



## PGME Guidelines: Reporting of Assessment Results

### Preamble:

Reporting and communication of assessment results needs to be done in a manner that is consistent with other documents including PGME guidelines and standards, University of Toronto policies, and accreditation standards.

The PGME office and residency programs will develop and update report templates in consultation of the stakeholder groups. This will allow the groups to determine whether the report provides the appropriate information in a format that is useful for their particular needs.

PGME and programs will be cautious to develop reports that might overwhelm the utility of the report to the user (e.g. overly large volumes of assessment data; misleading graphical presentations)

Programs will determine what support or resources are needed to help each resident understand their assessment reports and provide guidance for improvement and progress.

Programs and the PGME office will support faculty and learner development in the interpretation of assessment data reports for the program's different user groups.

### Minimum Standards

- 1) Access to individual assessment data reports will be consistent with other PGME and University policies and be restricted to those having a specific responsibility for resident oversight or promotion decisions (e.g. resident, the Program Director, remediation coordinators, any individual or Committee responsible for making Promotion decisions or overseeing the resident's education such as Site Directors, Supervising Faculty).
- 2) Assessment data synthesis must be made available, reviewed by program, and discussed with residents a minimum of twice a year.
- 3) Programs will determine the specifics of tools/documents to be included in assessment data synthesis reports for each stakeholder group (e.g. residents, Competence Committees).

- 4) Tools/documents and the assessment data reports should reflect the resident's educational experiences, and be consistent with the established residency program assessment plan. These may include: Fieldnotes, Entrustable Professional Activities (EPAs), ITERs, ITARs, Multi-source Feedbacks (MSFs,) oral exams, OSCEs, in-training exams, or other relevant assessments.
- 5) The PGME Office will ensure programs have access to basic reports from the PGME systems (e.g. Elentra, POWER) to meet the need of key stakeholders.
- 6) Assessment data reports must include:
  - a) Identifying data.
  - b) Resident pattern of performance over time.
  - c) Information on patient safety concerns (e.g. critical incident reports), if available.
  - d) Areas of strength and gaps.
  - e) Sufficient information to enable interpretation of the reports.
- 7) Individual resident results should include appropriate normative or comparative data and be mapped to performance standards.
- 8) Results should, where possible, be mapped to performance standards (e.g. specialty specific EPAs or milestones).

## Overview of Reporting on Assessment Data

### Purpose(s) of assessment data

Who receives assessment data ↓	Purpose(s) of assessment data ↓
<b>Assessee</b>  E.g. Residents	<ul style="list-style-type: none"> <li>- Document assessments completed</li> <li>- Monitor own progress of assessments</li> <li>- Document own progress across residency program</li> <li>- Comparison against program expectations</li> <li>- Comparison against peer/cohort group (i.e. current or historical)</li> </ul>
<b>Assessor</b>  E.g. Faculty, Co-residents	<ul style="list-style-type: none"> <li>- Document assessments completed</li> <li>- Monitor own progress of assessments completed, as compared to other assessors</li> <li>- Document own progress over time in terms of number of assessments completed</li> <li>- Comparison against departmental or divisional expectations</li> <li>- Comparison against peer/cohort group (i.e. current or historical)</li> </ul>
<b>Program Director (PD)</b>	For all residents, assessors, and by site and rotation: <ul style="list-style-type: none"> <li>- Monitor progress of individual residents</li> <li>- Identify residents who are struggling</li> <li>- Identify residents who need more challenges</li> <li>- Monitor progress of individuals against groups, program expectations</li> <li>- Monitor reasonableness of program expectations of residents, and assessors</li> <li>- Monitor assessors who are struggling or exceeding expectations</li> </ul>
<b>Site Director And/or Rotation Director</b>	For specific site, rotation and those residents and assessors: <ul style="list-style-type: none"> <li>- Monitor progress of individual residents</li> <li>- Identify residents who are struggling</li> <li>- Identify residents who need more challenges</li> <li>- Monitor progress of individuals against groups, program expectations</li> <li>- Monitor reasonableness of program expectations of residents, assessors</li> <li>- Monitor assessors who are struggling or exceeding expectations</li> </ul>
<b>Resident Advisor</b> (may be separate person or PD)	<ul style="list-style-type: none"> <li>- Help individual resident understand their assessment reports and progress</li> <li>- Provide guidance and access to resources for improvement and progress.</li> <li>- Identify appropriate educational opportunities to either address gaps or enhance strengths.</li> </ul>
<b>Competence Committee Chair &amp; members</b>	<ul style="list-style-type: none"> <li>- Make individual resident decisions on progress and promotion</li> <li>- Identify residents in need of additional or different experience (e.g. remedial, enrichment)</li> <li>- Monitor for trends across assessment tools, sites, rotations, assessors</li> </ul>
<b>Residency Program Committees</b>	<ul style="list-style-type: none"> <li>- Monitor for trends across residents, assessment tools, sites, rotations, assessors</li> </ul>
<b>Remediation Coordinators</b>	<ul style="list-style-type: none"> <li>- Assist program in developing appropriate remediation plans</li> <li>- Provide coaches/tutors with information to assist them in working with the resident</li> </ul>

