analysis of ‘the post examination psychometric data’. Problematic test items which are either too difficult or too easy are identified by using Student-item maps. Test-score reliability (Cronbach’s alpha), item discrimination index (ID) and standard error of measurement (SEM) are also used to analyses knowledge papers.

In addition, for each paper, not only frequency and discrimination (U-L), but also learning objective analysis are used to inspect. Item difficulty (p) and discrimination value (d) are used to calculate for each item. If the items have low discrimination value (d <0.15) and high level of difficulty (p <0.2), they are excluded. Generalizability (G) theory (student x item) is used to measure the reliability of the test. Furthermore, descriptive statistics (sample characteristics according to gender/course) and item analysis (item

discrimination, generalizability and decision studies) are used to check for each item. Whether there is any correlation between cases and mean marks are also considered.

External and internal examiners review the paper carefully. If there is any comment for some questions, those comments are considered carefully to decide whether further action is needed to be taken. The final scores are provided after these steps are taken. The following tables provide the overall results for the CP1 and CP3 (table 4 and table 5).

Table 4: Ranges of CP1 marks from 2012 to 2014

CP1	Marks			
	CR	NCR	Total	CR % of Total
2012	85	101	186	46
2013	73	112	185	39
2014	116	79	195	59

Table 5: Ranges of paper 1 and paper 2 of CP3 marks from 2014 to 2016

CP3	Marks Paper 1				Mark Paper 2			
	CR	NCR	Total	CR % of Total	CR	NCR	Total	CR % of Total
2014	151	41	192	79	87	83	170	51
2015	85	85	170	50	91	79	170	54
2016	113	57	170	66	98	77	175	56

These exam papers are not designed specifically for this research but the researcher used the opportunity to review these exam papers to identify CR questions and compare the effect on gender so there are many data sets with different numbers of students, different summative knowledge exam papers and different components of CR marks in each cohort. The nature of the data we have, we are conducting an independent sample *t*-test wherein we are comparing the group means instead of conducting a pairwise analysis of the difference of scores. The total raw scores reflect the weight of the correct answers.

A level of significance of 0.05 was also used in the independent sample *t*-test.

The ethical committee confirmed that the approval from the department was not required as this kind of project is classified as service evaluation.

Results

Normality

An independent sample *t*-test was conducted to address the research objectives. Therefore, normality testing was conducted by investigation of the skewness and kurtosis statistics and histogram to check the distribution of data of the different dependent variable. To determine whether the data follows normal distribution, skewness statistics greater than three indicate strong non-normality and kurtosis statistics between 10 and 20 also indicate non-normality [16]. As can be seen in Table 4 and 5, the range of values of the skewness (-0.35 to 0.63) and kurtosis (-0.79 to -0.13) statistic for the dependent variables of CR scores, NCR scores, and total scores of summative written exam for the CP1 dataset; and also the range of values of the skewness (-0.30 to 0.26) and kurtosis (-1.02 to 0.49) statistic for the dependent variables of CR scores, NCR scores, and total scores of summative written exam for papers 1 and 2 for the CP3 dataset were in the acceptable range which showed that all the data of these dependent variables exhibited normal distribution. Thus, the parametric statistical analyses can be conducted [Table 6,7]. According to the nature of our study, we are not comparing each year but we put all CP1 from 3 cohorts and all CP3 from 3 cohorts together and that accounts 577 female students and 441 male in CP1 [Table 8]. For CP3, 574 female and 425 male in total are involved in this study [Table 9].

Table 6: *Skewness and Kurtosis Statistics of CR Scores, NCR Scores, and Total Scores of Summative Written Exam for CP1 Dataset*

	N	Skewness		Kurtosis	
	Statistic	Statistic	Std. Error	Statistic	Std. Error
CR	1022	0.63	0.08	-0.15	0.15
NCR	1020	0.04	0.08	-0.79	0.15
Summative written exams	1021	-0.35	0.08	-0.13	0.15

Table 7: *Skewness and Kurtosis Statistics of Scores of CR Scores, NCR Scores, and Total Scores of Summative Written Exam for CP3 Dataset*

	N	Skewness		Kurtosis	
	Statistic	Statistic	Std. Error	Statistic	Std. Error
CR Score (Paper 1)	1003	0.26	0.08	-1.02	0.15
NCR Paper 1, 41/192	1003	0.10	0.08	-1.02	0.15
Summative written exams (Paper 1)	1003	-0.28	0.08	0.49	0.15
CR (Paper 2)	1003	-0.11	0.08	-0.01	0.15
NCR paper 2, 98/185	1003	0.23	0.08	-0.35	0.15
Summative written exams (Paper 2)	1003	-0.30	0.08	0.01	0.15

