Development and Validation of a Formative Assessment Tool for Nephrology Fellows' Clinical Reasoning

Suzanne M. Boyle,[1] James Martindale,[2] Andrew S. Parsons,[3] Stephen M. Sozio,[4] Rachel Hilburg,[5] Jehan Bahrainwala,[6] Lili Chan,[7] Lauren D. Stern,[8] and Karen M. Warburton[9]

Abstract

Background Diagnostic errors are commonly driven by failures in clinical reasoning. Deficits in clinical reasoning are common among graduate medical learners, including nephrology fellows. We created and validated an instrument to assess clinical reasoning in a national cohort of nephrology fellows and established performance thresholds for remedial coaching.

Methods Experts in nephrology education and clinical reasoning remediation designed an instrument to measure clinical reasoning through a written patient encounter note from a web-based, simulated AKI consult. The instrument measured clinical reasoning in three domains: problem representation, differential diagnosis with justification, and diagnostic plan with justification. Inter-rater reliability was established in a pilot cohort (n=7 raters) of first-year nephrology fellows using a two-way random effects agreement intraclass correlation coefficient model. The instrument was then administered to a larger cohort of first-year fellows to establish performance standards for coaching using the Hofstee method (n=6 raters).

Results In the pilot cohort, there were 15 fellows from four training program, and in the study cohort, there were 61 fellows from 20 training programs. The intraclass correlation coefficients for problem representation, differential diagnosis, and diagnostic plan were 0.90, 0.70, and 0.50, respectively. Passing thresholds (% total points) in problem representation, differential diagnosis, and diagnostic plan were 59%, 57%, and 62%, respectively. Fifty-nine percent (n=36) met the threshold for remedial coaching in at least one domain.

Conclusions We provide validity evidence for a simulated AKI consult for formative assessment of clinical reasoning in nephrology fellows. Most fellows met criteria for coaching in at least one of three reasoning domains, demonstrating a need for learner assessment and instruction in clinical reasoning.

CJASN ■: 1–9, 2023. doi: https://doi.org/10.2215/CJN.0000000000000315

Due to the number of contributing authors, the affiliations are listed at the end of this article.

Correspondence:
Dr. Suzanne M. Boyle, Section of Nephrology, Hypertension, and Kidney Transplantation, Lewis Katz School of Medicine, Temple University, 3440 North Broad Street, Suite 100 Kresge West, Philadelphia, PA 19140. Email: Suzanne.Boyle@tuhs.temple.edu

Introduction

Diagnostic errors are common and account for 40,000–80,000 deaths per year.[1,2] There are multiple causes of diagnostic errors, but the majority are driven by cognitive failures, which may reflect inadequate medical knowledge or application of knowledge.[3,4] The application of knowledge to patient care is known as clinical reasoning—the process by which clinicians integrate, organize, and interpret data to arrive at a diagnosis and management plan.[5,6] Clinical reasoning is a complex, iterative process, but for educational and remediation purposes, it can be thought of as a linear pathway that includes hypothesis generation, data gathering, problem representation, prioritized differential diagnosis, selection of a working diagnosis, and formation of a management plan.[5] The pathway has conscious and unconscious components and is subject to cognitive biases that predispose to diagnostic error (Supplemental Figure 1).[7,8]

Clinical reasoning can go awry at any step in the pathway, from eliciting the patient's chief concern to development of the management plan. Cognitive biases are frequent culprits among experts who commonly use pattern recognition.[7,9,10] Among novices, who often default to analytic thinking given their limited experience, failures can stem from a primary weakness in any step(s) of the pathway.[11,12] For example, some learners gather clinical data haphazardly rather than in a hypothesis-driven manner, leading to failure of illness script activation (*i.e.*, a clinician's organized knowledge of disease and associated presentations). Other learners might struggle with problem representation, that is, an inability to concisely summarize the key features of a patient's primary problem during or after data gathering.[5] Lack of this skill can impede generation of a prioritized differential diagnosis and lead to cognitive overload, particularly when the learner is caring for multiple complex patients. Still, other learners might lack schema to compare and contrast features of competing illness scripts and, consequently, struggle to generate or prioritize a differential diagnosis.

Research supports deliberately teaching clinical reasoning to learners.[13] Understanding the cognitive steps

that drive decision making promotes metacognition and may reduce diagnostic error.[9,10] There is evidence that some learners do not adequately develop their reasoning through clinical exposure alone.[14] Among undergraduate and graduate medical learners who require remediation to meet Accreditation Council for Graduate Medical Education (ACGME) competencies, clinical reasoning deficits are documented in 25%–45%.[15,16] These learners often come to attention because of disorganized presentations and patient hand-offs or "missing the big picture" with cases. They are often "misdiagnosed" as having medical knowledge deficits or problems with efficiency. If correctly identified, remediation of reasoning deficits is often successful but more time-intensive than other deficits.[17]

Nephrology is a highly cognitive specialty that serves the most medically complex patients.[18] Most nephrology fellows are responsible for an average of 20 hospitalized patients per day.[19] This combination of high volume and medical complexity makes nephrologists vulnerable to diagnostic errors. Indeed, a national survey of nephrology training program directors indicated that 88% remediated at least one fellow over 5 years, and in 35% of the cases, the indication was clinical reasoning.[20] Therefore, it is critical to identify fellows who have significant reasoning deficits early in their training so that they can receive remediation.[21] Given that the term remediation is stigmatizing, we use "clinical coaching" to signify a process by which a learner receives individualized guidance for a deficient skill. Currently, there are no validated methods for assessing clinical reasoning in nephrology fellows. Here, we describe the creation and validation of an instrument for formative assessment of nephrology fellows' clinical reasoning using a simulated consult for AKI. We also established performance thresholds to identify fellows who will benefit from clinical coaching to improve reasoning.

