Validity, Reliability, Feasibility and Acceptability of Using the Consultation Letter Rating Scale to Assess Written Communication Competencies Among Geriatric Medicine Postgraduate Trainees

Victoria YY Xu
PGY-3 Internal Medicine
University of Toronto

Supervisor: Dr. Camilla Wong
Presenter Disclosure

- Grants/Research Support: Postgraduate Innovation Fund (PGIF), Department of Medicine, University of Toronto
- No conflicts of interest identified
Background
Consultation Letters

► Writing effective consultation letters is an important competency

► A well-written consultation letter:
  ▪ Informs the referring physician about diagnosis and management
  ▪ Can be a valuable source of continuing medical education

► A poorly written consultation letter may result in:
  ▪ Delayed diagnoses
  ▪ Redundant investigations
  ▪ Erosion of inter-physician relationships

► Most physicians have not received feedback on their letters

---

Assessment

“A strategy to support and inform further learning”

6 items rated on a 5-point Likert scale:

- **3 on letter content:** history, physical examination and impression & plan
- **2 on letter style:** clarity & brevity and organization of letter
- **Overall rating**

Free-text area:

- **3 areas of strength**
- **3 areas for improvement**
- **Comments**
Study Objectives

- Validity
- Consultation Letter Rating Scale
- Geriatric Medicine training program
- Acceptability
- Reliability
- Feasibility
Methods
Study Setting and Participants

- Multisite initiative across the Division of Geriatric Medicine at 4 academic hospitals

- Eligible participants: 11 postgraduate trainees in Geriatric Medicine at the University of Toronto

- Exclusion criteria: fewer than 5 letters available at the time of recruitment

- Ethics approval was obtained from each institutional Research Ethics Board

- Informed consent was obtained from all study participants
Data Collection

- Each trainee provided 5 letters for new consults seen in the 2017-2018 academic year
  - Either outpatient or inpatient setting
  - Avoided letters for which the staff was one of the raters

- All letters were de-identified and coded

- The 6 raters participated in a workshop on how to use the tool, including an online module, “Creating Effective Consultation Letters”
## Data collection

<table>
<thead>
<tr>
<th>Raters</th>
<th>PGY4/5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>E</td>
<td></td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

**A** B C D E F
# Outcome Measures and Data Analysis

<table>
<thead>
<tr>
<th>1. Validity</th>
<th>Raters completed a survey on whether or not there were any missing components and/or components that should be removed. Responses were reviewed independently by 2 authors for thematic content.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Reliability</td>
<td>• Inter-rater reliability was calculated for each item and for overall scores, using the Kappa coefficient and its corresponding 95% confidence interval.</td>
</tr>
</tbody>
</table>
| 3. Feasibility | • The mean time to complete the tool for each letter with standard deviation was calculated.  
• Proportion of scales with incomplete data was determined. |
| 4. Acceptability | • The collated feedback was provided to the participants, who then completed a survey on the tool’s usefulness.  
• Responses were reviewed independently by 2 authors for thematic content. |
Results
Feasibility

- A very small portion (4%, n = 12) of data was incomplete.
- An average of 4.82 minutes (SD = 3.17) was used to complete the tool:
  - Minimum: 0.83 minutes
  - Maximum: 20 minutes
Reliability

- **Multi-rater kappa** was calculated for each item and overall scores
- Both unweighted and weighted kappa were calculated
  - **Weighted kappa**: accounts for the degree to which the raters agree or disagree
    - E.g. Ratings of 5 and 1 are considered to represent a larger degree of disagreement than ratings of 2 and 1
Reliability

- Kappa values range from -1 to 1, where 1 is considered perfect agreement, while 0 is what would be expected by chance.
- In our interpretation of the kappa results, we used the following commonly used categories as guidance:
  - <0: poor agreement
  - 0-0.2: negligible to slight agreement
  - 0.21-0.4: fair agreement
  - 0.41-0.6: moderate agreement
  - 0.61-0.8: substantial agreement
  - 0.81-1.0: strong agreement

### Table 1: Pairwise agreement (between pairs of raters) for overall scores, with unweighted kappa shaded in red and weighted kappa shaded in pink

