



The Critical Thinking Skills of Practicing Family Physicians: A Population-Based Cross-Sectional Study

David Ross, MD; Michelle Morros, MD; Efrem Violato, MSc

BACKGROUND AND OBJECTIVES: Critical thinking (CT) skills are an important aspect of clinical reasoning and diagnosis. The goals of this study were to (1) examine levels of CT skills of practicing family physicians, (2) compare the CT skills of practicing family physicians to family medicine residents, and (3) identify individual variables and practice characteristics predictive of CT skills. .

METHODS: We used a population-based, cross-sectional design to compare practicing and resident family physicians and examine the predictors of CT skills in practicing family physicians. Sixty-two practicing family physicians were recruited across Canada. We used data from 59 family medicine residents at a single institution in Canada. We used the California Critical Thinking Skills Test (CCTST) to measure CT skills. We analyzed data using descriptive and univariate analysis, multivariate analysis of variance, and hierarchical multiple linear regression. CT skills were further examined in follow-up analysis using polynomial regression.

RESULTS: Residents performed better than practicing physicians on nearly all aspects of CT ($P<.005$). Age was the strongest predictor of CT skills in practicing physicians ($P<.005$); CT skills declined with age as a quadratic function ($P<.005$).

CONCLUSIONS: As a group, practicing family physicians exhibited lower scores on the CCTST compared to family medicine residents. CT skills showed a decline with age, accelerating after approximately age 60 years. The results of the study have implications for continuing education and assessment of physicians' clinical skills. Further research is required to better understand what other predictors may be important for CT skills of practicing family physicians.

(Fam Med. 2020;52(9):635-41.)

doi: 10.22454/FamMed.2020.271478

Cognitive errors due to cognitive and affective biases, faulty or inaccurate reasoning, and misunderstanding probabilities can lead to diagnostic errors and patient harm, and are some of the most complex issues surrounding patient safety.¹ To begin to ameliorate

some of these biases and errors, it has been suggested that education, beginning in medical school, teach physicians analytical and rational thinking.^{2,3} While the processes of clinical reasoning (CR) are debated,⁴ it is generally accepted that the analytical and rational aspects

of CR consist of the ability to identify a problem, select and evaluate pertinent information, recognize assumptions, formulate appropriate hypotheses, and draw valid conclusions and critical inferences.⁵⁻⁸ These aspects of CR are proximal to scientific thinking and the cognitive and metacognitive processes and dispositions to make meaning of information.⁹ Scientific thinking is a domain of the broader construct of critical thinking (CT),¹⁰ with CT being essential for CR. CT consists of six skills: interpretation, analysis, evaluation, inference, explanation, and self-regulation.^{11,12}

CT skills serve as a good predictor of academic success in health care,¹³⁻¹⁶ yet it is unknown what the level of CT skills of practicing physicians are and how continued development of CT skills can be supported. Critical thinking is essential to the safe and effective practice of family medicine, as family physicians (FP) must diagnose and manage numerous undifferentiated conditions¹⁷ in complex systems¹⁸ that require a broader general knowledge base than those in a specialist field.^{19,20} As the initial point of contact with a patient, it is important for a FP

From the Department of Family Medicine (Drs Ross and Morros), and Department of Educational Psychology (Mr Violato), University of Alberta, Edmonton, Canada.

to be able to engage CR bolstered by skillful CT to make an initial diagnosis and pursue an appropriate management plan.^{17,21} Deficits in CT skills can contribute to the inability to make effective clinical decisions when faced with clinical uncertainty, a situation with potentially devastating consequences.²²

Previously, CT skills have been examined in residents, showing no change between entering and completing residency.²³ The next step is to examine CT skills in practicing family physicians. Currently, there is limited knowledge regarding the level of CT skills of practicing physicians, and FPs specifically. It will be important to examine if change in CT occurs in practice when compared to residency and understand what potential variables may be predictive of CT skills in practicing FPs.

To better understand the CT skills of practicing FPs, and potential changes in CT, we designed a population-based, cross-sectional observational study. Three aims for the present research were developed: (1) examine the level of CT skills of practicing family physicians using the California Critical Thinking Skills Test (CCTST), (2) compare the CT skills of practicing family physicians with the CT skills of family medicine residents using a cross-sectional design to determine group differences, and (3) identify which individual variables and practice characteristics are predictive of CT skills.

