Methods to Improve Diagnostic Reasoning in Undergraduate Medical Education in the Clinical Setting: a Systematic Review

Hongyun Xu, MBBS (S'pore)[1,3] , *Benson W. G. Ang, MBBS (S'pore)[1,3],*
Jian Yi Soh, MBBS (S'pore), MMed (Paeds) (S'pore), MRCPCH (Paeds) (UK)[1,2], and
Gominda G. Ponnamperuma, MBBS (Colombo), Dip. Psychology (Colombo), MMEd (Dundee), PhD (Dundee)[3]

[1]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore; [2]Khoo Teck Puat-National University Children's Medical Institute, National University Health System, Singapore, Singapore; [3]Centre for Medical Education (CenMED), Yong Loo Lin School of Medicine, National University of Singapore, Block MD 11, #05-10, Clinical Research Centre 10 Medical Drive, Singapore, Singapore.

BACKGROUND: In recent years, there have been numerous studies exploring different teaching methods for improving diagnostic reasoning in undergraduate medical students. This systematic review examines and summarizes the evidence for the effectiveness of these teaching methods during clinical training.

METHODS: PubMed, Embase, Scopus, and ERIC were searched. The inclusion criteria for the review consist of articles describing (1) methods to enhance diagnostic reasoning, (2) in a clinical setting (3) on medical students. Articles describing original research using qualitative, quantitative, or mixed study designs and published within the last 10 years (1 April 2009–2019) were included. Results were screened and evaluated for eligibility. Relevant data were then extracted from the studies that met the inclusion criteria.

RESULTS: Sixty-seven full-text articles were first identified. Seventeen articles were included in this review. There were 13 randomized controlled studies and four quasi-experimental studies. Of the randomized controlled studies, six discussed structured reflection, four self-explanation, and three prompts for generating differential diagnoses. Of the remaining four studies, two employed the SNAPPS[1] technique for case presentation. Two other studies explored schema-based instruction and using illness scripts. Twelve out of 17 studies reported improvement in clinical reasoning after the intervention. All studies ranked level two on the New World Kirkpatrick model.

DISCUSSION: The authors posit a framework to teach diagnostic reasoning in the clinical setting. The framework targets specific deficiencies in the students' reasoning process. There remains a lack of studies comparing the effectiveness of different methods. More comparative studies with standardized assessment and evaluation of long-term effectiveness of these methods are recommended.

INTRODUCTION

Clinical reasoning (CR) can be defined as "a process that operates toward the purpose of arriving at a diagnosis, treatment, and/or management plan."[1] It is a complex process that involves a series of steps and cognitive functions.[2] Effective CR is an essential skill for clinicians to acquire[2] as it reduces cognitive errors,[3] which contribute to the bulk of mistakes made in medicine[4] and compromise patient safety.[5, 6]

This systematic review will focus on examining existing methods in the literature to teach diagnostic reasoning (i.e., the process of arriving at a diagnosis). A suggested framework for breaking down the process into its various components is shown:[7–9]

1. Data acquisition
2. Problem representation
3. Hypothesis generation
4. Searching for and selection of illness script (comparing and contrasting illness scripts)
5. Diagnosis

Without the right diagnosis, empiric therapy will likely be ineffective for patients and, therefore, greater emphasis should be placed on this process.[10]

CR is often assumed to be a process that can be learnt through direct observation of expert clinicians.[11] This leads to students receiving irregular feedback and varying quality of supervision,[12] causing poor knowledge structures and an inability to synthesize clinical features during patient encounters.[13] The lack of feedback also leads to poor insight.[14]

Timely and regular feedback based on observation of students' reasoning skill with an emphasis on correctable deficit is recommended.[15, 16] Therefore, this review will focus on teaching methods that can be used in clinical settings (wards, clinics) as opposed to classroom-based techniques (e.g., problem-based learning). It will also provide suggestions on how tutors can give personalized feedback to students to refine their clinical judgment by targeting specific components of CR.[17]

To evaluate the impact of these teaching methods, the authors will apply the New World Kirkpatrick Model (NWKPM) to categorize the studies. The NWKPM consists of four levels (Table 1); a higher level reflects stronger evidence for the effectiveness of an intervention and its impact in bringing about a demonstrable improvement in the learner's behavior.[18]

Hence, the two questions investigated by this systematic review are:

1. How can different components of diagnostic reasoning be taught to medical students in the clinical setting?
2. How impactful are the interventions based on the NWKPM?

METHODS
Protocol, Information, and Search Strategy

We performed a systematic review in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline.[19] This study was not registered under PROSPERO. Two reviewers (AB, XH) conducted electronic database searches (on 6 Mar 2019) on PubMed, Embase, Scopus, and ERIC, using keywords related to CR, teaching, learning, and medical students. A hand search was also performed by examining the references of the articles that were reviewed in full. The exact search terms are presented under Supplementary Information. Figure 1 shows the complete search and study selection strategy.

Study Selection, Eligibility Criteria

The inclusion criteria were articles describing (1) methods to enhance diagnostic reasoning, (2) in a clinical setting, (3) on medical students. These articles are (4) either original research using qualitative, quantitative, or mixed study designs and (5)

Table 1 The Four Levels of the New World Kirkpatrick Model

Level		Description
1	Reaction	Evaluates the impression of the program. Does not measure outcomes but gauges the motivation and interest of participants.
2	Learning	Measures what participants have learnt in terms of knowledge/skills by using assessments and role plays to demonstrate the skills.
3	Behavior	Involves assessment of newly learnt knowledge/skill in the workplace. Determines if participants apply the new skills in the working environment.
4	Results	Measures the overall impact of the training at the organizational level (e.g., misdiagnosis rates, patient adverse events, etc.)

published within the last 10 years (1 April 2009–2019). Only studies over the past decade were chosen because the authors are keen to focus on evaluating more recent developments in teaching CR.[20] Another article has reviewed studies further in the past.[21]

The exclusion criteria were (1) CR not in the field of medicine (e.g., nursing, physiotherapy); (2) papers that only addressed patient management; (3) not involving medical students; and (4) editorials/commentaries/letters.

