Supplemental Digital Appendix 2
Data Extraction Form for a 2016 Scoping Review of Clinical Reasoning Assessment Methods

1. Should this article be included in review?
   - Yes, include
   - Is this article a review article on the assessment method?
     (Note: If you answer yes, use for discovering additional articles and interpretation, but do not extract)
     - No, exclude
     - Flag for third party review due to questions

2. Does the citation explicitly or implicitly use one or more conceptual frameworks?
   Note: you may select both explicit and implicit (e.g., dual processing theory explicitly discussed, cognitive load implicitly discussed)

   - Yes explicitly
   - Yes implicitly
   - No
   - Uncertain-explain:
   - Not applicable

3. If you selected yes above, please select all the conceptual frameworks described either explicitly or implicitly in the article.

   - Cognitive load – comments:
   - Dual processing theory – comments:
   - Expert performance theory (e.g., deliberate practice) – comments:
   - Motivation and emotion (e.g., control-value theory) – comments:
   - Probability theory (Bayesian reasoning – e.g., pre-test probability estimation, likelihood ratios, etc) – comments:
   - Script theory (e.g., illness scripts) – comments:
   - Self-regulation – comments:
   - Situativity theory – comments:
   - Other:

4. What assessment method(s) was used? (Select all that apply)

   - Biologic (cortisol levels, pupil dilation, functional MRI) – comments if needed:
   - Chart stimulated recall – comments if needed:
   - Clinical reasoning problem (exact phrase must be used in article) – comments if needed:
   - Comprehensive integrative puzzle – comments if needed:
   - Concept map – comments if needed:
   - Direct observation (Mini-CEX, clinical examination exercise) – comments if needed:
   - Extended matching questions – comments if needed:
   - Free text responses/short / long essay – comments if needed:
   - Global assessment – comments if needed:
Key features testing – comments if needed:
Multiple choice questions – comments if needed:
Objective structural clinical examination (OSCE) – comments if needed:
Oral case presentation – comments if needed:
Oral examination – comments if needed:
Patient management problem – comments if needed:
Script concordance testing – comments if needed:
Self-regulated learning/microanalysis techniques (SRL-MAT) – comments if needed:
Stimulation with technology (simulation) – comments if needed:
Think aloud protocol – comments if needed:
Written notes (charted documents e.g. admission notes, OR post-encounter form) – comments if needed:
Other – list method and explain it:

5. Please select the stimulus format. **Select all that apply.**

- Real patient
- Standardized patient
- Virtual patient (e.g. computer-based avatar) – describe if necessary:
- Written case vignette – describe if necessary:

6. Please choose response format. **Select all that apply.**

- Selected response (i.e. answers provided)

  What selected response format was used? **Select all that apply**
  - Single best answer from a short list of <6 options
  - Single best answer from a short list of >5 options
  - Greater than 1 correct answer – please describe:
  - Other – please describe:

- Constructed response/free text

  What was the format of the constructed response/free text?
  - Verbal response

    Please select the format of the verbal response. **Select all that apply.**
    - Examiner/teacher-driven
    - Learner-driven

- Written response

  What was the format of the written response?
  - Clinical documentation – describe:
  - Diagram/graphic depiction (e.g. concept map) – describe:
  - Long answer/essay (>3 sentences) – describe:
  - Post-encounter form (e.g., write-up of differential diagnosis, working diagnosis,
7. What scoring activity was used specifically for clinical reasoning? **Select all that apply.**
   - Fixed answer (e.g., MCQ, EMQ)
   - Global rating scale only
   - Global rating scale followed by itemized rating scale only
   - Itemized (analytic) rating scale only (e.g., Likert scale)
   - Itemized (analytic) rating scale followed by global rating scale
   - Dichotomous items (e.g., performed yes/no checklist)
   - Pure narrative (e.g., some think alouds) – describe:
   - Other – describe:
   - Uncertain – explain:
   - Not applicable