Methods

Participants

We recruited family physicians from Alberta, British Columbia, Ontario, and Quebec through professional list serves, at professional conferences, and by site leads for primary care networks. Family physicians were recruited between May and December 2019. An initial contact email was distributed; interested participants were sent a consent form and a link to the CCTST. Participants completed the test on personal devices. Completion of the test ended participants' involvement in the

study. Data collected from residents in June 2013 at the end of a residency program at a single institution (previously presented by Ross et al²³) were used for comparison purposes. Research Ethics Board 2 at the University of Alberta granted ethics approval for this study. We used STROBE cross-sectional reporting guidelines.²⁴

Measures

We used the CCTST test as a measure of critical thinking skills. This 34-item test demonstrates strong validity evidence, is predictive of academic success in health professions, and has been used extensively in research in health professions.^{15,16,25,26}

The CCTST is comprised of an overall score and five subscales, each measuring a different aspect of CT.^{11,12} The overall score describes strength in using reasoning to form reflective judgments about what to believe or what to do (scored: 0-35). Analytical skills are used to identify assumptions, reasons, themes, and the evidence used in making arguments or offering explanations (scored: 0-8). Inference skills are the ability to draw conclusions from reasons, evidence, observations, experiences, or values and beliefs (scored: 0-15). Evaluative skills are used to assess the credibility of the claims people make and to assess the quality of the reasoning people display when they make arguments or give explanations (scored: 0-10). Inductive reasoning relies on estimating likely outcomes and decision making in contexts of uncertainty (scored: 0-15). Deductive skills are used to determine the precise logical consequences of a given set of rules, conditions, beliefs, values, policies, principles, procedures, or terminology (scored: 0-15).²⁷ The CCTST requires approximately 45 minutes to complete.

Analysis

Data were analyzed using SPSS Version 25.²⁸ We used descriptive statistics to examine practicing FP CCTST scores and demographic characteristics, and we conducted univariate

analysis on categorical variables. The primary analysis consisted of two steps. First, we conducted a one-way multivariate analysis of variance (MANOVA) to determine differences between practicing FPs and residents on the CCTST and its subscales. Second, we conducted a hierarchical multiple linear regression with the practicing physicians' data to determine predictor variable importance for CT skills in practicing physicians.

Assumption Checks. Assumptions for MANOVA were checked. Examination of Q-Q plots and Shapiro-Wilks test indicated multivariate normality. Residual plots and Q-Q plots identified a borderline outlier. One resident was an outlier on percentile rank and the induction subscale, though not on other scores; we decided to remove the participant. We identified high multicollinearity among the variables, subscale relationships were examined, and we removed the analysis subscale to carry the test forward.²⁹ Box's test indicated homogeneity of covariance matrices, $\chi^2(10), P=.460$. We checked assumptions for regression; residual plots and Cooks Distance indicated there were no outliers. We identified no violations for the Durbin-Watson test for autocorrelation, variance inflation factor for collinearity, and Q-Q plots and measures of skewness for normality. We deemed the sample size adequate for the number of participants per predictor included in the model.²⁹

Results

Demographics

We recruited 62 practicing family physicians. The sample was composed primarily of participants of Caucasian (41, 66%) and Asian (14, 23%) backgrounds. Thirty-six of the participants were female (58%). The majority of the sample worked in full service/broad scope family practice (33, 53%), followed by solely office-based family practice (11, 18%). Sixty percent (38) of the sample worked in population centers with >500K

people; 10% (6) worked in centers of 50K-100K people, 20% (12) worked in centers with 10K-50K people, and 10% (6) worked in centers of <10K people. There were 53 Canadian medical graduates, eight international medical graduates, and one Canadian who had studied abroad. Twenty-seven percent of the sample had a graduate degree; two PhDs, and 15 master's degrees. Participants engaged in a wide range of continuing medical education (CME) with the majority indicating their primary form of CME was conferences (28) and self-study (16). For further sample characteristics see Table 1.

Secondary data from 60 residents was available from the Ross study.²³ Thirty-five (56%) of the residents were female with a mean (SD) age of 27.9 (5.1) years.

