Development of a Clinical Reasoning Documentation Assessment Tool for Resident and Fellow Admission Notes: a Shared Mental Model for Feedback

Verity Schaye, MD, MHPE[1,2] , Louis Miller, MD[3], David Kudlowitz, MD[1], Jonathan Chun, MD[4], Jesse Burk-Rafel, MD, MRes[1], Patrick Cocks, MD[1], Benedict Guzman, MS[1], Yindalon Aphinyanaphongs, MD, PhD[1], and Marina Marin, MSc[1]

[1]NYU Grossman School of Medicine, New York, NY, USA; [2]NYC Health + Hospitals/Bellevue, New York, NY, USA; [3]Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, USA; [4]Stanford University School of Medicine, Stanford, CA, USA.

BACKGROUND: Residents and fellows receive little feedback on their clinical reasoning documentation. Barriers include lack of a shared mental model and variability in the reliability and validity of existing assessment tools. Of the existing tools, the IDEA assessment tool includes a robust assessment of clinical reasoning documentation focusing on four elements (interpretive summary, differential diagnosis, explanation of reasoning for lead and alternative diagnoses) but lacks descriptive anchors threatening its reliability.

OBJECTIVE: Our goal was to develop a valid and reliable assessment tool for clinical reasoning documentation building off the IDEA assessment tool.

DESIGN, PARTICIPANTS, AND MAIN MEASURES: The Revised-IDEA assessment tool was developed by four clinician educators through iterative review of admission notes written by medicine residents and fellows and subsequently piloted with additional faculty to ensure response process validity. A random sample of 252 notes from July 2014 to June 2017 written by 30 trainees across several chief complaints was rated. Three raters rated 20% of the notes to demonstrate internal structure validity. A quality cut-off score was determined using Hofstee standard setting.

KEY RESULTS: The Revised-IDEA assessment tool includes the same four domains as the IDEA assessment tool with more detailed descriptive prompts, new Likert scale anchors, and a score range of 0–10. Intraclass correlation was high for the notes rated by three raters, 0.84 (95% CI 0.74–0.90). Scores ≥6 were determined to demonstrate high-quality clinical reasoning documentation. Only 53% of notes (134/252) were high-quality.

CONCLUSIONS: The Revised-IDEA assessment tool is reliable and easy to use for feedback on clinical reasoning documentation in resident and fellow admission notes with descriptive anchors that facilitate a shared mental model for feedback.

INTRODUCTION

Clinical reasoning is a core competency of medical training.[1] There are multiple components of the clinical reasoning process for trainees to master including information gathering, hypothesis generation, forming a problem representation (a concise synthesis of the patient's presentation including key demographics and risk factors, temporal pattern of illness, and the clinical syndrome), generating a differential diagnosis, selecting a leading or working diagnosis, providing a diagnostic justification, and developing a management or treatment plan.[2–4] Written notes are a potential avenue for assessing a trainee's competency in the clinical reasoning process.[3] To this end, the ACGME Internal Medicine milestone ICS3 is "appropriate utilization and completion of medical records including effective communication of clinical reasoning."[5] While the hope was that the advent of electronic health records (EHRs) would help facilitate improvement in documentation quality, it has not achieved this goal and failure to demonstrate clinical reasoning in documentation persists.[6–9]

Feedback plays a key role in trainees' development of competency in documenting their clinical reasoning.[3,10,11] Despite this relationship, feedback on notes is typically infrequent, with barriers including time limitations of supervising faculty, lack of a shared mental model for high-quality clinical reasoning, and variability in the reliability of assessment tools that do exist.[12–17]

There are several note rating instruments that have been validated to assess documentation quality such as QNOTE, PDQI-9, the RED checklist, the HAPA form, the P-HAPEE rubric, the IDEA assessment tool, and the CRANAPL assessment rubric.[8,11,18–22] However, these note rating instruments possess varying degrees of detailed evaluation of clinical

reasoning. Of these note rating instruments, the IDEA assessment tool developed by Baker et al. includes a robust assessment of clinical reasoning documentation. The rating tool focuses on four elements of the assessment section of clinical notes, including Interpretive summary (I), Differential diagnosis with commitment to the most likely diagnosis (D), Explanation of reasoning in choosing the most likely diagnosis (E), and Alternative diagnosis with explanation of reasoning (A).[20] Each of these components is rated on a 3-point Likert scale. While a useful framework, the IDEA assessment tool lacks descriptive anchors for each of these domains which threatens its reliability, and its authors identified standard setting for item rating stringency as a future direction.[20]

Our goal was to develop a valid and reliable assessment tool for clinical reasoning documentation building off the IDEA assessment tool. Here, we discuss the process of developing and validating the Revised-IDEA assessment tool with standard setting for item rating in order to achieve increased reliability and create a shared mental model for feedback on clinical reasoning documentation.