Table 8: Descriptive Statistics Summaries of CR Scores by Gender for CP1 Dataset

	Gender	N	Mean	Std. Deviation	Std. Error Mean
CR Score	0 Female	577	63.63	13.84	0.58
	1 Male	441	65.17	14.09	0.67

Table 9: Descriptive Statistics Summaries of CR Scores by Gender for CP3 Dataset

	Gender	N	Mean	Std. Deviation	Std. Error Mean
CR Score (Paper 1)	0 Female	574	86.13	21.52	0.90
	1 Male	425	84.79	20.89	1.01
CR Score (Paper 2)	0 Female	574	67.40	9.08	0.38
	1 Male	425	67.31	9.15	0.44

In order to prevent data overload, only CR questions are compared and NCR questions are out of our study

Does Gender Influence Outcomes as Measured by Summative Written Exams?

An independent sample *t*-test was conducted to determine whether the gender influence outcomes as measured by the CR score in the summative written exam. Also, different analyses were conducted on the dataset of CP1 and CP3.

Table 10 summarized the results of the independent samples *t*-test for the difference of the CR score in the summative written exams by gender in the dataset of CP1. The results of independent sample *t*-test showed that CR score of the summative written exam ($t(1016) = -1.75, p = 0.08$) were not significantly different between males and females. Table 11 summarized the results of the independent samples *t*-test for the difference of the CR score in the summative written exams by gender in the dataset of CP3. The results of independent sample *t*-test also showed that CR scores of the summative written exam in paper 1 ($t(997) = 0.99, p = 0.32$) and paper 2 ($t(997) = 0.15, p = 0.88$) were both not significantly different between males and females. With this result, the null hypothesis for research question three that there is no significant effect from gender on outcomes as measured by summative written exams was not rejected.

Table 10: Independent Sample t-test Results of CR Scores by Gender for CP1 Dataset

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
CR Score	Equal variances assumed	0.004	0.95	-1.75	1016	0.08	-1.54	0.88	-3.27	0.19

Table 11: Independent Sample t-test Results of CR Scores by Gender for CP3 Dataset

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
CR Score (Paper 1)	Equal variances assumed	0.51	0.48	0.99	997	0.32	1.35	1.36	-1.32	4.01
CR Score (Paper 2)	Equal variances assumed	0.11	0.74	0.15	997	0.88	0.09	0.58	-1.06	1.23

Discussion

The research question asked “Does gender influence outcomes as measured by summative written exams?” The results of the data analysis revealed that there were no significant difference between males and females in terms of outcomes as measured by summative written exams.

Based on the literature, gender differences in CR have been found, with men exhibiting higher CR. For instance, Karwowski *et al.* (2015) [3] found that the development of CR is different with respect to the gender of students. The male students have been found having better CR skills than their female student counterparts have [4]. In a similar study, gender also played a role in rationality of decision-making, wherein male students are more rational decision-makers as compared to female counterparts [5]. These studies did not support the results of current study, revealing no gender differences in the CR of men and women.

One of the unique results that came out from this study was the lack of gender differences in the CR of male and female students in terms of improvements in their CR. This means that no curricular modifications may be necessary to be taken into consideration to take into account possible gender differences in the development of CR. This finding is a contrast to previous studies suggesting that CR tends to be different between men and women [3,4].

The recommendation is to formulate an assessment plan or rubric that captures the conceptual components of CR. The generation of an assessment rubric of CR can facilitate better redesigning of courses that would target specific weaknesses based on student outcomes. Assessing students for CR should come from many sources because it consists of multiple traits such as knowledge, skills, problem solving etc., and each trait is best assessed by a specific tool [17-20].

Limitation

There is no gold standard for the assessment of CR. The data in this paper uses purely summative knowledge papers as a proxy measure; all the different aspects of CR could not be assessed. There are many data sets with different numbers of students, different summative knowledge exam papers and different components of CR marks in each cohort. The nature of the data we have, we are conducting an independent sample *t*-test wherein we are comparing the group means instead of conducting a pairwise analysis of the difference of scores. The total raw scores reflect the weight of the correct answers.

Conclusions

In conclusion, it can be stated that both genders have potential to develop the CR skills. Some studies provide enough evidence about the effective development of CR amongst male students, whereas, others favor that the female students have more potential than the male students do. Meanwhile, some show lack of effect of gender on development of CR. All these possible scenarios clearly demand more research in order to understand the extent to which gender can influence on the development of CR in the learning environment and/or decision making in real world practice.