Methods

Instrument Development and Study Design

We sought to develop and provide validity evidence for an instrument (Reasoning Evaluation in Nephrology Education) to measure the construct of clinical reasoning in first-year nephrology fellows. We chose to measure reasoning through a written patient encounter note as part of a simulated, web-based AKI consult. The advantages to this approach are that it does not require a proctor, produces an enduring end product, and mimics a real-world activity. Furthermore, we were able to model the instrument on the basis of several previously validated instruments that assess reasoning through written notes (interpretative summary, differential diagnosis, explanation of reasoning, and alternatives, Clinical Reasoning in Admission Note Assessment and Plan, and Diagnostic Justification).[22–24]

The simulated encounter was accessed *via* a website created with WordPress and hosted by Temple University (https://sites.temple.edu/rene/case-2/; not publicly searchable). It included a 2-minute video of a standardized patient in an emergency department describing the history of her present illness with text captions. Drop-down menus showed home medications, vital signs, intake/output, physical examination, laboratory/imaging results, and an emergency department event log. The case was designed to prompt a broad differential diagnosis and consideration of both intrinsic causes of AKI and extrinsic causes from systemic conditions. There was no intended correct diagnosis; rather, the case was intentionally designed to be ambiguous because the objective was to measure the learner's thought process. The topic of AKI was chosen to reflect a commonly encountered condition in medical residency that does not require specialized nephrology knowledge. Fellows were directed to write (*1*) a problem representation in the form of an interpretive summary statement, (*2*) a differential diagnosis with justification (including a lead diagnosis), and (*3*) a diagnostic plan with justification. Each of these domains map to a step(s) in the reasoning pathway and was scored (Figure 1; see Scoring Instrument in the Supplemental Appendix).

As noted above, several previously validated instruments for the assessment of clinical reasoning were used to inform the development of the tool, including the domains and the elements tested within each domain. For example, interpretative summary, differential diagnosis, explanation of reasoning, and alternatives is a well-known framework for assessing diagnostic reasoning, predominantly in medical students. Its "interpretative summary" is synonymous with the problem representation, and the "differential diagnosis, explanation of reasoning, and alternatives" are consistent with what is tested in the differential diagnosis with justification domain.[22] Similarly, the Diagnostic Justification framework requires medical students to provide written justification for their differential diagnosis with a lead diagnosis and alternates, including pertinent positives and negatives.[24] Finally, the Clinical Reasoning in Admission Note Assessment and Plan tool was validated among hospitalists. It tests the diagnostic reasoning domains of problem representation and differential diagnosis (assessing for a lead diagnosis, alternate diagnoses, justification for diagnoses, and the level of uncertainty for the lead diagnosis). It also assesses management reasoning through justification of a diagnostic and treatment plan.[23] Figure 1 presents our instrument's three domains and the elements tested within them.

We recruited two cohorts of first-year nephrology fellows from ACGME-accredited US training programs to participate in the validation of the instrument. The pilot cohort (target enrollment, 15–30 fellows) was recruited in the spring of 2022 from nephrology training programs at the investigators' home institutions and was used to establish inter-rater reliability. The study cohort (target enrollment, 50–100 fellows) was used to measure fellows' performance on the instrument and set thresholds for coaching. It was recruited in the fall of 2022 by offering participation to randomly sampled programs in each mainland US time zone to achieve balanced geographic representation. This was followed by a nationwide invitation to training program directors on the American Society of Nephrology's Training Program Director Exchange.

Fellows received information on the objectives, risks, and benefits of the study. Their written responses were transmitted from the website to a neutral third party ("honest broker"), who exported them to a secure, web-based data manager (Redcap) and deidentified them for the investigators to score. Fellows were given the option of providing their name and program, which would allow them to