METHODS

Study Design

This study was conducted at a large academic medical center in New York City. A team of physicians with expertise in clinical reasoning, assessment, and feedback was assembled to iteratively review admission notes written by internal medicine residents and medicine subspecialty fellows using the IDEA assessment tool and other evidence-based assessment frameworks in order to develop and validate the Revised-IDEA assessment tool with descriptive anchors for each of the four core domains: interpretive summary, differential diagnosis, explanation of lead diagnosis, alternative diagnosis explained.[3,20] Validity evidence for the Revised-IDEA assessment tool was generated using components of Messick's validity framework: content validation, response process, internal structure, and consequences.[23]

Content Validation

The panel of physicians consisted of an internal medicine chief resident and three clinician educators, including two hospitalists, one who is the site director for the internal medicine residency program and the other assistant director of curricular innovation for the Institute for Innovations in Medical Education, and a cardiologist who was the senior associate program director for the internal medicine residency program. This group had expertise in assessment and feedback and development of clinical reasoning curricula across the educational continuum—undergraduate medical education, graduate medical education, and faculty development.[24–26] This panel iteratively reviewed the assessment and plans of randomly selected de-identified admission notes from the NYU Langone EHR

written by internal medicine residents and medicine subspecialty fellows using the IDEA assessment tool core domains. To help ensure the selected notes were diagnostic in nature rather than more management related, pre-procedure notes and admission notes written after 3 days of the admitting date were excluded. Both residents and fellows were included in the review to help assess whether the tool would be relevant for both of these learner groups. Not having access to the full medical chart, the panel focused on assessing for demonstration of clinical reasoning and was not reviewing for accuracy of clinical reasoning although the intent in practice for a supervising faculty member would be to assess for both. After each round of note rating, the panel met to discuss discordant note ratings, identify reasons for differences, and establish consensus for modifications and descriptive anchors in each domain of the IDEA assessment tool grounded in clinical reasoning theory.[2,3,27] After 3 rounds and a review of 90 notes in total, the Revised-IDEA assessment tool was created.

Response Process

The Revised-IDEA assessment tool generated from the above process was then piloted with several faculty outside of the initial panel of physicians. They discussed with members of the panel if the rubric was sufficiently detailed, measured the intended construct of clinical reasoning documentation quality, and what training would be needed for future raters to use this tool. This feedback was incorporated into the final version of the Revised-IDEA assessment tool.

Data Collection

Once the Revised-IDEA assessment tool was finalized, 258 notes from July 2014 to June 2017 were randomly selected from the NYU Langone EHR written by medicine residents and medicine subspecialty fellows. There were no major changes in the EHR that would have significantly impacted note content during this time period. Of the 258 notes, 6 were excluded given the assessment and plan section of the notes were not present. The 252 notes were de-identified and the assessment and plan sections of the admission notes were reviewed and rated by the panel of physicians.

Internal Structure

To build further validity evidence and assess for inter-rater reliability, 3 physicians from the panel rated 50 (20%) of the 252 notes. For these notes rated by multiple raters, the average score of the raters was used in the final data analysis. The remainder of the notes were divided up evenly among 3 raters and rated by a single rater on the panel.

Consequences

In order to determine a cut-off for what constituted high- vs. low-quality clinical reasoning documentation, the members of the panel used the Hofstee standard setting method.[28] Each of