For papers which discussed other areas of CR, the authors only focused on extracting information that is related to improving diagnostic reasoning.

AB and XH independently screened the title and abstract of each article for relevance. Any discrepancies were resolved through discussion. An extended discussion involving GP was employed if no agreement could be reached. Full-text articles were then reviewed and analyzed independently by AB and XH.

Data Collection Process

We piloted a data extraction sheet on five random papers included in this review. Piloting did not change any of the constituents of the sheet. AB and XH independently extracted data for the studies and compared the data, with any discrepancies resolved through discussion. From each study, the following information was extracted (if applicable): (1) study design; (2) participants; (3) teaching method(s); (4) assessment tools used to evaluate the method(s); (5) reported improvement in CR; and (6) NWKPM level based on study design and methodology. An assessment was also performed using the Cochrane Collaboration's Tool for Assessing Risk of Bias.[22]

RESULTS
Study Selection and Characteristics

Database searches yielded 1550 records. One article was identified by hand search. This was narrowed down to 1070 after removal of duplicates. Screening of abstracts identified 67 articles for full-text reading. Of these, 17 satisfied the review criteria. There were 13 randomized controlled studies (five discussed reflection, four self-explanation, one on generating differential diagnoses, one on schema-based instruction, one on illness scripts, one SNAPPS[1]). Of the remaining four, there were three quasi-experimental studies (two on generating differential diagnoses, one on SNAPPS,) and one within-subjects study (reflective reasoning). Twelve studies reported improvement in CR. Article demographics are summarized in Table 2. The teaching methods, their effectiveness, and NWKPM level are summarized in Table 3. The risk of bias for "incomplete outcome data" of the studies was generally

---

[1] SNAPPS stands for summarize history and findings; narrow the differential; analyze the differential; probe preceptor about uncertainties; plan management; select case-related issues for self-study.
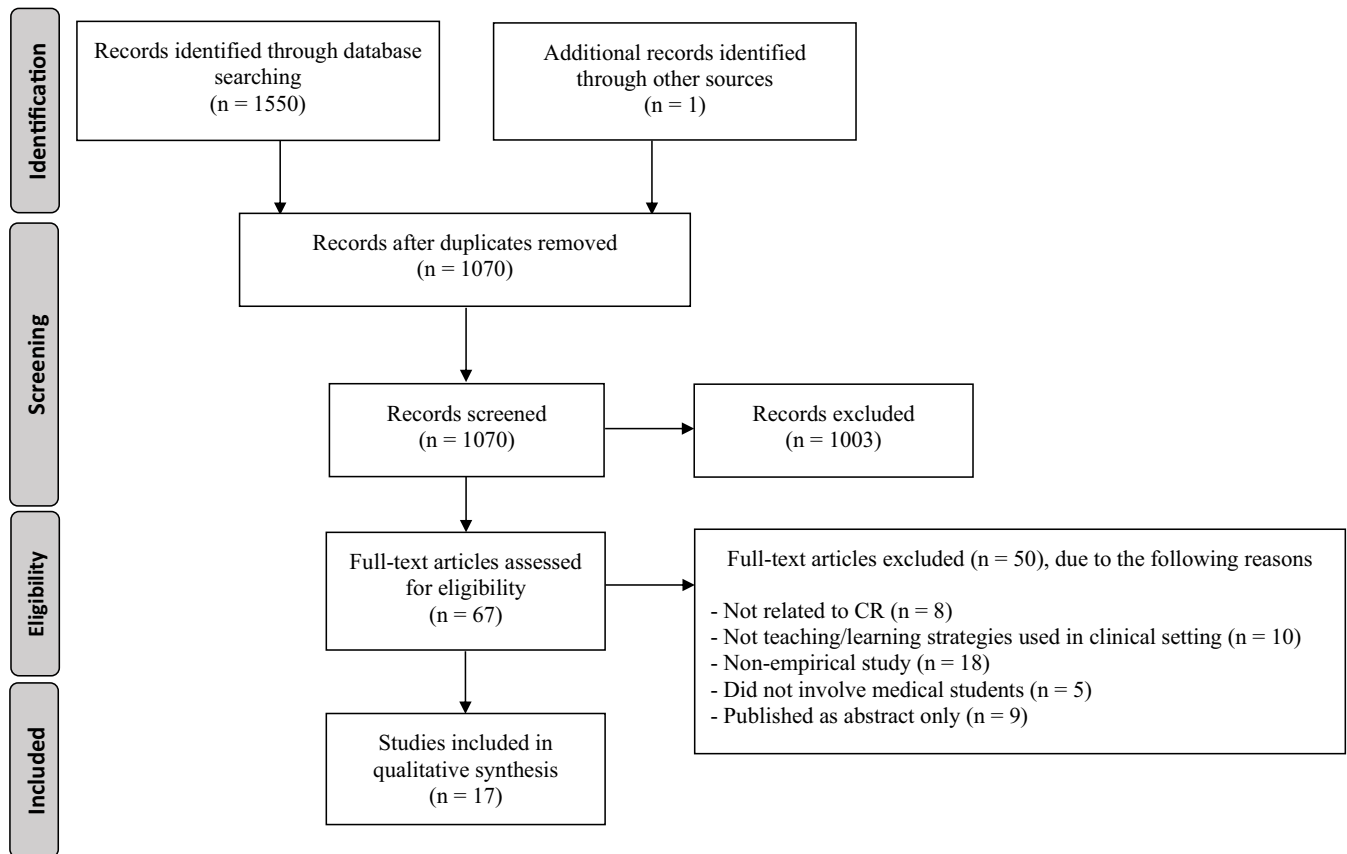
Figure 1 PRISMA workflow.

low, and there was generally scarce information or unclear risk of bias for the remaining domains.