Initial Analysis

Family medicine residents and practicing FPs exhibited a wide distribution of scores on overall CCTST score and the subscales. Residents tended to perform better than practicing physicians on all measures, except for the analysis subscale (Table 2) and scored higher on mean percentile rank (Table 2). Percentile rank is a norm referenced score based on an aggregated sample of test takers.³⁰ We conducted univariate analysis on the categorical demographic variables, and found no significant differences for overall CCTST score or the subscales on any categorical variables. The second highest scoring participant, scoring in the 97th percentile, was 64 years old; only one resident scored higher, scoring in the 98th percentile.

Main Analysis I: MANOVA—Comparison of Practicing Family Physicians and Residents

We ran a one-way MANOVA (independent variable: practicing physicians vs residents and dependent variables: inference, evaluation, induction and deduction). Results indicated there was a significant difference between residents and practicing family physicians on the subscales of the CCTST ($F [4,116]=4.75$; Wilk's $\lambda=.86$; $P<.001$).

Follow-up univariate tests identified significant differences on the evaluation subscale ($F [1,119]=14.56$; $P\leq.001$), and induction subscale ($F [1,119]=9.92$; $P=.002$; Table 2). There were no significant differences on the analysis subscale ($F [1,119]=.20$, $P=.659$), but there were differences on the overall CCTST score ($F [1,119]=5.02$, $P<.027$); and percentile rank ($F [1,119]=4.97$, $P<.028$). Subscale effect sizes indicate small to large effects for the subscales and overall score, except for analysis.

Main Analysis II: Hierarchical Multiple Linear Regression—Predictor Variables of CT Skills in Practicing Family Physicians

We used only data from practicing FPs for a series of hierarchical multiple regressions with overall CCTST score and each subscale as dependent variables. We used age, clinical days per week, patients per day, research hours per week, and hours of teaching per week as predictor (independent) variables. Overall model fit was significant for overall score, evaluation, and induction. Age was found to be a significant predictor of lower CT scores for overall score and for each subscale, except for analysis. Research hours per week and clinical

days per week were also predictive of lower scores on evaluation and induction, respectively (Table 3).

Follow-up Analysis

As age was the most common and powerful predictor of CT skills, we conducted a follow-up analysis to better understand the relationship between age and CT. For comparability across subscales, all scores were scaled (z score, mean=0, SD=1). We used polynomial regression to test nonlinear relationships, quadratic, cubic, and quartic. Results indicated that the data were best fit by a quadratic model for all scores, except for analysis. Analysis displayed no significant linear or nonlinear relationships (Table 4). No other higher order polynomials were significant. The parameters of the quadratic regression were used to predict scores on the overall and subscale scores at 5-year age intervals (Figure 1). The follow-up analysis indicates a negative quadratic relationship between age and CT skills.

Discussion

We found that family medicine residents and practicing FPs differ in CT skills. Practicing physicians generally exhibited lower scores on the CCTST compared to residents, with age as the strongest negative predictor of CCTST scores. Group differences are likely due to the homogeneity of age for the resident sample compared with the heterogeneity of age amongst the practicing physicians. CT skills appear to decline at the greatest rate after 60 years of age. It appears the cognitive demands of the medical profession do not protect against normal age-related decline in the aspects of

Table 1: Demographic Characteristics of Practicing Family Physicians' and Residents' Age

	Residents Age (Years)	FP Age (Years)	Years in Practice	Patients per Day	Clinical Days per Week	Research Hours per Month	Hours Teaching per Week
Mean	27.9	47.7	17.9	20.9	3.97	3.50	12.2
SD	5.1	12.0	12.9	7.61	1.42	7.68	13.6
Range	-	27-73	2-43	0-40	0-7	0-50	0-80

Abbreviation: FP, family physician.

Table 2: Results of a One-Way MANOVA and *t* Tests^a

Score		Residents (n=59)	Practicing Physicians (n=62)	Effect Size η^2
Percentile rank*	Mean	69th	61st	0.41
	SD	18.1	21.2	
	Range	30-98	6-97	
Overall*	Mean	23.9	22.3	0.41
	SD	3.83	4.35	
	Range	16-32	11-31	
Analysis	Mean	4.97	5.06	0.08
	SD	1.25	1.20	
	Range	2-7	2-7	
Inference	Mean	11.9	11.4	0.2
	SD	2.03	2.58	
	Range	8-15	5-16	
Evaluation***	Mean	7.08	5.82	0.70
	SD	1.78	1.85	
	Range	3-11	1-9	
Induction**	Mean	12.8	11.7	0.55
	SD	1.87	2.14	
	Range	8-16	7-15	
Deduction	Mean	11.1	10.6	0.17
	SD	2.91	2.91	
	Range	6-17	4-17	

Abbreviation: MANOVA, multivariate analysis of variance.