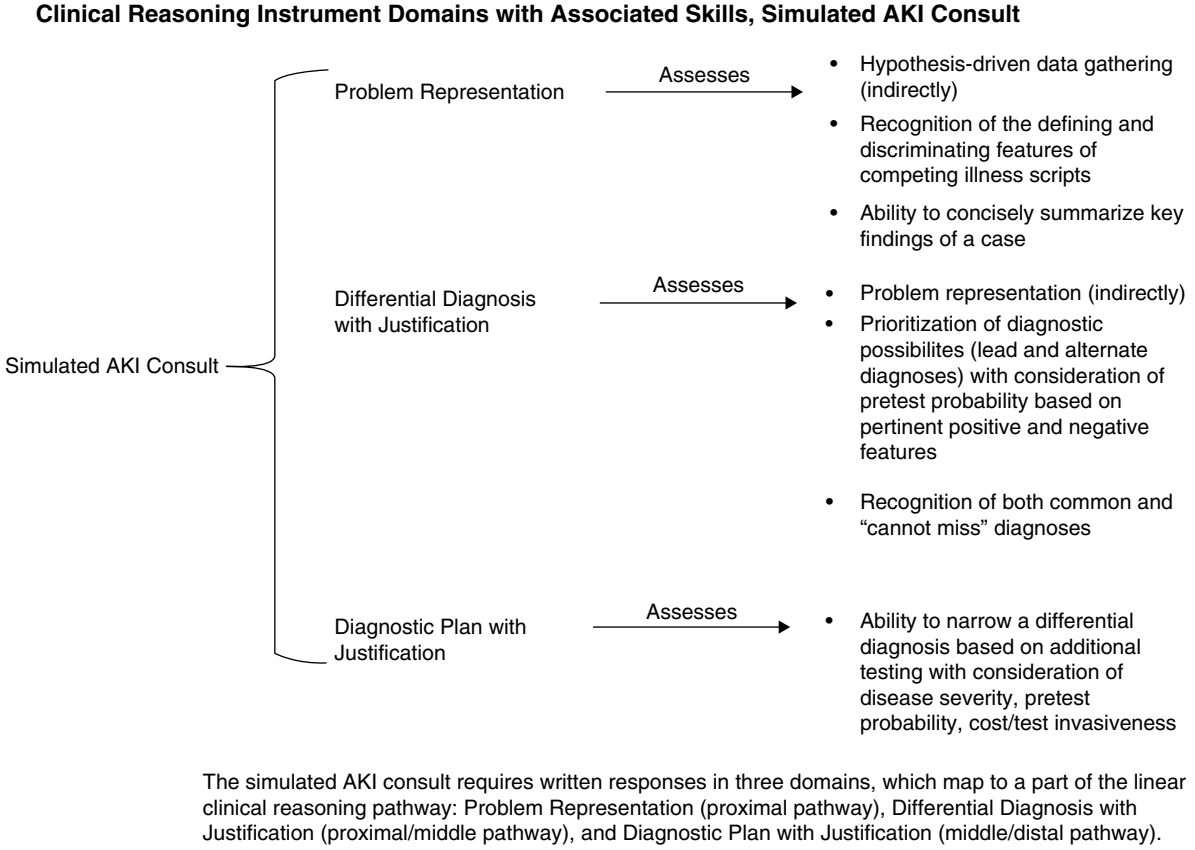

Figure 1. Clinical reasoning instrument domains with associated skills, simulated AKI consult. The simulated AKI consult requires written responses in three domains, which map to a part of the linear clinical reasoning pathway: problem representation (proximal pathway), differential diagnosis with justification (proximal/middle pathway), and diagnostic plan with justification (middle/distal pathway).

receive feedback on their performance at the conclusion of the study through their program director. The investigators were blinded to the identity of the fellow and their program. The honest broker linked the scores to the fellow's identity at the conclusion of the study and emailed them to the program director. The study protocol was approved by the Temple University Institutional Review Board.

Validity

The instrument and scoring protocol were developed by four clinician educators (S.M. Boyle, S.M. Sozio, A.S. Parsons, and K.M. Warburton). Three have expertise in nephrology (S.M. Boyle, S.M. Sozio, and K.M. Warburton). Two have expertise in the formal assessment and remediation of learners with clinical reasoning deficits (A.S. Parsons and K.M. Warburton). The scoring protocol was modeled on previously validated instruments and modified to incorporate features that were specific to our simulated patient encounter.[22–24] We initially administered the instrument to six volunteer nephrology educators and revised the instrument iteratively on the basis of their responses. We further revised the protocol after scoring the pilot cohort. All revisions were finalized before measuring inter-rater reliability in the pilot cohort.

To ensure response process validity, we surveyed participants on whether the case simulation was similar to AKI consults in their hospital and its perceived level of difficulty. We also hosted focus groups with five volunteer members of the pilot cohort to get feedback on the comprehensibility of the directions and the user friendliness of the web interface. No changes to the content of the case simulation were made based on these steps. To determine the generalizability of performance on our instrument to the broader first-year fellow population, we surveyed participants on the following: English as a first language, type of medical school (allopathic versus osteopathic versus international), completion of a US-based internal medicine residency, and previous formal instruction in clinical reasoning.

The internal consistency of the items tested in each domain was measured with Cronbach alpha.

Reliability

Once the scoring protocol was finalized, seven raters independently scored the pilot cohort to establish inter-rater reliability. Four developed the instrument and scoring protocol (S.M. Boyle, S.M. Sozio, A.S. Parsons, and K.M. Warburton). Three were expert nephrology educators (L. Chan, J. Bahrainwala, and L.D. Stern), who were trained by the principal investigator (S.M. Boyle) in application of the scoring protocol. Inter-rater reliability for the seven raters was measured using a two-way random effects agreement intraclass correlation coefficient model.[25]

Standard Setting

After establishing inter-rater reliability in the pilot cohort, we used the study cohort to identify a score in each domain (problem representation, differential diagnosis, and diagnostic plan) below which coaching is recommended.

The threshold score for coaching (*i.e.*, "pass/fail" score) was established based on the study cohort using the Hofstee approach to standard setting. The Hofstee uses both criterion- and norm-referenced methods. It is criterion-referenced in that set performance ranges are determined *a priori* based on expert determination. The methodology is norm-referenced in that the predetermined set performance expectations are plotted against aggregated learner performance—the entire distribution of examination takers is taken into consideration, which combines expert clinician judgment with aggregated learner performance.[26] J. Martindale, an expert in assessment methodology, trained six raters (S.M. Boyle, K.M. Warburton, A.S. Parsons, S.M. Sozio, J. Bahrainwala, and L. Chan) in the Hofstee method before its application.