Teaching Methods Useful in the Clinical Setting

A variety of teaching strategies for CR surfaced through this review. We now present these methods in the order of frequency that they are studied.

Structured Reflection. Structured reflection (SR) improves diagnostic accuracy in complex problems and counteracts cognitive errors among physicians.[23–25] In SR, students first generate their own list of differential diagnoses without prompts. They then analyze the supporting and rejecting features, expected but absent findings, and alternative differential diagnoses, and prioritize the differential diagnoses.[26] When students use SR for developing CR, they engage in elaboration and refinement of their own repertoire of illness scripts. Illness scripts are "mental cue cards" consisting of key features of individual diseases.[27] These mental representations are crucial for future problem-solving. Structured reflection has the potential to improve students' diagnostic competency by encouraging students to compare different illness scripts in a clinical encounter.[28]

Mamede et al.[26] studied this technique in 2012. The SR group outperformed other groups ($p < 0.01$, $p = 0.01$) in the delayed test phase despite performing worse in the immediate test phase ($p = 0.012$ and 0.04). Subsequently, Myung et al.[29] studied a similar technique and demonstrated a significant improvement in diagnostic accuracy scores ($p = 0.016$) in the group who utilized the technique. In another 2014 study by Mamede et al.,[30] diagnostic performance in the SR group improved significantly ($p < 0.01$) compared to others when approaching new exemplars of criterion diseases and cases of new diseases. In 2019, they extended the technique by comparing between free, cued, and modeled reflection.[31] The cued reflection group performed significantly better than others when diagnosing criterion cases ($p < 0.02$), but this improvement was not extended to adjacent diseases.

Two studies utilizing SR did not demonstrate improvement in diagnostic accuracy. Lambe et al.[32] did not find any significant differences ($p > 0.05$) between the control and interventional groups. They proposed that the cases used in the study were not structurally complex, while SR has been shown to be effective in more complex cases.[25] Mamede et al.[33] also found that reflective reasoning (compared to non-analytical reasoning and "deliberation without attention") did not increase diagnostic accuracy of students in the face of salient distracting clinical features. However, it was useful among residents ($p = 0.03$). They suggested that students are novices who do not yet possess an adequate knowledge base about supporting and rejecting features of diseases, making comparisons difficult.[33]

Table 2 Article Demographics of the 17 Included Empirical Studies

No.	Author(s) and year	Journal	Study participants	Study design
1.	Mamede et al., 2012	*BMJ Quality and Safety*	4th year medical students (n = 50) and IM residents (n = 34)	Within-subjects, post-test only
2.	Mamede et al., 2012	*Medical Education*	4th year medical student (n = 46)	Randomized, three-group, pre-post comparison
3.	Mamede et al., 2014	*Academic Medicine*	4th year medical students (n = 110)	Randomized, three-group, pre-post comparison
4.	Mamede et al., 2019	*Medical Education*	3rd year medical students (n = 80)	Randomized, three-group, pre-post comparison
5.	Myung et al., 2013	*Medical Teacher*	4th year medical students (n = 145)	Randomized, two-group, post-test only
6.	Chamberland et al., 2015	*Medical Education*	3rd year medical students (n = 54)	Randomized, three-group, pre-post comparison
7.	Chamberland et al., 2015	*Advances in Health Science Education*	3rd year medical students (n = 53)	Randomized, three-group, pre-post comparison
8.	Chamberland et al., 2011	*Medical Education*	3rd year medical students (n = 36)	Randomized, two-group, pre-post comparison
9.	Peixoto et al., 2017	*Advances in Health Sciences education: Theory and Practice*	4th year medical students (n = 39)	Randomized, two-group, post-test only
10.	Chai et al., 2017	*The Clinical Teacher*	3rd year medical students (n = 48)	Randomized, two-group, pre-post comparison
11.	Chew et al., 2017	*BMC Medical Education*	5th year medical students (n = 88)	Quasi-experimental (non-equivalent two-group, post-test only)
12.	Lambe et al., 2018	*Frontiers in Psychology*	4th and 5th year medical students (n = 184)	Randomized, three-group, post-test only
13.	Shimizu et al., 2013	*Medical Teacher*	4th to 6th year medical students (n = 188)	Quasi-experimental (two-group, post-test only)
14.	Sawanyawisuth et al., 2015	*Medical Teacher*	5th and 6th year medical students (n = 303)	Quasi-experimental (four-group, post-test only)
15.	Wolpaw et al., 2009	*Academic Medicine: Journal of the Association of American Medical Colleges*	3rd year medical students (n = 64)	Randomized, three-group post-test only
16.	Lee et al., 2010	*Family Medicine*	4th year medical students (n = 53)	Randomized, two-group, pre-post comparison
17.	Blisset et al., 2012	*Medical Education*	2nd year medical students (n = 53)	Randomized, two-group, post-test

Self-Explanation. Self-explanation (SE) is an active process in which the learner uses "remarks…to oneself after reading… statements in a text…or a problem…." SE allows the student to revise his/her biomedical knowledge and build a more coherent representation of the disease. In SE, learners actively link clinical features of a disease to their prior knowledge. This facilitates knowledge assimilation, elaboration, and restructuring, leading to a better development of illness scripts.[34] This method is more useful in the early stages of clinical training where the learner does not have a robust set of illness scripts yet.[12]

Chamberland et al. conducted three studies on SE.[34–36] The 2011 study[35] showed that SE resulted in better diagnostic performance than the control group after one week (p < 0.05) and was especially useful for less familiar cases. A follow-up 2015 study[34] compared between SE alone and listening to SE models (example cases using SE, performed by peers and experts) in addition to students' own SE. All groups showed significant improvement in diagnostic performance. The addition of other SE models did not produce significantly more improvement as hypothesized. Another 2015 study[36] compared students' SE alone and listening to residents' SE models, with/without addition of prompts (justification prompts and mental model revision prompts). All three groups showed significant improvement. The group that benefitted from other residents' SE along with the prompts

produced by themselves showed significantly more improvement than the other two groups. Performance in transfer cases also improved significantly, reflecting deep and meaningful learning. This showed that choosing the right level of expertise in providing SE models is important.