* $P < .05$; ** $P < .01$; *** $P < .001$.

[†]Cohen's *d*: small = .2, medium = .5, large = .8⁴¹

^a Analysis was also conducted with the borderline resident outlier in the data set ($n=60$, $M=23.8$, $SD=4.09$). Results of the *t* test varied marginally for overall score, $t(120)=1.93$, $P=.056$, $d=.3$, and percentile rank $t(120)=1.89$, $P=.061$, $d=.34$. Results of the MANOVA were not affected.

cognition related to CT skills. The results of this present study converge with other findings on aging and cognition showing a normal pattern of senescence, especially when considering the quadratic relationship.³¹⁻³³ Deficits in the cognitive skills related to CT are most strongly related to declines in working memory (WM).³² Though the sample used was restricted to family physicians, it can be hypothesized that the results are generalizable to other specialties in medicine.

Practicing physicians performed similarly to residents on the analysis subscale. Analysis measures

the ability to identify assumptions, reasons, themes, and the evidence used in making arguments or offering explanations. These skills are concordant with the use of pattern recognition and illness scripts in the diagnostic process and are more reliant on knowledge.³⁴ As humans age and fluid intelligence declines, crystallized intelligence remains fairly stable until approximately the eighth decade.³³ The amount of knowledge that a person has, and can develop, does not decline at the same rate as WM and so the aspects of CT that do not rely on WM remain more stable.

Though individual variance is important when considering aging and CT skills, as shown in Figure 1, CT skills are not as strong after age 60. Previously, a small but significant negative relationship has been found between diagnostic accuracy and age.^{35,36} Lower diagnostic accuracy and consideration of the need for CT skills in CR,¹ and the need to engage System II processing for CT,³⁷ indicate that it is important to assess CT skills longitudinally; stability in pattern recognition may not be protective for optimal diagnostic accuracy. Declining CT skills could be a harbinger of reduced clinical ability.

Three unexpected findings were identified in the present study. The first two are the negative relationship between research hours and evaluation, and clinical days worked and induction. It could be expected the two variables would be positively related with CT skills as engaging in research and spending more time in the clinic may hone CT skills. However, patients seen per day was not correlated with clinical days worked, so spending more time in the clinic may not increase the opportunity to practice CT skills. It is possible more time spent in the clinic is related to less time engaged in reading or other activities that would support induction skills, and may be concurrently associated with greater fatigue, which can negatively impact cognitive abilities.³⁸⁻⁴⁰ The negative relationship between research hours and evaluation scores is also counterintuitive. More investigation is required to understand how conducting research in conjunction with practicing medicine influences CT skills. Finally, no relationship was found between population size served and patients seen per day and no differences were found for CT skills based on population size served; therefore, based on the current data, it does not appear these variables interact. Further investigation of the influence of clinical days worked, research hours, and populations served is necessary. A more granular approach along

Table 3: Results of Hierarchical Multiple Linear Regressions for the CCTST and Its Subscales*

Dependent Variable	Predictor Variables	b (95% CI)	SE	β (95% CI)	P
Overall	Age	-0.14 (-0.23, -0.005)	.05	-0.40 (-0.66, -0.14)	.03
		$F(5, 53)=2.34, P=.05, R^2=.18, R^2_{Adjusted}=.1$			
Inference	Age	-0.07 (-0.06, -0.01)	.03	-0.31 (-0.58, -0.05)	.023
		$F(5, 53)=1.43, P=.23, R^2=.12, R^2_{Adjusted}=.04$			
Evaluation	Age	-0.06 (-0.10, -0.02)	.02	-0.41 (-0.66, -0.15)	.002
	Research hours/week	-0.04 (-0.06, -0.003)	.02	-0.28 (-0.53, -0.03)	.03
	$F(5, 53)=2.95, P=.02, R^2=.22, R^2_{Adjusted}=.14$				
Induction	Age	-0.07 (-0.11, -0.03)	.02	-0.40 (-0.65, -0.15)	.002
	Clinical days/week	-0.42 (-0.80, -0.04)	.18	-0.28 (-0.54, -0.03)	.03
	$F(5, 53)=3.07, P=.02, R^2=.22, R^2_{Adjusted}=.15$				
Deduction	Age	-0.07 (-0.13, -0.007)	.03	-0.30 (-0.57, -0.03)	.03
		$F(5, 53)=1.30, P=.28, R^2=.11, R^2_{Adjusted}=.03$			