Please provide any additional details regarding scoring activity that are important

8. What range of tasks were assessed? **Select all that apply?**
   - Diagnosis

   What diagnostic tasks were assessed?
   - Data collection – describe if necessary
   - Data interpretation – describe if necessary
   - Diagnosis justification – describe if necessary
   - Diagnosis selection – describe if necessary
   - Hypothesis generation (e.g., differential diagnosis construction) – describe if necessary
   - Hypothesis refinement – describe if necessary
   - Pre-test probability estimation/Ranking differential diagnostic possibilities
   - Problem representation – describe if necessary
   - Other – please describe
   - Uncertain – explain:

   □ Treatment

   What treatment tasks were assessed?
   - Best therapeutic option selection
   - Therapeutic option prioritization (e.g., ranking)
Threshold to treat determination (e.g., at what probability of disease would the benefit of treating a patient outweigh the risk of further testing or treating someone with the disease)

Values and priorities identification and quantification (e.g., Quality Adjusted Life Year considerations)

Other – describe:
Uncertain – describe:
No applicable – explain:

9. What were the stakes of the assessment?
- High stakes (e.g., licensing examination, graduation requirement)
- Medium stakes (e.g., course requirement)
- Low stakes (e.g., no impact on pass/fail status)
- Uncertain – explain:
- Not applicable

10. Who were the participants studied?
- Medicine

What was the level(s) of training of participants studied? **Select all that apply.**
- Pre-medical
- Undergraduate, pre-clerkship
- Undergraduate, clerkship and beyond
- Postgraduate, resident
- Postgraduate, fellow
- Practicing physician

- Nursing

What are the level(s) of training of participants studied? **Select all that apply.**
- Undergraduate nursing degree trainees
- Advanced nursing degree trainees
- Practicing nurses
- Other:

- Dentistry – describe if necessary:
- Nutrition – describe if necessary:
- Occupational Therapy – describe if necessary:
- Osteopathic medicine – describe if necessary:
- Physical therapy – describe if necessary:
- Physician assistants – describe if necessary:
- Speech/language pathology – describe if necessary:
- Other – describe:

11. Was the feasibility of designing, administering, and/or scoring the assessment method described in the article?
- Yes
- No
- Uncertain
Please select which aspects of feasibility were discussed in the article. Select all that apply.

- Design – describe key elements discussed (e.g., number of designers, hours spent on design, piloting, etc.) and challenges faced if any
- Administration – describe key elements discussed: (e.g., number of administrators, hours spent on administration, piloting, etc.) and challenges faced if any
- Scoring – describe key elements discussed: (e.g., number of scorers, hours spent on scoring) and challenges faced if any
- Other:

12. Was reliability calculated?

- Yes
- No
- Uncertain
- Not applicable

How was reliability calculated?

- Consistency over items (e.g., Cronbach’s alpha)
- Consistency over judges (e.g., inter-rater reliability [kappa], intra-class correlation coefficient [ICC])
- Consistency over time (e.g., intra-rater)
- Other – describe:
- Uncertain

13. Please discuss any other important aspects of reliability.

14. Was validity evaluated?

- Yes explicitly
- Yes implicitly
- No
- Uncertain: explain
- Not applicable

Select all elements of validity assessed (as per Messick’s validity framework)

- Content (i.e., relationship between content of assessment method and construct of interest)
- Response process (i.e., analyses of responses of individual respondents or observers; Also includes instrument security, scoring, and reporting of results)
- Internal structure (i.e., the degree to which individual items within the instrument fit the underlying constructs, typically measured by reliability or factor analysis)
- Relationship to other variables (i.e., the relationship between scores and other variable relevant to the construct being measured)
- Consequences (e.g., assessments are expected to have intended and unintended effects; are these reported?

Additional comments regarding validity:

15. Please describe any other themes regarding clinical reasoning assessment that emerged from the article.
16. Please list important findings (i.e., take-home points) of the article.