Peixoto et al.[37] conducted another study with a similar design but focusing on SE of pathophysiological mechanisms of diseases. They did not find any significant improvement in diagnostic performance between both groups. They postulated that the positive effect depends on the studied diseases sharing similar pathophysiological mechanisms.

Prompts for Differential Diagnoses. A core aspect of clinical reasoning is to generate a broad range of initial differential diagnoses, then narrowing it down as more clinical information is obtained. Many students face difficulties in the initial generation of differential diagnoses. This is not necessarily due to insufficient content knowledge, but rather a deficiency in metacognitive skills.[38]

Chai et al.[39] assessed and compared the utility of mnemonics (surgical sieve: VITAMIN CD^2) and Compass Medicine to increase the number of differential diagnoses generated. Compass Medicine is a handheld wheel which consisted of

---

[2] VITAMIN CD stands for Vascular, Infectious, Traumatic, Autoimmune, Metabolic, Iatrogenic, Neoplastic, Congenital, Degenerative.

Table 3 Results of the 17 Included Empirical Studies

No.	Source	Teaching/learning method	Assessment/measurement tools	Improvement in CR?	KPM level
1.	Mamede et al., 2012	- Reflective reasoning	- Participants solved 12 clinical cases using either: (1) Non-analytical reasoning (2) Reflective reasoning (3) Deliberation without attention - Their results for non-analytical reasoning and reflective reasoning were evaluated by 5 board-certified internists	No (students did not demonstrate an increased number of correct diagnoses)	2
2.	Mamede et al., 2012	- Structured reflection	- Medical students diagnosed six clinical cases using either: (1) Structured reflection (2) Immediate diagnosis (3) Differential diagnosis generation - Students' diagnostic performance during their learning phase, immediate post-test phase, and delayed (1 week) post-test phase were compared	Yes (improvement in performances for delayed test phase was observed)	2
3.	Mamede et al., 2014	- Structured reflection	- Medical students diagnosed four clinical of two criterion diseases using either: (1) Structured reflection (2) Single diagnosis (3) Differential diagnosis generation - One week later, students were requested to diagnose: (a) Two novel exemplars of each criterion disease (b) Four cases of new diseases that were plausible alternative diagnosis to the criterion diseases in the learning phase	Yes (students obtained higher mean diagnostic accuracy score when diagnosing new exemplars of criterion diseases)	2
4.	Mamede et al., 2019	- Structured reflection	- Medical students diagnosed nine clinical cases using either: (1) Free reflection (2) Cued reflection (3) Modeled reflection - Two weeks later, students were requested to diagnose: (a) Four novel exemplars of diseases studied in first phase (b) Four cases of adjacent disease (c) Two fillers	Yes (students performed better in terms of diagnostic accuracy measured by the number of correctly diagnosed cases)	2
5.	Myung et al., 2013	- Structured reflection	- Mean diagnostic scores from four objective structured clinical examinations (OSCE) were compared between the control and intervention group	Yes (students have a significantly higher mean diagnostic accuracy score)	2
6.	Chamberland et al., 2015	SE - Residents' SEs with prompts - Residents' SEs without prompts	- Students' diagnostic performances were compared pre- and post-intervention, both immediately and after 1 week	Yes (students obtained a higher diagnostic accuracy and performance score)	2
7.	Chamberland et al., 2015	SE - Provided by peers - Provided by experts	- Students' diagnostic performances were compared pre- and post-intervention, based on 4 clinical cases, both immediately and after 1 week	No (diagnostic performance did not show any difference on transfer cases)	2
8.	Chamberland et al., 2011	- SE	- Students' diagnostic performances were compared pre- and post-intervention, both immediately and after 1 week, using a scoring grid developed to mark the questions	Yes (intervention group demonstrated better diagnostic performances for less familiar cases)	2
9.	Peixoto et al., 2017	SE - on pathophysiological mechanisms of disease	- Medical students solved 6 criterion cases with or without SE - One week later, the students solved 8 new cases of the same syndrome - Students' performances were compared between both groups for: (1) Accuracy of initial diagnosis for the cases in the training phase (2) Accuracy of final diagnosis after SE have taken place (3) Accuracy of the initial diagnosis provided for the cases in the initial assessment	No (no improvement in diagnostic performances for all diseases)	2
10.	Chai et al., 2017	Generating DDx using:			2

Table 3. (continued)