Domain	Description	Assessment	Points
I - Interpretive Summary	Provides a concise summary statement that uses semantic vocabulary to highlight the most important elements from history, exam, and testing and to interpret and represent the patient's main problem(s). The presence or absence of the following features is assessed: a) Key risk factors; b) Chief complaint; c) Illness time course; and d) Use of semantic qualifiers (e.g. monoarticular vs polyarticular) or unified medical concepts (e.g. volume overload, cardiovascular risk factors). NB: Some problems have an implied time course (e.g. syncope, seizure).	No features present	0
		1 feature present	1
		2 features present	2
		3 features present	3
		4 features present	4
D – Differential Diagnosis	Offers more than one relevant diagnostic possibility, committing to what is most likely and considering what is less likely or unlikely yet important to consider for the main chief complaint. If the chief complaint is a diagnosis or syndrome (e.g. acute on chronic systolic heart failure) then differential to rate may be around the differential for that exacerbation (e.g. medication non-compliance vs. arrhythmia).	No differential	0
		Differential is implicitly stated, given as a diagnostic category (e.g. "cardiac"), OR implicitly prioritized	1
		Differential is explicitly stated AND explicitly prioritized	2
E – Explanation of Lead Diagnosis	Explains the reasoning behind the lead diagnosis, including the epidemiology and key features and how these compare with the patient's presentation. If objective data points are not clearly linked to the lead diagnosis or alternative diagnosis, then only designate points to lead OR alternative diagnosis and NOT both.	No explanation	0
		1 objective data point in explanation of lead diagnosis	1
		≥2 objective data points in explanation of lead diagnosis	2
A – Alternative Diagnosis Explained	Explains the reasoning behind alternative diagnoses, including the epidemiology and key features and how these compare with the patient's presentation and alternative diagnosis. If objective data points are not clearly linked to the lead diagnosis or alternative diagnosis, then only designate points to lead OR alternative diagnosis and NOT both.	No explanation for any alternative diagnosis	0
		1 objective data point in explanation of at least one alternative diagnosis	1
		≥2 objective data points in explanation of at least one alternative diagnosis	2
Revised-IDEA Score	Overall evaluation of demonstration of clinical reasoning in the assessment section of admission notes.	Sum of I + D + E + A points (score ≥6 indicates high-quality clinical reasoning documentation)	0-10

Figure 1 The Revised-IDEA assessment tool for clinical reasoning documentation.

the four physicians determined what the minimally acceptable and maximally acceptable cut-off scores were for the Revised-IDEA assessment score as well as the minimally acceptable and maximally acceptable failure rates. Averages of the four physicians were used to then plot the possible cut-off points against acceptable failure rates. Then, point A, where the minimally acceptable cut-off score intersected with the maximally acceptable failure rate, and point B, where the maximally acceptable cut-off score intersected with the minimally acceptable failure rate, were determined. The point where the line joining A and B intersected with the distribution curve, generated from the rating results of the 252 notes, was the determined cut-off score for high- vs. low-quality clinical reasoning documentation.

Data Analysis

De-identified data were analyzed using a standard statistical software program (SPSS version 25, Chicago, IL, USA). Descriptive statistics were computed. For the notes rated by 3 raters using the final Revised-IDEA assessment tool, intraclass correlation was calculated for inter-rater reliability. Intraclass correlation was also calculated for the first round of ratings using the original IDEA assessment in the iterative process tool for comparison. To further assess for internal structure of the Revised-IDEA assessment tool, Cronbach α was calculated to measure consistency between rubric items. The study was approved by the New York University Grossman School of Medicine institutional review board.
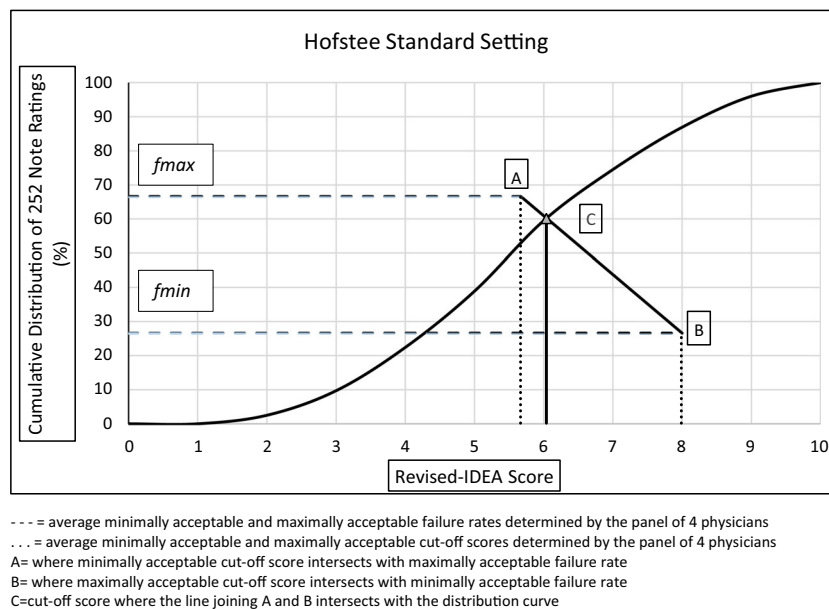
Figure 2 Revised-IDEA cut-off score ≥ 6 determined by the Hofstee standard setting method. Dashed lines indicate average minimally acceptable and maximally acceptable failure rates determined by the panel of 4 physicians. Dotted lines indicate average minimally acceptable and maximally acceptable cut-off scores determined by the panel of 4 physicians. A, where minimally acceptable cut-off score intersects with maximally acceptable failure rate. B, where maximally acceptable cut-off score intersects with minimally acceptable failure rate. C, Revised-IDEA cut-off score where the line joining A and B intersects with the distribution curve.