<table>
<thead>
<tr>
<th>Rater 1</th>
<th>Rater 2</th>
<th>Rater 3</th>
<th>Rater 4</th>
<th>Rater 5</th>
<th>Rater 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.84</td>
<td>0.78</td>
<td>0.73</td>
<td>0.73</td>
<td>0.76</td>
<td>0.80</td>
</tr>
<tr>
<td>0.87</td>
<td>0.86</td>
<td>0.73</td>
<td>0.75</td>
<td>0.68</td>
<td>0.68</td>
</tr>
<tr>
<td>0.77</td>
<td>0.77</td>
<td>0.98</td>
<td>0.38</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>0.86</td>
<td>0.77</td>
<td>0.87</td>
<td>0.73</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>0.89</td>
<td>0.85</td>
<td>0.87</td>
<td>0.81</td>
<td>0.87</td>
<td></td>
</tr>
</tbody>
</table>

Mean weighted kappa = 0.83 (95% CI: [0.75, 0.88])

Mean unweighted kappa = 0.65 (95% CI: [0.55, 0.75])
Reliability

<table>
<thead>
<tr>
<th>Categories</th>
<th>Multiple Rater Kappa [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Un-weighted</td>
</tr>
<tr>
<td>Overall</td>
<td>0.65 [0.55, 0.75]</td>
</tr>
<tr>
<td>Content Rating</td>
<td></td>
</tr>
<tr>
<td>History</td>
<td>0.23 [0.13, 0.35]</td>
</tr>
<tr>
<td>Physical Examination</td>
<td>0.25 [0.18, 0.33]</td>
</tr>
<tr>
<td>Impression and Plan</td>
<td>0.29 [0.20, 0.40]</td>
</tr>
<tr>
<td>Style Rating</td>
<td></td>
</tr>
<tr>
<td>Clarity and Brevity</td>
<td>0.21 [0.14, 0.32]</td>
</tr>
<tr>
<td>Organization of Letter</td>
<td>0.27 [0.18, 0.38]</td>
</tr>
</tbody>
</table>

Table 2: Weighted and unweighted kappa values for individual Likert scale items
Validity

What components were missing?

1) More room for comments
2) Evaluation of professionalism (spelling, grammar, tone)
3) Wording should reflect components of the CBD model/level of training

What components should be removed?

1) None
2) The 5-point Likert scale is not useful for feedback or differentiation, as it results in most ratings being 4 or 5
Acceptability

- All participants found the feedback from the tool useful

- The numeric Likert scale was less useful than the written comments

- Appreciated the opportunity for dedicated feedback on consult letters and written communication skills

- Interested in seeing examples of exemplary consult letters
Discussion
Strengths

- Multi-site
- Analysis of 300 assessments with minimal missing data
- Qualitative analysis: both raters’ and participants’ perspectives
Limitations

- Selection bias: trainees chose the letters they submitted
- Does not represent real-time use in the clinical setting
- Time recorded did not include time taken to read the letter and may underestimate faculty time commitment
- Did not account for differences in consult letter generation
  - Dictated vs. typed note
  - Opportunity to edit note
Our results support the use of the Consultation Letter Rating Scale to assess the quality of consult letters in postgraduate Geriatric Medicine training

- Inter-rater reliability was high for overall scores but lower for the other items; consistent with study on original tool
- Both raters and participants found the comments more useful than the numerical ratings

This tool is most useful for facilitating feedback on consult letters to help trainees improve written communication skills over time
Impact

- Based on our results, the U of T geriatric medicine program is recommending regular use of this tool in longitudinal clinics.
- A session on writing effective consult letters has been incorporated into the program’s academic half day.
Acknowledgements

Study team:
- Maia von Maltzahn, Terumi Izukawa, Mireille Norris, Vicky Chau and Barbara Liu (Raters & Co-authors)
- Jemila Hamid (Statistician)
- Camilla Wong (Research Supervisor & Principal Investigator)

Study Participants

Postgraduate Innovation Fund, Department of Medicine, University of Toronto
References


Thank you!

Contact info:
vxu@qmed.ca
Extra slides
Creating an Effective Consultation Letter: Essential Content

Subject Matter Expert:
Sue Dojeiji, MD, MEd, FRCPC

Attending, Department of Physical Medicine and Rehabilitation, Bruyère Continuing Care and The Ottawa Hospital Rehabilitation Centre
Associate Professor, Division of Physical Medicine and Rehabilitation, University of Ottawa

https://www.bruyere.org/en/Consultationletters
Intra-Rater Kappa

- Intra-rater kappa was used to evaluate internal consistency

- Overall weighted kappa = 0.84 (95%CI: [0.77, 0.89])
  - Overall unweighted kappa = 0.65 (95%CI: [0.55, 0.74])

- High intra-rater agreement (consistency)