No.	Source	Teaching/learning method	Assessment/measurement tools	Improvement in CR?	KPM level
		- Surgical sieve - Compass medicine (handheld tool)	- 30-min written test to generate possible DDx - The numbers of DDx generated before and after the teaching were compared	Yes (significantly greater number of differential diagnosis was generated)	
11.	Chew et al., 2017	Generating DDx using: - TWED checklist	- Script concordance tests	No (intervention group did not score significantly greater)	2
12.	Lambe et al., 2018	Generating DDx using: - Short guided reflection process - Long guided reflection process	- A series of clinical cases were diagnosed by using first impressions, or by using a short or long guided reflection process - Participants were asked to rate their confidence at intervals	No (did not elicit more accurate final diagnoses than diagnosis based on 1st impression)	2
13.	Shimizu et al., 2013	Generating DDx using: (1) DDXC Structured framework using: (1) GDBC	- Medical students were tasked to complete 5 diagnostic cases (2 difficult, 3 easy) using either DDXC or GDBC - The diagnostic mean score was compared between both groups	Yes (increased in proportion of correct diagnoses using DDXC)	2
14.	Sawanyawisuth et al., 2015	SNAPPS - In ambulatory clinic	- 12 outcome measures were used to assess an audio-recorded case presentation of medical students - Relevant measures were: (1) Number of basic clinical attributes of the chief complaint and history of present illness (2) Number of differential diagnoses considered (3) Number of justified diagnosis in the differential diagnosis (4) Number of basic attributes used to support the diagnosis (5) Number of presentations containing both history of present illness and physical exam findings	Yes (students had a greater number of differential diagnosis and more features to support the differential diagnosis)	2
15.	Wolpaw et al., 2009	- SNAPPS	- The content of each audio recording was coded for 10 presentation elements - Relevant measures are: (1) Basic attributes of chief content and history of present illness (2) Inclusion of both history and examination findings (3) Formulation of a differential diagnosis (4) Justification of the hypothesis in the differential (5) Comparing and contrasting hypotheses	Yes (students had a greater number of differential diagnosis and more features to support the differential diagnosis)	2
16.	Lee et al., 2010	3-h workshop on CR that used illness scripts, conducted with small-group teaching	- Using DTI to assess students' CR style and attitude pre- and post-intervention - Students' diagnostic performances in solving CR problems were compared pre- and post-intervention	Yes (improvement in diagnostic performances for CRP score)	2
17.	Blisset et al., 2012	- Use of schemas	- Written and practical test to assess diagnostic accuracy measured by percentage of questions answered correctly	Yes (students performed better on structured knowledge questions and had higher diagnostic success)	2

SE self-explanation, DDx differential diagnoses, DDXC differential diagnoses checklist, GDBC general debiasing checklist, DTI diagnostic thinking inventory

three concentric moving disks of different sizes, one each for anatomy, pathology, and etiology. It is a modification of the more commonly used surgical sieve methods, and it is based on the visual, auditory, reading and writing, kinesthetic (VARK) concept of learning.[39] A significant increase in the number of differential diagnoses was noted post-intervention for both groups compared to pre-intervention (p = 0.001), with no difference between the methods.

Shimizu et al.[40] studied the effects of differential diagnoses checklists (DDXCs) and general debiasing checklists (GDBC)

on diagnostic accuracy, and found that both can reduce diagnostic errors.[41] DDXC users were guided to consider alternative diagnoses that should not be/are commonly missed. Medical students using GDBC were guided to optimize their cognitive load by providing a reproducible approach to diagnosis. The effect of both interventions was measured by the number of correct diagnoses for five different cases. DDXC led to more correct diagnoses in complex cases, while using GDBC did not result in any significant improvement in diagnostic performance.[40]

Chew et al. proposed the TWED[3] checklist[42] to highlight limb/life-threatening conditions, evidence to make a clinical diagnosis, alternative diagnoses and the consequences of being incorrect. Other dispositional factors that could potentially influence decision-making were included, forcing users to consider additional patient information. Students that used the TWED checklist performed better (p = 0.008) compared to those who did not. However, this effect was only observed in the first half of the trial and the authors postulated that the TWED checklist could not be properly utilized due to time constraints. Overall, no significant difference in diagnostic performance between both groups was detected (p = 0.159).[42]

Case Presentation Techniques (SNAPPS). Case presentation techniques provide students with a structured way to express their clinical reasoning. This allows tutors to assess the learner's CR process and identify specific weaknesses.[43] Such techniques also encourage students to focus less on factual information and more on the CR process as well as their uncertainties behind the case.

One example is SNAPPS, which is a "learner-centered approach" to case presentation that encourages concise reporting of facts and expression of CR.[43–45] It is a six-step process, one for summarizing the case and the remainder to facilitate expression of CR. A 2009 study[43] evaluated different presentation elements (e.g., number of differential diagnoses, comparing and contrasting hypotheses) related to CR using SNAPPS. Students using SNAPPS gave twice as many differential diagnoses (p < 0.000), justified their differential diagnoses five times more frequently (p < 0.000), and compared and contrasted hypotheses more often (p < 0.000) when compared with a comparison group (focused on techniques for obtaining feedback from preceptors) or a "usual and customary group" (any presentation format the student preferred with no specific training). Another 2015 study[46] evaluated the generalizability of SNAPPS among Asian medical students. Echoing the findings of the 2009 study,[43] SNAPPS users generated more differential diagnoses (p = 0.016) and provided more evidence to support their differentials (p = 0.001), in comparison to the control group.

Illness Scripts and Knowledge Encapsulation/Structure. As mentioned above, illness scripts are "mental cue cards" representing individual diseases. They provide a structured template for learners to organize their knowledge, in relation to an illness, facilitating both problem representation and comparing and contrasting of key clinical features in making a diagnosis.[8, 47] Key features of an illness script include epidemiological factors, signs and symptoms of a disease, as well as the pathophysiology behind it.[27]

Lee et al.[13] conducted a workshop teaching students the important elements of CR, problem representation, and how to develop and select an appropriate illness script. Diagnostic

performance was assessed by objectively using 10 CR problems (CRP) and subjectively using the diagnostic thinking inventory (DTI). The study found no significant difference in the overall DTI score, and the post-flexibility and post-structure subscales pre- and post-intervention between the intervention and control groups. The intervention group fared significantly better in CRP scores. In the intervention group, 43% scored the maximum possible CRP score, compared to 29% of the control group.[13]

Teaching and Learning Techniques Using Schemas. A "schema" is a mental framework used to capture the gist of real-world information and store them as more coherent forms for use.[47–49] It helps to identify key features that can be used to include or exclude a set of diagnoses. Schemas differ from illness scripts as illness scripts are formulated around an individual disease. When encountering diagnostic challenges, multiple illness scripts are recalled, and then compared. On the other hand, schema is a mental framework that groups multiple diseases together, along with their distinguishing features, hence helping to reduce the cognitive load.[48] Blissett et al.[48] performed a trial using a schema-based instructional approach for cardiac valvular lesions and assessed students using both written and practical tests. For the written examination, the schema-based instruction group had significantly higher scores on structured knowledge questions (p < 0.0001) but not on factual knowledge questions (p = 0.93), when compared to the group that received traditional instructions. This result was also seen in a follow-up test 2 to 4 weeks later. For the practical test, the intervention group had greater diagnostic accuracy (p < 0.001) for both taught and untaught conditions.