Program directors were given a Score Interpretation Guide (Supplemental Appendix), which explained what skills were being assessed in each domain and the rationale behind each question. The Score Interpretation Guide also describes coaching exercises to strengthen skills within each domain.[27]

Statistical analyses were performed using SPSS v28.

Results

Study Participants

The pilot cohort included 15 first-year fellows from four nephrology training programs (spring 2022). The study cohort included 61 first-year fellows from 20 programs (fall 2022), both university- and community-based. Participants were largely international medical graduates (53%, pilot; 66%, study cohort) who completed a US-based internal medicine residency (93%, pilot; 90%, study cohort). English was a first language in 47% of the pilot cohort and 59% of the study cohort. Table 1 presents the characteristics of both cohorts. Their medical school training was similar to nephrology fellows nationwide (66%, international medical graduate; 21%, allopathic; 12%, osteopathic; 93%, US-based residency).[28,29]

Fifty-three percent (n=8) of the pilot cohort and 72% (n=44) of the study cohort took between 16 and 45 minutes to complete the simulation. It was rated as "moderately difficult" by 60% (n=9) of the pilot cohort and 62% (n=38) of the study cohort. Ninety-three percent (n=14) of the pilot cohort and 98% (n=60) of the study cohort thought the simulation was at least "somewhat similar" to AKI consults at their hospital. No one reported that the directions for the simulation were "not easy to understand."

Inter-Rater Reliability, Internal Consistency, and Instrument Administration

In the pilot cohort, the intraclass correlation coefficient values for inter-rater reliability in the three domains were problem representation, 0.9 ("good"); differential diagnosis with justification, 0.7 ("moderate"); and diagnostic plan with justification, 0.5 ("moderate").

Cronbach alpha for the internal consistency of items tested within each of the three domains in both the pilot and study cohorts were problem representation, 0.97 ("excellent"); differential diagnosis with justification, 0.95 ("excellent"); and diagnostic plan with justification, 0.88 ("good").

On average, the instrument took approximately 15 minutes for the raters to score.

Instrument Performance

The Hofstee method was used to identify a minimum score in each domain (expressed in percentage of total possible points) below which coaching is recommended. For problem representation, it was 59%; for differential diagnosis with justification, it was 57%; and for diagnostic plan with justification, it was 62% (Figure 2). These performance thresholds were applied to the study cohort (N=61). Twenty-three percent of fellows (n=14) fell below the threshold for problem representation; 33% (n=20) for differential diagnosis with justification; and 23% (n=14) for diagnostic plan with justification. Fifty-nine percent (n=36) of fellows fell below the performance threshold in at least one domain. Twenty-five fellows fell below in one domain (41%), ten fellows (16%) in two domains, and one fellow (2%) in three domains. The mean percentage of total possible points (by domain) is presented in Table 2.

Discussion

Our expert team of nephrology educators and clinical reasoning coaches developed a novel instrument for formative assessment of clinical reasoning using a simulated AKI consult. We provide validity evidence to support its use in first-year nephrology fellows through administration in a cohort, comprised of approximately 15% of first-year fellows in ACGME training programs.[28] Most fellows thought the instrument was a realistic representation of an AKI consult and reported it to be moderately difficult. Directions were easily understood. Inter-rater reliability for problem representation was good, and for differential diagnosis and diagnostic plan, it was moderate.

Clinical reasoning deficits in struggling residents and fellows are at least as common as deficits in medical knowledge.[15,16,30] However, these deficits are often "misdiagnosed" as problems with knowledge or efficiency. Evaluators may be conditioned to blame knowledge first when they encounter a struggling learner. "Needs to read more," is a common component of constructive feedback for the struggling learner and is often a signal that something is wrong—but it is often not knowledge. Common phenotypes associated with primary clinical reasoning deficits include a learner who "lacks the big picture," gives disorganized presentations, is unprepared for rounds or constantly behind in clinic, and cannot triage tasks or recognize clinical urgency.[31]

In a recent survey of nephrology training program directors, 88% of respondents reported remediating at least one fellow in 5 years. Approximately 35% of these fellows had clinical reasoning deficits.[20] Given the challenges with recognizing clinical reasoning deficits, this is likely an underestimate of its true prevalence. Nephrology fellows are expected to care for a high volume of medically complex

Table 1. Characteristics of first-year fellows who completed the clinical reasoning assessment instrument