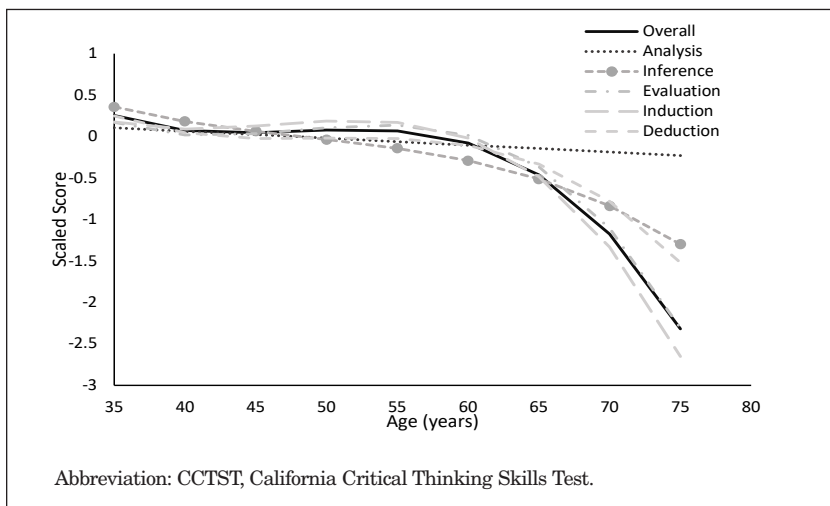
Abbreviation: CCTST, California Critical Thinking Skills Test.

*Only significant results are shown, for results of all variables see the supplemental material.

Table 4: Results of a Quadratic Regression

Score	Intercept	b	SE	P Value	R ²
Overall	7.4e-17	-.300	.092	.003	0.20
Analysis*	.4	-.008	.010	.440	0.01
Inference	1.8e-16	-.300	.094	.003	0.15
Evaluation	1.3e-16	-.230	.094	.020	0.16
Induction	1.1e-16	-.260	.091	.005	0.21
Deduction	6.6e-17	-.230	.096	.020	0.11

*Linear regression

Figure 1: Quadratic Regression for Prediction of Scaled CCTST Scores

with purposeful sampling is likely required to target these issues.

Future Directions

It is important to assess CT skills longitudinally across physician populations due to the importance of CT skills in the diagnostic process and the potential negative implications of declining CT skills for diagnostic accuracy. Large sample frames and longitudinal data collection could employ data mining techniques to better identify the contributing factors to CT and changes in CT skills. Consistent assessment of CT skills could be used to inform physicians of changes in their cognitive skill

set and identify physicians that may need continued professional development and support to remain in practice, or ultimately direction away from clinical practice. A measure of CT alone would not be appropriate for this but would function as one aspect of a broader evaluation of a physician's practice.

The present study sampled family physicians exclusively. It is reasonable to expect that since the pattern of decline is comparable to samples from the general population, similar results would be observed across medical specialties. Further research on a wider range of specialties is necessary to address this hypothesis. Finally, direct comparisons to the general population or among other professions should provide further insights for physician CT skills.

Limitations

This study has four main limitations. First, there was a relatively small sample size. Nevertheless, there were robust effect sizes for the differences between residents and practicing physicians indicating adequate statistical power for these analyses. Secondly, we assumed that the entire sample of family physicians was healthy with no neurodegenerative disorders. At present it is not possible to know if any of the participants had any age-related or other pathologies that may have influenced scores. Thirdly, the environmental testing conditions were not identical between residents and FPs. Residents completed the test during protected time while FPs completed the test at their own discretion. However, test forms, instructions, and allotted time to complete the test were identical. Finally, all residents were from a single institution. At present it is not possible to know if the CT skills of residents between institutions is uniform, limiting generalizability.