RESULTS

The final Revised-IDEA assessment tool had the same four core domains of the original IDEA assessment tool—interpretive summary, differential diagnosis, explanation of lead diagnosis, alternative diagnosis explained—with more detailed descriptions in the prompts and descriptive anchors for the Likert rating scale in each of the four domains (Fig. 1). The final Revised-IDEA assessment tool had a score range from 0 to 10 with a possible score of 0–4 for interpretive summary and 0–2 in each of the remaining 3 categories of differential diagnosis, explanation of lead diagnosis, and alternative diagnosis explained for a total possible score of 0–6 in the DEA sections combined (Fig. 1). The mean total Revised-IDEA score of note ratings was 5.75 (SD 2.01). Mean scores of each of the Revised-IDEA categories were I = 3.37 (SD 0.73), D = 0.99 (SD 0.74), E = 0.88 (SD 0.76), and A = 0.49 (SD 0.78). A total of 13 internal medicine residents and 17 medicine subspecialty fellows wrote these notes with an average of 8 notes per learner. In total, 185/252 notes had chief complaint data available and many notes had multiple chief complaints for a total of 197 chief complaints. The most frequent chief complaints were shortness of breath, 61/197 (31%); chest pain, 39/197 (20%); loss of consciousness, 10/197 (5%); and dizziness, 9/197 (5%).

Response Process

Given the increased specificity of the Revised-IDEA assessment tool including descriptive anchors for item rating, we found that minimal training was required for faculty to use the Revised-IDEA assessment tool. After a 15-min overview using the Revised-IDEA rubric as the only training instrument

(Fig. 1), faculty were able to use the tool with ease. Additionally, all generally found it a helpful framework in order to give feedback to their trainees on clinical reasoning documentation.

Internal Structure

Intraclass correlation was high on the 20% of notes that were scored by three raters, 0.84 (95% CI 0.74–0.90). For the 15 notes rated in the first round of the iterative process using the original IDEA assessment tool, intraclass correlation was moderate, 0.55 (95% CI −0.08 to 0.84). The Revised-IDEA assessment tool Cronbach α was 0.53, indicating moderate agreement between item scores. The agreement between the D, E, and A scores was higher with a Cronbach α 0.69.

Consequences

Through the Hofstee standard setting method, a cut-off score of ≥6 was determined to indicate high-quality clinical reasoning (Fig. 2). In total, 134/252 notes (53%) were rated as high-quality.

DISCUSSION

In this study, we developed and validated the Revised-IDEA assessment tool, which includes a more detailed rubric with descriptive anchors for item rating than the original IDEA assessment tool. We believe this enhances the existing tool as the addition of the descriptive anchors creates a shared mental model for feedback on clinical reasoning documentation focused on residents and fellows.[11,20–22] We demonstrated validity evidence for this rubric using Messick's

framework, ultimately creating a rubric that was easy to use with minimal training of faculty, thus easily implementable. Additionally, we demonstrated good reliability with an intraclass correlation between raters of 0.84, which is higher than what was described with the original IDEA assessment tool and higher than what we experienced using the original IDEA assessment tool in the first round of the iterative process of development.[20] Finally, the decision for score weighting (the interpretive summary from 0 to 4 and the D, E, and A scores from 0 to 2 each) was derived from the individual competencies a trainee needs to achieve to master the clinical reasoning process. The ability to formulate a problem representation or interpretive summary is in itself an essential skill and distinct from the ability to produce a prioritized differential diagnosis that is explained and justified.[29] The D, E, and A components of the rubric are interrelated representing the two competencies of prioritizing and justifying a differential diagnosis. Therefore, the combined DEA score component of the Revised-IDEA score holds a slightly higher weight than the interpretive summary in the overall score.