Characteristic	Pilot Cohort (*N*=15)	Study Cohort (*N*=61)
Type of medical school, *n* (%)		
US allopathic	6 (40)	13 (21)
US osteopathic	1 (7)	8 (13)
International	8 (53)	40 (66)
Did you complete an internal medicine residency in the United States? *n* (%)		
Yes	14 (93)	55 (90)
No	1 (7)	6 (10)
Is English your first language? *n* (%)		
Yes	7 (47)	36 (59)
No	8 (53)	25 (41)
How difficult was the simulated AKI case? *n* (%)		
Not difficult at all	1 (7)	4 (7)
Of little difficulty	3 (20)	14 (23)
Moderately difficult	9 (60)	38 (62)
Difficult	2 (13)	5 (8)
Very difficult	0 (0)	0 (0)
How similar was the simulated AKI case to a typical AKI case at your hospital? *n* (%)		
Not at all similar	1 (7)	1 (2)
Somewhat similar	3 (20)	16 (26)
Similar	8 (53)	20 (33)
Very similar	2 (13)	20 (33)
Extremely similar	1 (7)	2 (6)
Were the directions for completing the simulated AKI case easy to understand? *n* (%)		
Not easy to understand at all	0 (0)	0 (0)
Somewhat easy to understand	2 (13)	6 (9)
Easy to understand	9 (60)	20 (33)
Very easy to understand	3 (20)	0 (0)
Extremely easy to understand	1 (7)	35 (57)
How much time (min) did you spend completing the simulated AKI case? *n* (%)		
0–15	4 (26)	9 (15)
16–30	7 (47)	19 (31)
31–45	1 (7)	25 (41)
46–60	3 (20)	7 (11)
>60	0 (0)	1 (2)
Have you ever had formal instruction in clinical reasoning (*e.g.*, a course or didactic lecture)? *n* (%)		
Yes	5 (33)	27 (44)
No	7 (47)	21 (35)
Not sure	3 (20)	13 (21)

patients, often in a busier clinical environment than that of their internal medicine residency. Therefore, fellows who performed satisfactorily as residents might first come to attention as fellows, particularly if they have not developed systems for effectively gathering and applying large amounts of data in a compressed time frame. Furthermore, busy faculty, who juggle multiple competing demands for their time, might unintentionally value arriving at a diagnosis—rather than the process of reasoning through the diagnosis—during teaching rounds.[32,33] Early identification of struggling learners is critical in fellowship given the relatively short training duration available for remediation. Clinical reasoning remediation is time-consuming. Therefore, a formative assessment of clinical reasoning in nephrology training that allows for early identification of potential weaknesses in the reasoning pathway is important.

Most fellows (59%) fell below the performance threshold in at least one of the three clinical reasoning domains. Fifteen percent were below in two domains, and only one fellow was below in three. This is notable because each domain represents a unique but inter-related step intrinsic to the clinical reasoning process. For example, an effective problem representation is a precursor to generation a differential diagnosis and diagnostic plan, yet some fellows met the threshold for coaching in problem representation but were still able to create a reasonable differential diagnosis and/or diagnostic plan. One explanation for this is that, in practice, problem representations are not always outwardly articulated, but rather subconscious mental characterizations of the patient's primary problem. The oral or written articulation of the problem representation might be an unfamiliar construct for many fellows. In practice, the failure to articulate an accurate and concise problem representation can have negative ramifications for clinician-to-clinician communication that can lead to diagnostic error.

Because strong clinical reasoning is paramount to the safe practice of nephrology, we believe there is a benefit to administering our instrument within the first several months of training. It can be used similarly to the nephrology in-training examination, which is a formative knowledge assessment that predicts board passage.[34] Medical knowledge is necessary to execute sound clinical reasoning