DISCUSSION

Heterogeneity of Interventions and Implications for Undergraduate Medical Education

There is vast heterogeneity in the methods used to teach CR to medical students. This leads to a lack in standardization for teaching CR, as well as differing methods of assessing CR. This is expected as CR is a complicated process with no recognized gold standard.[8, 50, 51]

Lambe et al.[21] conducted a review examining studies from 1994 to 2014, and found several methods that could be useful to improve CR. In our review, we examined more recent studies that include a wider range of techniques (e.g., diagnostic tools and case presentation techniques) that target different components of CR. Furthermore, the prior review did not explore how these techniques could be used to teach medical students in a clinical setting. To build on this, we propose an exemplar framework to guide tutors on how this could be done. The framework proposed anchors on the principle of deliberate practice.[17, 52] While there have been multiple methods utilized to teach CR, these methods only target specific component(s) of

---

[3] TWED stands for Threat, What else, Evidences, Dispositional influence.

Table 4 Teaching Methods to Enhance Diagnostic Reasoning

Component	Teaching method	How it helps
Data acquisition	SNAPPS	A technique that forces students to thoroughly think through all aspects of a case
	Illness scripts	Good organization and structure of knowledge allow student to be aware of what are the pertinent features of history/examination to obtain and pay attention to
	Schemas	Highlights key features that differentiate between different conditions, allow a student to be aware of what are the pertinent features of history/examination to obtain and pay attention to
Problem representation	SNAPPS	As above
	Self-explanation	Facilitates knowledge assimilation and organization
		Aids development of illness scripts
	Illness scripts	Good organization and structure of knowledge allow student to be aware of the pertinent features to pay attention to
	Schemas	Highlights key features that differentiate between different conditions, allow a student to be aware of the pertinent features to pay attention to
Hypothesis generation (generation of DDx)	SNAPPS	As above
	DDx checklists Mnemonics (Surgical Sieve) Handheld tools (Compass Medicine)	Aids students in generating a broad base of differentials using a structured framework
Hypothesis testing (comparing and contrasting illness scripts)	SNAPPS	As above
	Self-explanation	As above
	Illness scripts	Good organization and structure of knowledge will facilitate comparison of different illnesses
	Schemas	Highlights key features that differentiate between different conditions
	Structured reflection	Forces students to analyze supporting and rejecting features of conditions and prioritize differentials

SNAPPS summarize history and findings; narrow the differential; analyze the differential; probe preceptor about uncertainties; plan management; select case-related issues for self-study; DDx differential diagnoses

CR (Table 4).[53–55] This may not be ideal as students have varying strengths and weaknesses. Educators should utilize different methods to target specific deficiencies. This allows personalized feedback to be provided promptly for recalibration in the student's reasoning process to occur.[17]

This exemplar framework involves using SNAPPS to aid students in expressing their CR process. By listening to the presentation, tutors can identify areas in which the student is weaker in, and subsequently employ the suggested methods to improve the student's deficit:

(1) An inability to summarize a patient's key history and physical examination findings in a succinct manner could suggest a deficit in either data acquisition or problem representation. To improve on this, tutors can strengthen students' illness scripts by highlighting key clinical and epidemiological features that should be present in the summary statement. The students' learning can be reinforced by encouraging them to practice self-explanation thereafter on how the disease's pathophysiology may lead to the clinical presentation. Upon receiving this feedback, students can then clerk another patient with a similar complaint. This can be repeated until the tutor is satisfied with the summary statement.

(2) Difficulties in providing a list of differential diagnosis points to a deficiency in hypothesis generation. To overcome this, prompts for differential diagnoses could be taught, This is paramount for preventing anchoring bias and premature closure, especially for students early in their training.[56]

(3) Being unable to analyze the differentials or provide the correct diagnosis suggests a weakness in searching for and selecting suitable illness scripts in their mental repository. Students can use structured reflection if they are deficient in this area. Based on the differential diagnosis that has been generated, students can be asked to verbalize key features that are present/absent in the patient's presentation. With this, tutors can guide students to "weigh" diagnostic probabilities and hence rank the differential diagnoses in order of likelihood.[26,30] This improves the student's clinical insight.[10]

While the authors did propose specific methods to improve each component of CR, a limitation of this framework is the lack of synthesized evidence comparing the effectiveness of different methods. This makes it difficult for tutors to choose the best strategies to teach students despite being able to identify their weaker areas. Hence, future studies should be performed to focus on comparing the effectiveness of different strategies in improving a specific component of CR.

Furthermore, the real-world impact of these interventions on patient care remains unclear as all studies are ranked KPM level 2. This echoes the findings of Lambe et al.[21] Further studies designed to target KPM level 3/4 are suggested. Such studies will require more resources and planning, and it is questionable whether these findings can be scaled up to practical teaching in the entire medical school. There may also be

potential ethical and practical difficulties—it is neither fair nor practical to limit a student to only a particular style of learning for a prolonged period, and therefore compromising their training as a clinician. Given the importance of CR to optimal patient care, such an effort, with suitable methodological alternations to overcome the said practical and ethical issues, is justifiable and overdue.

Lastly, the effectiveness of these techniques in atypical cases or presentations is also unclear. In the clinical environment, doctors encounter various problems and presentations. Any tool utilized should cover these variations as much as possible.[31, 48, 57–63]. Test cases in many studies are curated to classical presentations, but patients may not present with such discrete features.

Limitations of Review

CR can be taught in a large variety of ways worldwide, but these might not have been studied or published. Gray literature was also not included in our search, and we would have also missed those not published in formal academic channels. We also chose not to perform a metanalysis due to the significant heterogeneity of the studies.

