

Supplemental Digital Appendix 1

Semistructured Interview Guide From a Study of the Hazards in the Interpretation of Written Assessment Comments 2017^a

Hi (Name):

Our research study is investigating the potential utility of analyzing the written comments on internal medicine residents' ITERs. We're testing various methods for analyzing the nature of written comments. We'll be going through two separate exercises. Do you have any questions before we begin?

Before we do the exercises, I'd just like to record some demographic data about you and where you work. This information will be kept strictly confidential and will only be used to understand patterns of responses. (Asks about institution, rank, previous degrees.)

1. So, for this first exercise, I'm going to ask you to fill out a survey. This survey will consist of a list of words and phrases that are commonly seen in residents' ITERs. I'd like you to think about each word/phrase and consider: If you read this on a resident's ITER would it raise a "red flag" about the resident's performance? The options will be a standard Likert scale (strongly disagree – disagree – slightly disagree – neutral – slightly agree – agree – strongly agree).

After completing the survey... Now we just have a few follow-up questions before we move on to the second exercise.

- Do you have any additional thoughts on the words?
 - Are there any "red flag" words or phrases that you commonly encounter that are missing from the list?
 - What do you think about using "red flag" words in ITERs?
Probe if necessary: Do you find them challenging to interpret? Do you ever use them when you really mean something else (e.g., something stronger or more critical)?
 - Do you think residents see these as red flags as well? (Probe: How do you think residents interpret this kind of language—you can use the word "lukewarm" if it helps.)
 - How did you learn about "red flag" words?
2. For our second exercise, I'm going to give you a list of words and phrases that have been judged by a group of researchers as reflective of "feedback language." That is, these are words or phrases commonly seen on ITERs that indicate that the writer is attempting to give feedback to the resident. Take your time to go over the words. Can you tell me your initial thoughts on these feedback words?

Note: If they ask what we mean by feedback, try asking them what they think it means first. We could then offer the following: While virtually any language on

ITERS can be considered loosely as feedback in the broadest sense, what we are really interested in are words/phrases that suggest constructive feedback intended to influence future performance or behavior.

- In your opinion, do these words/phrases indicate “feedback”?
- Are there any you feel do not represent feedback? If yes, what would you say these words represent?
- Are there any feedback words or phrases missing from the list?
- Have you ever been in a position to “use” written comments, e.g., as a program director, selection committee member, for remediation, etc.?

Abbreviation: ITER indicates in-training evaluation report.

^aParticipants were faculty from the University of Toronto, Toronto, Ontario, Canada; the University of British Columbia, Vancouver, British Columbia, Canada; Western University, London, Ontario, Canada; and the University of Pennsylvania, Philadelphia, Pennsylvania.