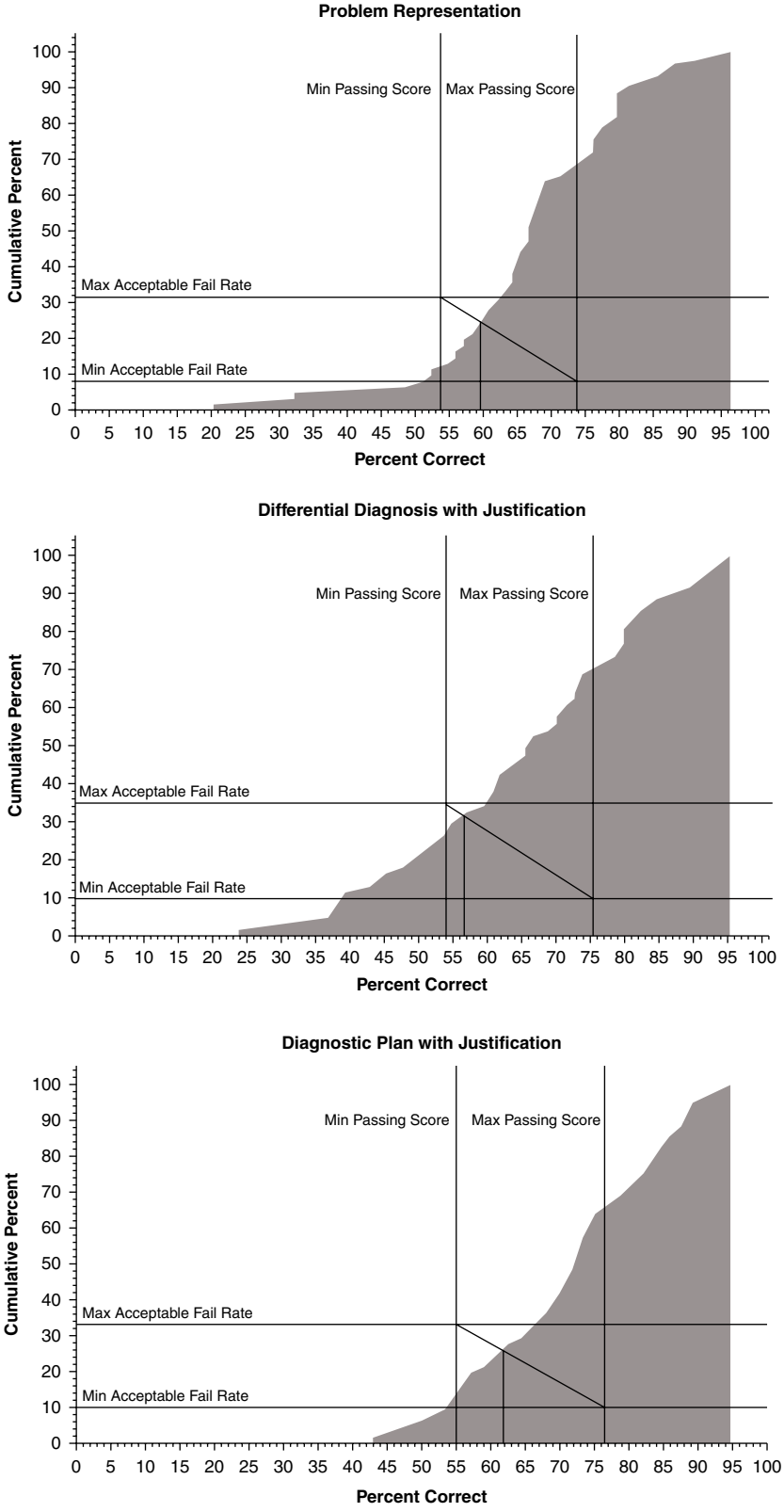

Figure 2. Establishment of passing standard with the Hofstee method, simulated AKI consult. The vertical lines represent the expert-established range of acceptable passage scores (by % total points). The horizontal lines represent the expert-established range of acceptable failure rates. The area under the curve represents cumulative performance within the study cohort (by % total points). The cut points for % total possible points below which targeted coaching is recommended were problem representation, 57%; differential diagnosis with justification, 59%; and diagnostic plan with justification, 62%.

Table 2. Mean percentage correct in each clinical reasoning domain

Domain	Pilot Cohort (N=15)	Study Cohort (N=61)
	Mean % Correct (SD)	Mean % Correct (SD)
Problem representation	71 (11)	67 (14)
Differential diagnosis with justification	71 (12)	65 (17)
Diagnostic plan with justification	80 (13)	72 (13)

but, alone, is insufficient. Therefore, using the in-training examination as the only objective metric to assess progress toward independent practice may be insufficient. Performance on our instrument underscores an unmet need for clinical reasoning coaching among nephrology fellows. Given the recent decline in nephrology board examination passage rates, it is even more critical that we institute metrics to ensure that our graduates are holistically prepared to practice independently.[35]

The strengths of our instrument are that it was developed by content experts, validated in a cohort of first-year nephrology fellows with training backgrounds similar to fellows nationwide, reflects a real-world AKI consult, and has acceptable inter-rater reliability. Time is probably a rate-limiting factor in assessing and coaching clinical reasoning in nephrology's busy clinical learning environment. Our instrument alleviates this by using a simulated consult that can be administered off the wards during an hour of conference time. The Score Interpretation Guide provides faculty coaches with an explanation of scores by question and domain, with domain-specific exercises for deliberate practice.[27]

There are limitations to our instrument. Primarily, it was designed to assess reasoning and not medical knowledge. In some instances, lack of knowledge may have limited a fellow's ability to recognize key discriminating features of the case. However, because the instrument mimics the real-word activity of note writing and requires a rationale for the differential diagnosis and diagnostic plan, a coach can easily use the instrument's output to assess for gaps in medical knowledge. In addition, our inter-rater reliability was "good" for problem representation but "moderate" for differential diagnosis and diagnostic plan. This might reflect previously identified challenges with assessing clinical reasoning.[6,36] Namely, management reasoning, as opposed to diagnostic reasoning, is a relatively new construct to be formally taught and assessed in learners. It is also inherently more complex than diagnostic reasoning because there are multiple successful management options, which consider things such as cost and individual patient values.[37] The diagnostic plan with justification assesses management reasoning and, therefore, might explain the relatively lower intraclass correlation coefficient. Finally, it is possible that some program directors enrolled their programs because they suspected clinical reasoning deficits in some of their fellows, thereby potentially overestimating the proportion with coaching needs compared with the base population. However, it is plausible that performance on the instrument is an accurate reflection of nephrology fellow clinical reasoning in early training, particularly because clinical reasoning deficits have a high rate of underdiagnosis or misdiagnosis. Furthermore, criterion-referenced formative assessments like this one are intended to function as screening tools, that is, with the aim of identifying all learners who have clinical reasoning skills below expected for their level of training and who will benefit from coaching, even at the expense of "overdiagnosing" a few.

There are few studies that systematically assess clinical reasoning in graduate medical learners the way ours does. The previously validated instruments from which our instrument was adapted were based on medical students and attending physicians.[22–24] Published data from formal remediation programs suggest that among all learners (*i.e.*, medical students and graduate medical learners), 7%–28% require remediation in the form of an individualized learning plan to achieve competence.[15] Among these, approximately 30% have clinical reasoning deficits.[15] The need for remediation is often identified from direct observation, discussion at annual summative assessment meetings, or critical events.[38,39] Formative assessment tools, like the one we developed, are not commonly used to inform remediation needs. This might be one reason that clinical reasoning deficits are underreported and likely underdiagnosed.

In conclusion, we provide validity evidence for a simulated AKI consult for formative assessment of clinical reasoning in first-year nephrology fellows. Fifty-nine percent of the cohort met the criteria for coaching by scoring below the passing standard in at least one of the three domains. This demonstrates a need for formative assessment, coaching, and faculty development in clinical reasoning. Implementation of our instrument can begin to address this need.