Conclusion

This review provides synthesized evidence for teaching CR to medical students. The example framework provides tutors with tools to improve students' deficits in a targeted manner. It considers students who might be in different phases of their training and hence their skills in CR. All studies have been ranked at NWKPM level two. We recommend more carefully planned study designs and evaluation of long-term effectiveness of these teaching methods using standardized assessment methods.

Corresponding Author: *Hongyun Xu, MBBS (S'pore); Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore (e-mail: hyx2196@gmail.com).*

Declarations:

Ethics Approval: *Not applicable.*

Disclaimer: *None.*

Conflict of Interest: *None.*

REFERENCES

1. Young M, Thomas A, Lubarsky S, et al. Drawing Boundaries: The Difficulty in Defining Clinical Reasoning. *Academic Medicine.* 2018;93(7):990-995.
2. Lateef F. Clinical Reasoning: The Core of Medical Education and Practice. *International Journal of Internal and Emergency Medicine.* 2018;1(2).
3. Scott IA. Errors in clinical reasoning: causes and remedial strategies. *BMJ.* 2009;338:b1860.
4. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med.* 2005;165(13):1493-1499.
5. Singh H, Schiff GD, Graber ML, Onakpoya I, Thompson MJ. The global burden of diagnostic errors in primary care. *BMJ quality & safety.* 2017;26(6):484-494.
6. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med.* 1991;324(6):370-376.
7. Yazdani S, Hosseinzadeh M, Hosseini F. Models of clinical reasoning with a focus on general practice: A critical review. *Journal of advances in medical education & professionalism.* 2017;5(4):177-184.
8. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med.* 2006;355(21):2217-2225.
9. Foshay W. *Medical Problem Solving, an Analysis of Clinical Reasoning by Arthur S. Elstein; Lee S. Shulman; Sarah A. Sprafka.* Vol 281980.
10. Cunha BA. The Master Clinician's Approach to Diagnostic Reasoning. *The American Journal of Medicine.* 2017;130(1):5-7.
11. Gay S, Bartlett M, McKinley R. Teaching clinical reasoning to medical students. *Clinical Teacher.* 2013;10(5):308-312.
12. Schmidt HG, Mamede S. How to improve the teaching of clinical reasoning: a narrative review and a proposal. *Medical education.* 2015;49(10):961-973.
13. Lee A, Joynt GM, Lee AKT, et al. Using illness scripts to teach clinical reasoning skills to medical students. *Family medicine.* 2010;42(4):255-261.
14. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med.* 2008;121(5 Suppl):S2-23.
15. Anderson PAM. Giving feedback on clinical skills: are we starving our young? *Journal of graduate medical education.* 2012;4(2):154-158.
16. Ramani S, Krackov SK. Twelve tips for giving feedback effectively in the clinical environment. *Medical teacher.* 2012;34(10):787-791.
17. Rajkomar A, Dhaliwal G. Improving diagnostic reasoning to improve patient safety. *Perm J.* 2011;15(3):68-73.
18. Kirkpatrick JD, Kirkpatrick WK. Kirkpatrick's four levels of training evaluation. 2016.
19. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ.* 2009;339:b2700.
20. Lum LHW, Poh K-K, Tambyah PA. Winds of change in medical education in Singapore: what does the future hold? *Singapore Med J.* 2018;59(12):614-615.
21. Lambe KA, O'Reilly G, Kelly BD, Curristan S. Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review. *BMJ quality & safety.* 2016;25(10):808-820.
22. Higgins JPT TJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). . Cochrane Handbook for Systematic Reviews of Interventions version 6.1 (updated September 2020). Cochrane, 2020. Available from www.training.cochrane.org/handbook. In.
23. Mamede S, Schmidt HG, Rikers RM, Custers EJ, Splinter TA, van Saase JL. Conscious thought beats deliberation without attention in diagnostic decision-making: at least when you are an expert. *Psychol Res.* 2010;74(6):586-592.
24. Mamede S, Van Gog T, Van Den Berge K, et al. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA - Journal of the American Medical Association.* 2010;304(11):1198-1203.
25. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Medical education.* 2008;42(5):468-475.

26. Mamede S, van Gog T, Moura AS, et al. Reflection as a strategy to foster medical students' acquisition of diagnostic competence. *Medical education.* 2012;46(5):464-472.

27. Gavinski K, Covin YN, Longo PJ. Learning How to Build Illness Scripts. *Academic Medicine.* 2019;94(2):293.

28. Chamberland M, Mamede S, Bergeron L, Varpio L. A layered analysis of self-explanation and structured reflection to support clinical reasoning in medical students. *Perspectives on medical education.* 2020.

29. Myung SJ, Kang SH, Phyo SR, Shin JS, Park WB. Effect of enhanced analytic reasoning on diagnostic accuracy: a randomized controlled study. *Medical teacher.* 2013;35(3):248-250.

30. Mamede S, Van Gog T, Sampaio AM, De Faria RMD, Maria JP, Schmidt HG. How can students' diagnostic competence benefit most from practice with clinical cases? the effects of structured reflection on future diagnosis of the same and novel diseases. *Academic Medicine.* 2014;89(1):121-127.

31. Mamede S, Figueiredo-Soares T, Santos S, Faria R, Schmidt H, Gog T. Fostering novice students' diagnostic ability: the value of guiding deliberate reflection. *Medical education.* 2019;53.

32. Lambe KA, Hevey D, Kelly BD. Guided reflection interventions show no effect on diagnostic accuracy in medical students. *Frontiers in psychology.* 2018;9(NOV).

33. Mamede S, Splinter TAW, Van Gog T, Rikers RMJP, Schmidt HG. Exploring the role of salient distracting clinical features in the emergence of diagnostic errors and the mechanisms through which reflection counteracts mistakes. *BMJ Quality and Safety.* 2012;21(4):295-300.

34. Chamberland M, Mamede S, St-Onge C, Setrakian J, Bergeron L, Schmidt H. Self-explanation in learning clinical reasoning: the added value of examples and prompts. *Medical education.* 2015;49(2):193-202.