Nearly half of the resident and fellow notes were scored as low quality using the Revised-IDEA assessment tool. This is consistent with what is seen more globally with a decline in clinical reasoning documentation quality since the advent of the EHR.[6–8] Clinician educators agree that utilizing the EHR to assess learners' clinical reasoning and provide feedback is an important teaching strategy.[15] However, in order to give effective feedback, faculty and residents need a shared mental model of high-quality clinical reasoning documentation and feedback needs to be specific.[27,30] Furthermore, if the assessment tool used for feedback lacks reliability, this could result in poor understanding of expectations and impact buy-in from learners.[3,11,21] With the development of the Revised-IDEA assessment tool, we have created a reliable tool with a shared mental model to provide feedback to residents and fellows so they may achieve the milestone of effectively communicating their clinical reasoning in medical records upon completion of training.[5]

There were several limitations of this study. There was only moderate agreement between all four domains of item scores of the Revised-IDEA assessment tool. However, with the interpretative summary score removed, the agreement between the other three domains was high. A reason for this discrepancy might be that it is easier to rate the D, E, and A without clinical knowledge of the case rather than the interpretive summary. While the A scores tended to be lower than the D and E scores, this can likely be explained by the fact that learners did not spend as much time explaining their alternative diagnosis and not that this is a fundamentally different skill. Further evaluation of using the tool with faculty familiar with each case could build further validity evidence and demonstrate agreement between all four domains including the interpretive summary and provide further validity evidence for the weighting of the Revised-IDEA score.

Another limitation was we did not include the relation to other variable domains of Messick's validity framework. The sample size of learners was not sufficient to include this domain but would be an area of focus for next steps to add additional validity evidence. Lastly, this was at a single institution and the distribution of chief complaints was limited. Further validation with a note set containing a broader range of chief complaints would be additive.

CONCLUSION

The Revised-IDEA assessment tool is a reliable and easy-to-use assessment tool for feedback on clinical reasoning documentation, with descriptive anchors that facilitate a shared mental model for feedback. Next steps are to continue to refine this tool and build additional validity evidence. We ultimately plan to use this tool as the human rating gold standard to develop a machine learning algorithm for automated clinical reasoning documentation feedback, which will help overcome the barriers of time and training related to human note review.

Corresponding Author: *Verity Schaye, MD, MHPE; NYC Health + Hospitals/Bellevue, New York, NY, USA (e-mail: Verity. Schaye@nyulangone.org).*

Declarations:

Conflict of Interest: *The authors declare that they do not have a conflict of interest.*

REFERENCES

1. Accreditation Council for Graduate Medical Education. ACGME common program requirements. Available at: http://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/CPRs_2017-07-01.pdf.Revised July 1. Accessed January 17, 2020.
2. **Young M, Thomas A, Lubarsky S**, et al. Drawing boundaries: the difficulty in defining clinical reasoning. Acad Med. 2018;93(7):990-995.
3. **Daniel M, Rencic J, Durning SJ**, et al. Clinical reasoning assessment methods: a scoping review and practical guidance. Acad Med. 2019;94(6):902-912.
4. **Bowen JL**. Educational strategies to promote clinical diagnostic reasoning. N Engl J Med. 2006;355(21):2217-2225.
5. Accreditation Council for Graduate Medical Education. ACGME core program requirements. Available at: http://www.acgme.org/portals/0/pdfs/milestones/internalmedicinemilestones.pdf. Accessed January 17, 2020.
6. **March CA, Scholl G, Dversdal RK**, et al. Use of electronic health record simulation to understand the accuracy of intern progress notes. J Grad Med Educ. 2016;8(2):237-40.
7. **Colicchio TK, Cimino JJ**. Clinicians' reasoning as reflected in electronic clinical note-entry and reading/retrieval: a systematic review and qualitative synthesis. J Am Med Inform Assoc. 2019;26(2):172-184.
8. **Bierman JA, Hufmeyer KK, Liss DT, Weaver AC, Heiman HL**. Promoting responsible electronic documentation: validity evidence for a checklist to assess progress notes in the electronic health record. Teach Learn Med. 2017;29(4):420-432.