Disclosures

L. Chan reports consultancy for CSL Vifor Pharma, Inc. and research funding from NIH. R. Hilburg's spouse reports employment with and ownership interest in Lockheed Martin. A.S. Parsons reports consultancy for National Board of Medical Examiners and NEJM Education Group. L.D. Stern reports advisory or leadership role for Lightline Medical—Medical Advisory Board, speakers bureau for Baxter Inc., and other interests or relationships with Home Dialysis University. All remaining authors have nothing to disclose.

Funding

This research was supported by the William and Sandra Bennett Clinician Scholar Award from the ASN.

Acknowledgments

The investigators thank the participating nephrology training programs (Supplemental Appendix).

Author Contributions

Conceptualization: Suzanne M. Boyle, Andrew S. Parsons, Karen M. Warburton.

Data curation: Suzanne M. Boyle.

Formal analysis: James Martindale.

Funding acquisition: Suzanne M. Boyle, Karen M. Warburton.
Investigation: Jehan Bahrainwala, Suzanne M. Boyle, Lili Chan, Andrew S. Parsons, Stephen M. Sozio, Lauren D. Stern, Karen M. Warburton.
Methodology: Suzanne M. Boyle, James Martindale, Andrew S. Parsons, Karen M. Warburton.
Project administration: Suzanne M. Boyle.
Resources: Rachel Hilburg, Stephen M. Sozio.
Software: Rachel Hilburg.
Supervision: Stephen M. Sozio, Karen M. Warburton.
Visualization: Rachel Hilburg.
Writing – original draft: Suzanne M. Boyle.
Writing – review & editing: Lili Chan, James Martindale, Andrew S. Parsons, Stephen M. Sozio, Lauren D. Stern, Karen M. Warburton.

Data Sharing Statement

All data are included in the manuscript and/or supporting information.

Supplemental Material

This article contains the following supplemental material online at http://links.lww.com/CJN/B815.

Supplemental Appendix.

Supplemental Figure 1. The clinical reasoning pathway.

References

1. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med.* 2008;121(5 suppl l):2–23. doi:10.1016/j.amjmed.2008.01.001
2. Graber ML, Wachter RM, Cassel CK. Bringing diagnosis into the quality and safety equations. *JAMA.* 2012;308(12):1211–1212. doi:10.1001/2012.jama.11913
3. Croskerry P. From mindless to mindful practice–cognitive bias and clinical decision making. *N Engl J Med.* 2013;368(26):2445–2448. doi:10.1056/NEJMp1303712
4. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med.* 2005;165(13):1493–1499. doi:10.1001/archinte.165.13.1493
5. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med.* 2006;355(21):2217–2225. doi:10.1056/NEJMra054782
6. 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. doi:10.1097/ACM.0000000000002618
7. Croskerry P. Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Adv Health Sci Educ Theor Pract.* 2009;14(suppl 1):27–35. doi:10.1007/s10459-009-9182-2
8. Charlin B, Tardif J, Boshuizen HP. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med.* 2000;75(2):182–190. doi:10.1097/00001888-200002000-00020
9. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 1: origins of bias and theory of debiasing. *BMJ Qual Saf.* 2013;22(suppl 2):ii58–ii64. doi:10.1136/bmjqs-2012-001712
10. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 2: impediments to and strategies for change. *BMJ Qual Saf.* 2013;22(suppl 2):ii65–ii72. doi:10.1136/bmjqs-2012-001713
11. Audetat MC, Laurin S, Dory V, Charlin B, Nendaz MR. Diagnosis and management of clinical reasoning difficulties: Part I. Clinical reasoning supervision and educational diagnosis. *Med Teach.* 2017;39(8):792–796. doi:10.1080/0142159X.2017.1331033
12. Audetat MC, Laurin S, Sanche G, et al. Clinical reasoning difficulties: a taxonomy for clinical teachers. *Med Teach.* 2013;35(3):e984–e989. doi:10.3109/0142159X.2012.733041
13. Hawks MK, Maciuba JM, Merkebu J, et al. Clinical reasoning curricula in preclinical undergraduate medical education: a scoping Review. *Acad Med.* 2023;98(8):958–965. doi:10.1097/ACM.0000000000005197
14. Kassirer JP. Teaching clinical reasoning: case-based and coached. *Acad Med.* 2010;85(7):1118–1124. doi:10.1097/acm.0b013e3181d5dd0d
15. Guerrasio J, Garrity MJ, Aagaard EM. Learner deficits and academic outcomes of medical students, residents, fellows, and attending physicians referred to a remediation program, 2006-2012. *Acad Med.* 2014;89(2):352–358. doi:10.1097/ACM.0000000000000122
16. Warburton KM, Shahane AA. Mental health conditions among struggling GME learners: results from a single center remediation program. *J Grad Med Educ.* 2020;12(6):773–777. doi:10.4300/JGME-D-20-00007.1
17. Guerrasio J, Aagaard EM. Methods and outcomes for the remediation of clinical reasoning. *J Gen Intern Med.* 2014;29(12):1607–1614. doi:10.1007/s11606-014-2955-1
18. Tonelli M, Wiebe N, Manns BJ, et al. Comparison of the complexity of patients seen by different medical subspecialists in a universal health care system. *JAMA Netw Open.* 2018;1(7):e184852. doi:10.1001/jamanetworkopen.2018.4852
19. Boyle SM, Subedi K, Pivert KA, et al. Nephrology fellows' and program directors' perceptions of hospital rounds in the United States. *Clin J Am Soc Nephrol.* 2020;15(4):474–483. doi:10.2215/CJN.10190819
20. Warburton KM, Mahan JD. Coaching nephrology trainees who struggle with clinical performance. *Clin J Am Soc Nephrol.* 2018;13(1):172–174. doi:10.2215/CJN.07270717
21. Rumack CM, Guerrasio J, Christensen A, Aagaard EM. Academic remediation: why early identification and intervention matters. *Acad Radiol.* 2017;24(6):730–733. doi:10.1016/j.acra.2016.12.022
22. 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):163–173. doi:10.1080/10401334.2015.1011654
23. 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(12):746–753. doi:10.12788/jhm.3233
24. Williams RG, Klamen DL. Examining the diagnostic justification abilities of fourth-year medical students. *Acad Med.* 2012;87(8):1008–1014. doi:10.1097/ACM.0b013e31825cfcff
25. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15(2):155–163. doi:10.1016/j.jcm.2016.02.012
26. Downing SM, Tekian A, Yudkowsky R. Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. *Teach Learn Med.* 2006;18(1):50–57. doi:10.1207/s15328015tlm1801_11
27. Parsons AS, Clancy CB, Rencic JJ, Warburton KM. Targeted strategies to remediate diagnostic reasoning deficits. *Acad Med.* 2022;97(4):616. doi:10.1097/ACM.0000000000004244
28. Brotherton SE, Etzel SI. Graduate medical education, 2021-2022. *JAMA.* 2022;328(11):1123–1146. doi:10.1001/jama.2022.13081
29. Pivert KBA, Chan L, Halbach SM, et al. *On Deck: results from the 2022 ASN Nephrology Fellow Survey*; 2022. https://data.asn-online.org/posts/2022_fellow_survey/
30. Warburton KM, Goren E, Dine CJ. Comprehensive assessment of struggling learners referred to a graduate medical education remediation program. *J Grad Med Educ.* 2017;9(6):763–767. doi:10.4300/JGME-D-17-00175.1
31. Parsons AS, Warburton KM. In: *Remediation in Medical Education: a Mid-Course Correction* Kalet A, Chou C, eds. In Press
32. Simpkin AL, Vyas JM, Armstrong KA. Diagnostic reasoning: an endangered competency in internal medicine training. *Ann Intern Med.* 2017;167(7):507–508. doi:10.7326/M17-0163
33. Abdulnour RE, Parsons AS, Muller D, Drazen J, Rubin EJ, Rencic J. Deliberate practice at the virtual bedside to improve clinical reasoning. *N Engl J Med.* 2022;386(20):1946–1947. doi:10.1056/NEJMe2204540