35. Chamberland M, St-Onge C, Setrakian J, et al. The influence of medical students' self-explanations on diagnostic performance. *Medical education.* 2011;45(7):688-695.

36. Chamberland M, Mamede S, St-Onge C, Setrakian J, Schmidt HG. Does medical students' diagnostic performance improve by observing examples of self-explanation provided by peers or experts? *Advances in health sciences education : theory and practice.* 2015;20(4):981-993.

37. Peixoto JM, Mamede S, de Faria RMD, Moura AS, Santos SME, Schmidt HG. The effect of self-explanation of pathophysiological mechanisms of diseases on medical students' diagnostic performance. *Advances in health sciences education : theory and practice.* 2017;22(5):1183-1197.

38. Leeds FS, Atwa KM, Cook AM, Conway KA, Crawford TN. Teaching heuristics and mnemonics to improve generation of differential diagnoses. *Medical education online.* 2020;25(1):1742967-1742967.

39. Chai J, Evans L, Hughes T. Diagnostic aids: the Surgical Sieve revisited. *The clinical teacher.* 2017;14(4):263-267.

40. Shimizu T, Matsumoto K, Tokuda Y. Effects of the use of differential diagnosis checklist and general de-biasing checklist on diagnostic performance in comparison to intuitive diagnosis. *Medical teacher.* 2013;35(6):e1218-e1229.

41. Ely JW, Graber ML, Croskerry P. Checklists to reduce diagnostic errors. *Academic medicine : journal of the Association of American Medical Colleges.* 2011;86(3):307-313.

42. Chew KS, van Merrienboer JJG, Durning SJ. Investing in the use of a checklist during differential diagnoses consideration: what's the trade-off? *BMC medical education.* 2017;17(1):234.

43. Wolpaw T, Papp KK, Bordage G. Using SNAPPS to facilitate the expression of clinical reasoning and uncertainties: a randomized comparison group trial. *Academic medicine : journal of the Association of American Medical Colleges.* 2009;84(4):517-524.

44. Wolpaw TM, Wolpaw DR, Papp KK. SNAPPS: a learner-centered model for outpatient education. *Academic medicine : journal of the Association of American Medical Colleges.* 2003;78(9):893-898.

45. Wolpaw T, Côté L, Papp KK, Bordage G. Student uncertainties drive teaching during case presentations: More so with SNAPPS. *Academic Medicine.* 2012;87(9):1210-1217.

46. Sawanyawisuth K, Schwartz A, Wolpaw T, Bordage G. Expressing clinical reasoning and uncertainties during a Thai internal medicine ambulatory care rotation: does the SNAPPS technique generalize? *Medical teacher.* 2015;37(4):379-384.

47. Custers EJ. Thirty years of illness scripts: Theoretical origins and practical applications. *Medical teacher.* 2015;37(5):457-462.

48. Blissett S, Cavalcanti RB, Sibbald M. Should we teach using schemas? Evidence from a randomised trial. *Medical education.* 2012;46(8):815-822.

49. van Merrienboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Medical education.* 2010;44(1):85-93.

50. Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication. *Academic medicine : journal of the Association of American Medical Colleges.* 1990;65(10):611-621.

51. Eva KW. What every teacher needs to know about clinical reasoning. *Medical education.* 2005;39(1):98-106.

52. Guerrasio J, Aagaard EM. Methods and Outcomes for the Remediation of Clinical Reasoning. *Journal of general internal medicine.* 2014;29(12):1607-1614.

53. Schubach F, Goos M, Fabry G, Vach W, Boeker M. Virtual patients in the acquisition of clinical reasoning skills: does presentation mode matter? A quasi-randomized controlled trial. *BMC medical education.* 2017;17(1):165.

54. Marei HF, Donkers J, Al-Eraky MM, van Merrienboer JJG. The effectiveness of sequencing virtual patients with lectures in a deductive or inductive learning approach. *Medical teacher.* 2017;39(12):1268-1274.

55. Wu B, Wang M, Grotzer TA, Liu J, Johnson JM. Visualizing complex processes using a cognitive-mapping tool to support the learning of clinical reasoning. *BMC medical education.* 2016;16(1):216.

56. Rylander M, Guerrasio J. Heuristic errors in clinical reasoning. *The clinical teacher.* 2016;13(4):287-290.

57. Nendaz MR, Gut AM, Louis-Simonet M, Perrier A, Vu NV. Bringing explicit insight into cognitive psychology features during clinical reasoning seminars: a prospective, controlled study. *Education for health (Abingdon, England).* 2011;24(1):496.

58. McLaughlin K, Heemskerk L, Herman R, Ainslie M, Rikers RM, Schmidt HG. Initial diagnostic hypotheses bias analytic information processing in non-visual domains. *Medical education.* 2008;42(5):496-502.

59. Baghdady M, Carnahan H, Lam EW, Woods NN. Test-enhanced learning and its effect on comprehension and diagnostic accuracy. *Medical education.* 2014;48(2):181-188.

60. Raupach T, Andresen JC, Meyer K, et al. Test-enhanced learning of clinical reasoning: a crossover randomised trial. *Medical education.* 2016;50(7):711-720.

61. Armstrong B, Spaniol J, Persaud N. Does exposure to simulated patient cases improve accuracy of clinicians' predictive value estimates of diagnostic test results? A within-subjects experiment at St Michael's Hospital, Toronto, Canada. *BMJ open.* 2018;8(2):e019241.

62. Blissett S, Morrison D, McCarty D, Sibbald M. Should learners reason one step at a time? A randomised trial of two diagnostic scheme designs. *Medical education.* 2017;51(4):432-441.

63. Lisk K, Agur AMR, Woods NN. Exploring cognitive integration of basic science and its effect on diagnostic reasoning in novices. *Perspectives on medical education.* 2016;5(3):147-153.