Conclusion

CT skills are an essential consideration in clinical competency. Changes in CT skills in practicing family physicians reflect normal human aging

indicating CT assessment could be a diagnostic method for identifying potential declines in physician performance. CT skills should not be the only consideration, or the ultimate determinant, as individuals will vary in clinical performance. It may also be possible to develop external supports and cues to help older doctors remain effective in practice when it is expected CT skills may be declining. Further research on CT skills employing other physician specialties, other professions, and the population in general may be highly instructive of life span changes in cognitive skills.

FUNDING SUPPORT: Funding was provided by the Northern Alberta Academic Family Medicine Fund and the University of Alberta Department of Family Medicine.

PREVIOUS PRESENTATIONS: Preliminary study results were presented at the North American Primary Care Research Group Conference, November 2019, Toronto, Canada.

CORRESPONDING AUTHOR: Address correspondence to Dr David Ross, 205 College Plaza, University of Alberta, Edmonton, Alberta, Canada T6G 2C8. 1-780-342-4038. Fax: 780-492-8191. djross@ualberta.ca.

References

- Royce CS, Hayes MM, Schwartzstein RM. Teaching critical thinking: a case for instruction in cognitive biases to reduce diagnostic errors and improve patient safety. *Acad Med.* 2019;94(2):187-194.
- Croskerry P, Cosby KS, Graber ML, Singh H. *Diagnosis: Interpreting the Shadows.* Boca Raton, FL: CRC Press, Taylor & Francis Group; 2017.
- Croskerry P. From mindless to mindful practice—cognitive bias and clinical decision making. *N Engl J Med.* 2013;368(26):2445-2448.
- Daly P. A concise guide to clinical reasoning. *J Eval Clin Pract.* 2018;24(5):966-972.
- Groves M, O'Rourke P, Alexander H. Clinical reasoning: the relative contribution of identification, interpretation and hypothesis errors to misdiagnosis. *Med Teach.* 2003;25(6):621-625.
- Groves M, Dick M-L, McColl G, Bilszta J. Analysing clinical reasoning characteristics using a combined methods approach. *BMC Med Educ.* 2013;13(1):144.
- Harasym PH, Tsai TC, Hemmati P. Current trends in developing medical students' critical thinking abilities. *Kaohsiung J Med Sci.* 2008;24(7):341-355.
- Zuriguél Pérez E, Lluich Canut MT, Falcó Peguerolles A, Puig Llobet M, Moreno Arroyo C, Roldán Merino J. Critical thinking in nursing: scoping review of the literature. *Int J Nurs Pract.* 2015;21(6):820-830.

- Dowd JE, Thompson RJ, Schiff LA, Reynolds JA. Understanding the complex relationship between critical thinking and science reasoning among undergraduate thesis writers. *CBE Life Sciences Education.* 2018;17(1).
- Holyoak KJ, Morrison RG. *The Cambridge Handbook of Thinking and Reasoning.* New York, NY: Cambridge University Press; 2005.
- Facione PA. *The California Critical Thinking Skills Test—College Level. Technical Report #4. Interpreting the CCTST, Group Norms, and Sub-Scores.* Millbrae, CA; California Academic Press; 1990. <http://files.eric.ed.gov/fulltext/ED327566.pdf>. Accessed July 13, 2020.
- Facione PA. *The California Critical Thinking Skills Test -- College Level Technical Report #1. Experimental Validation and Content Validity.* Millbrae, CA; California Academic Press 1990. <http://files.eric.ed.gov/fulltext/ED327549.pdf>. Accessed July 13, 2020.
- Allaire JL. Assessing critical thinking outcomes of dental hygiene students utilizing virtual patient simulation: a mixed methods study. *J Dent Educ.* 2015;79(9):1082-1092.
- Fesler-Birch DM. Critical thinking and patient outcomes: a review. *Nurs Outlook.* 2005;53(2):59-65.
- Ross D, Loeffler K, Schipper S, Vandermeer B, Allan GM. Do scores on three commonly used measures of critical thinking correlate with academic success of health professions trainees? A systematic review and meta-analysis. *Acad Med.* 2013;88(5):724-734.
- Lee J, Lee Y, Gong S, Bae J, Choi M. A meta-analysis of the effects of non-traditional teaching methods on the critical thinking abilities of nursing students. *BMC Med Educ.* 2016;16(1):240.
- Kostopoulou O, Sirota M, Round T, Samaranyaka S, Delaney BC. The role of physicians' first impressions in the diagnosis of possible cancers without alarm symptoms. *Med Decis Making.* 2017;37(1):9-16.
- Sturmberg JP. Systems and complexity thinking in general practice: part 1 - clinical application. *Aust Fam Physician.* 2007;36(3):170-173.
- Kelly M, Lecturer S, Bennett D, Kelly M, Micgp MA. General practice : the DREEM attachment ? Comparing the educational environment of hospital and general practice placements. *Education for Primary Care.* 2015;23:1, 34-40. 8
- Pereira Gray D. Role reversal between primary and secondary care. *Med Educ.* 2003;37(9):754-755.
- Donner-Banzhoff N, Seidel J, Sikeler AM, et al. The phenomenology of the diagnostic process: A primary care-based survey. *Med Decis Making.* 2017;37(1):27-34.
- Zayapragassarazan Z, Menon V, Kar SS, Batmanabane G. Understanding Critical Thinking to Create Better Doctors. *J Adv Med Educ Res.* 2016;1(3):9-13.
- Ross D, Schipper S, Westbury C, et al. Examining critical thinking skills in family medicine residents. *Fam Med.* 2016;48(2):121-126.