34. Jurich D, Duhigg LM, Plumb TJ, et al. Performance on the nephrology in-training examination and ABIM nephrology certification examination outcomes. *Clin J Am Soc Nephrol.* 2018; 13(5):710–717. doi:10.2215/CJN.05580517
35. Berns JS, Weng W, Jaar BG, Lipner RS, Brossman BG, McDonald FS. Analysis of performance trends from 2010-2019 on the American board of internal medicine nephrology certifying exam. *J Am Soc Nephrol.* 2021;32(11):2714–2723. doi:10.1681/ASN.2021020160
36. Young M, Thomas A, Lubarsky S, et al. Drawing boundaries: the difficulty in defining clinical reasoning. *Acad Med.* 2018;93(7):990–995. doi:10.1097/ACM.0000000000002142
37. Cook DA, Sherbino J, Durning SJ. Management reasoning: beyond the diagnosis. *JAMA.* 2018;319(22):2267–2268. doi:10.1001/jama.2018.4385
38. Yao DC, Wright SM. National survey of internal medicine residency program directors regarding problem residents. *JAMA.* 2000;284(9):1099–1104. doi:10.1001/jama.284.9.1099
39. Schwind CJ, Williams RG, Boehler ML, Dunnington GL. Do individual attendings' post-rotation performance ratings detect residents' clinical performance deficiencies?. *Acad Med.* 2004; 79(5):453–457. doi:10.1097/00001888-200405000-00016

Received: May 30, 2023 **Accepted:** October 2, 2023
Published Online Ahead of Print: October 18, 2023

AFFILIATIONS

[1]Section of Nephrology, Hypertension, and Kidney Transplantation, Lewis Katz School of Medicine at Temple University, Philadelphia, Pennsylvania
[2]Office of Medical Education, University of Virginia School of Medicine, Charlottesville, Virginia
[3]Division of General, Geriatric, Palliative, and Hospital Medicine, University of Virginia School of Medicine, Charlottesville, Virginia
[4]Division of Nephrology, Department of Medicine, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University School of Medicine, Baltimore, Maryland
[5]Renal, Electrolyte, and Hypertension Division, Perelman School of Medicine, The University of Pennsylvania, Philadelphia, Pennsylvania
[6]Division of Nephrology, Stanford University School of Medicine, Palo Alto, California
[7]Barbara T. Murphy Division of Nephrology, Mt. Sinai School of Medicine, New York, New York
[8]Renal Section, Boston University Chobanian and Avedisian School of Medicine, Boston, Massachusetts
[9]Division of Nephrology, University of Virginia School of Medicine, Charlottsville, Virginia