9. **Tierney MJ**, **Pageler NM**, **Kahana M**, **Pantaleoni JL**, **Longhurst CA**. Medical education in the electronic medical record (EMR) era: benefits, challenges, and future directions. Acad Med. 2013;88(6):748-52.

10. **Lessing JN**, **Rendón P**, **Durning SJ**, **Roesch JJ**. Approaches to clinical reasoning assessment. Acad Med. 2020;95(8):1285.

11. **Middleman AB**, **Sunder PK**, **Yen AG**. Reliability of the history and physical assessment (HAPA) form. Clin Teach. 2011;8(3):192-5.

12. **Habboush Y**, **Hoyt R**, **Beidas S**. Electronic health records as an educational tool: viewpoint. JMIR Med Educ. 2018;4(2):e10306.

13. **Varpio L**, **Day K**, **Elliot-Miller P**, et al. The impact of adopting EHRs: how losing connectivity affects clinical reasoning. Med Educ. 2015;49(5):476-86.

14. **Berndt M**, **Fischer MR**. The role of electronic health records in clinical reasoning. Ann NY Acad Sci. 2018;1434(1):109-114.

15. **Atwater AR**, **Rudd M**, **Brown A**, et al. Developing teaching strategies in the EHR era: a survey of GME experts. J Grad Med Educ. 2016;8(4):581-586.

16. **Schenarts PJ**, **Schenarts KD**. Educational impact of the electronic medical record. J Surg Educ. 2012;69(1):105-12.

17. **Pageler NM**, **Friedman CP**, **Longhurst CA**. Refocusing medical education in the EMR era. JAMA. 2013;310(21):2249-50.

18. **Burke HB**, **Hoang A**, **Becher D**, et al. QNOTE: an instrument for measuring the quality of EHR clinical notes. J Am Med Inform Assoc. 2014;21(5):910-6.

19. **Stetson PD**, **Bakken S**, **Wrenn JO**, **Siegler EL**. Assessing electronic note quality using the physician documentation quality instrument (PDQI-9). Appl Clin Inform. 2012;3(2):164-174.

20. **Baker EA**, **Ledford CH**, **Fogg L**, **Way DP**, **Park YS**. The IDEA assessment tool: assessing the reporting, diagnostic reasoning, and decision-making skills demonstrated in medical students' hospital admission notes. Teach Learn Med. 2015;27(2):163-73.

21. **King MA**, **Phillipi CA**, **Buchanan PM**, **Lewin LO**. Developing validity evidence for the written pediatric history and physical exam evaluation rubric. Acad Pediatr. 2017;17(1):68-73.

22. **Kotwal S**, **Klimpl D**, **Tackett S**, **Kauffman R**, **Wright S**. Documentation of clinical reasoning in admission notes of hospitalists: validation of the CRANAPL assessment rubric. J Hosp Med. 2019;14:E1-e8.

23. **Cook DA**, **Beckman TJ**. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med. 2006;119(2):166.e7-16.

24. **Schaye V**, **Janjigian M**, **Hauck K**, et al. A workshop to train medicine faculty to teach clinical reasoning. Diagn. 2019;6(2):109-113.

25. **Schaye V**, **Eliasz KL**, **Janjigian M**, **Stern DT**. Theory-guided teaching: implementation of a clinical reasoning curriculum in residents. Med Teach. 2019;41(10):1192-1199.

26. **Horlick M**, **Miller L**, **Cocks P**, **Bui L**, **Schwartz M**, **Dembitzer A**. Calling it like you see it: three-hour workshop improves hospitalists observation and feedback skills. Abstracts from the 38th annual meeting of the society of general internal medicine. J Gen Intern Med. 2015;30(Suppl 2):45-551.

27. **Thammasitboon S**, **Rencic JJ**, **Trowbridge RL**, **Olson AP**, **Sur M**, **Dhaliwal G**. The assessment of reasoning tool (ART): structuring the conversation between teachers and learners. Diagn. 2018;5(4):197-203.

28. **Bandaranayake RC**. Setting and maintaining standards in multiple choice examinations: AMEE Guide No. 37. Med Teach. 2008;30(9-10):836-845.

29. **Olson A**, **Rencic J**, **Cosby K**, et al. Competencies for improving diagnosis: an interprofessional framework for education and training in health care. Diagn. 2019;6(4):335-341.

30. **Ende J.** Feedback in clinical medical education. JAMA. 1983;250(6):777-781.