24. von Elm E, Altman D, Egger M, Pocock S, Gøtzsche PC, Vandenbroucke J. The strengthening of reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. 2007. Bull World Health Organ. 85(11): 867-872.
25. Reale MC, Riche DM, Witt BA, Baker WL, Peeters MJ. Development of critical thinking in health professions education: A meta-analysis of longitudinal studies. *Curr Pharm Teach Learn*. 2018;10(7):826-833.
26. Sharples JM, Oxman AD, Mahtani KR, et al. Critical thinking in healthcare and education. *BMJ*. 2017;357:j2234.
27. Insight Assessment. Assessment Report Package. <https://www.insightassessment.com/article/what-is-included-in-the-assessment-report-package>. Accessed July 13, 2020.
28. IBM SPSS Statistics for Windows. 2017. Armonk, NY: IBM Corporation.
29. Tabachnick BG, Fidell LS. *Using Multivariate Statistics*. Sixth Edition. Boston: Pearson; 2013.
30. Insight Assessment. What is a Percentile Score? <https://www.insightassessment.com/article/what-is-a-percentile-score>. Published 2020.
31. McArdle JJ, Ferrer-Caja E, Hamagami F, Woodcock RW. Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Dev Psychol*. 2002;38(1):115-142.
32. Mutter SA, Haggblom SJ, Plumlee LF, Schirmer AR. Aging, working memory, and discrimination learning. *Q J Exp Psychol (Hove)*. 2006;59(9):1556-1566.
33. Salthouse TA. Trajectories of normal cognitive aging. *Psychol Aging*. 2019;34(1):17-24.
34. Croskerry P. Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Adv Health Sci Educ Theory Pract*. 2009;14(S1)(suppl 1):27-35.
35. Norcini JJ, Lipner RS, Kimball HR. Certifying examination performance and patient outcomes following acute myocardial infarction. *Med Educ*. 2002;36(9):853-859.
36. St-Onge C, Landry M, Xhignesse M, et al. Age-related decline and diagnostic performance of more and less prevalent clinical cases. *Adv Health Sci Educ Theory Pract*. 2016;21(3):561-570.
37. Lilienfeld SO, Ammirati R, Landfield K. Giving Debiasing Away: Can Psychological Research on Correcting Cognitive Errors Promote Human Welfare? *Perspect Psychol Sci*. 2009;4(4):390-398.
38. Schnyer DM, Zeithamova D, Williams V. Decision-making under conditions of sleep deprivation: cognitive and neural consequences. *Mil Psychol*. 2009;21(sup1)(suppl 1):S36-S45.
39. Bérastégui P, Jaspar M, Ghuyssen A, Nyssen AS. Fatigue-related risk management in the emergency department: a focus-group study. *Intern Emerg Med*. 2018;13(8):1273-1281.
40. Olds DM, Clarke SP. The effect of work hours on adverse events and errors in health care. *J Safety Res*. 2010;41(2):153-162.
41. Cohen JD. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Erlbaum; 1988.