References

1. Christensen, N., Black, L., Furze, J., Huhn, K., Vendrely, A. & Wainwright, S. (2016). Clinical Reasoning: Survey of Teaching Methods, Integration, and Assessment in Entry-Level Physical Therapist Academic Education. *Physical therapy*, 97(2), 175-186.
2. Drain, P. K., Mock, C., Toole, D., Rosenwald, A., Jehn, M., et al. (2017). The emergence of undergraduate majors in global health: systematic review of programs and recommendations for future directions. *The American journal of tropical medicine and hygiene*, 96(1), 16-23.
3. Karwowski, M., Gralewski, J. & Szumski, G. (2015) Teachers' effect on students' creative self-beliefs is moderated by students' gender. *Learning and individual differences*, 44, 1-8.

4. Baron-Cohen, S., Bowen, D.C., Holt, R.J., Allison, C., Auyeung, B., *et al.* (2015). The “reading the mind in the eyes” test: complete absence of typical sex difference in ~400 men and women with autism. *PLoS One*, *10*(8), e0136521.
5. Courteille, O., Josephson, A. & Larsson, L.O. (2014). Interpersonal behaviors and socioemotional interaction of medical students in a virtual clinical encounter. *BMC medical education*, *14*(1), 64.
6. Siani, M. & Assaraf, O. B. Z. (2016). The moral reasoning of genetic dilemmas amongst Jewish Israeli undergraduate students with different religious affiliations and scientific backgrounds. *Journal of genetic counseling*, *25*(3), 596-609.
7. Doane, A. N., Kelley, M. L. & Pearson, M. R. (2016). Reducing cyberbullying: A theory of reasoned action-based video prevention program for college students. *Aggressive behavior*, *42*(2), 136-146.
8. Groves, M., O’rourke, P. & Alexander, H. (2003). The association between student characteristics and the development of clinical reasoning in a graduate-entry, PBL medical programme. *Medical Teacher*, *25*(6), 626-631.
9. Brosnan, M., Chapman, E. & Ashwin, C. (2014). Adolescents with autism spectrum disorder show a circumspect reasoning bias rather than ‘jumping-to-conclusions’. *Journal of autism and developmental disorders*, *44*(3), 513-520.
10. Anderson, K. (2006). *Factors Affecting the Development of Undergraduate Medical Students’ Clinical Reasoning Ability*. PhD dissertation University of Adelaide.
11. Staff, C. N. (2014). *Module Nine: Factors affecting clinical reasoning* [Online].
12. Ashoorion, V., Liaghatdar, M. & Adibi, P. (2012). What variables can influence clinical reasoning? *Journal of Research in Medical Sciences*, *17*(12), 1170-1175.
13. Da Silva, A. (2013). *Clinical Reasoning Development in Medical Students: An Educational and Transcultural Comparative Study*. PhD Dissertation. The University of Nottingham.
14. Shafaroodi, N., Kamali, M. & Parvizy, S. (2014). Factors affecting clinical reasoning of occupational therapists: a qualitative study. *Medical Journal of the Islamic Republic of Iran*, *28*, 1-8.
15. Nafea, E. (2015). *CR in dental students: A comparative cross-curricular study* Doctoral Dissertation. University of Nottingham.
16. Kline, T. J. B. (2005). *Psychological testing: a practical approach to design and evaluation*. Thousand Oaks, Calif. London, Sage Publications.
17. Schuwirth, L. (2009). Is assessment of clinical reasoning still the Holy Grail? *Medical Education*, *43*, 298-300.

-
18. AlHaqwi, A. I. (2014). Learning outcomes and tutoring in problem based-learning: how do undergraduate medical students perceive them? *International journal of health sciences*, 8(2), 125-132.
 19. Fingeret, A. L., Martinez, R. H., Hsieh, C., Downey, P. & Nowygrod, R. (2016). Watch what happens: using a web-based multimedia platform to enhance intraoperative learning and development of clinical reasoning. *The American Journal of Surgery*, 211(2), 384-389.
 20. Gross, M., Moore, Z. E., Gardner, F. L., Wolanin, A. T., Pess, R. & Marks, D. R. (2016). An empirical examination comparing the Mindfulness-Acceptance-Commitment approach and Psychological Skills Training for the mental health and sport performance of female student athletes. *International Journal of Sport and Exercise Psychology*, 16(4), 1-21.