**Samples of assessment data in reports**

<b>Who receives assessment data ↓</b>	<b>Samples of needed assessment data in reports ↓</b>
<b>Assessee</b>  E.g. Residents	<ul style="list-style-type: none"> <li>- Assessment scores including narratives for assessments completed about them</li> <li>- Completion of assessments compared to program expectations (e.g. target scores or frequencies; target variety of assessments)</li> <li>- Completions compared to peers (e.g. scores, frequency, variety)</li> </ul>
<b>Assessor</b>  E.g. Faculty, Co-residents	<ul style="list-style-type: none"> <li>- Assessment scores including narratives for assessments they completed</li> <li>- Completion rates and scores compared to other assessors.</li> <li>- Comparisons to departmental or divisional expectations</li> </ul>
<b>Program Director (PD)</b>	<ul style="list-style-type: none"> <li>- Assessment scores including narratives for assessments completed about each resident</li> <li>- Assessment reports that compare and contrast residents for each stage/year</li> <li>- Quantity and variety of assessments for each resident, compared to program expectations (e.g. target scores or frequencies; target variety of assessments, narratives)</li> </ul> Quantity and variety of assessments for each resident as compared to peers (e.g. scores, frequency, variety, narratives)
<b>Site Director And/or Rotation Director</b>	<ul style="list-style-type: none"> <li>- Assessment scores including narratives for assessments completed about each resident</li> <li>- Assessment reports that compare and contrast individual residents to their peers at that stage/year</li> <li>- Completion of required assessments for each resident for that rotation/educational experience</li> <li>- Completion rate of assessment reports by individual faculty</li> </ul>
<b>Resident Advisor</b> (may be separate person or PD)	<ul style="list-style-type: none"> <li>- Assessment scores including narratives for assessments completed about each resident they are advising</li> <li>- Assessment reports that compare and contrast each resident they are advising for each stage/year</li> <li>- Quantity and variety of assessments for each resident they are advising, compared to program expectations (e.g. target scores or frequencies; target variety of assessments, narratives)</li> <li>- Quantity and variety of assessments for each resident as compared to peers (e.g. scores, frequency, variety, narratives)</li> </ul>
<b>Competence Committee Chair &amp; members</b>	<ul style="list-style-type: none"> <li>- Assessment scores including narratives for assessments completed about each resident</li> <li>- Assessment reports that compare and contrast residents for each stage/year</li> <li>- Quantity and variety of assessments for each resident, compared to program expectations (e.g. target scores or frequencies; target variety of assessments, narratives)</li> <li>- Quantity and variety of assessments for each resident as compared to peers (e.g. scores, frequency, variety, narratives)</li> <li>- Completion rate of required assessments for each stage/year of education</li> </ul>
<b>Residency Program Committees</b>	<ul style="list-style-type: none"> <li>- Assessment scores including narratives for assessments completed about each resident</li> <li>- Assessment reports that compare and contrast residents for each stage/year</li> <li>- Quantity and variety of assessments for each resident, compared to program expectations (e.g. target scores or frequencies; target variety of assessments, narratives)</li> <li>- Quantity and variety of assessments for each resident as compared to peers (e.g. scores, frequency, variety, narratives)</li> <li>- Completion rate of assessment reports by individual faculty, as compared to peers</li> <li>- Quality of assessment reports by individual faculty</li> </ul>

<b>Who receives assessment data ↓</b>	<b>Samples of needed assessment data in reports ↓</b>
<b>Remediation Coordinators</b>	<ul style="list-style-type: none"><li>- Assessment scores, including narratives, for assessments completed about resident</li><li>- Assessment reports and that compare and contrast the resident with other residents at that level</li><li>- Quantity and variety of assessments for each resident, compared to Remediation Plan expectations (e.g. target scores or frequencies; target variety of assessments)</li><li>- Quantity and variety of assessments for resident as compared to peers (e.g. scores, frequency, variety)</li><li>- Areas of strength</li><li>- Gaps noted</li></ul>

## APPENDIX: Backgrounder: Reporting of Assessment Data

(NOTE: not part of Reporting Guideline)

### 1. Introduction

The purpose of this document is to provide background information to the Best Practices in Evaluation and Assessment (BPEA) Working Group so that they can make recommendations to the PGMEAC on the minimum standards for reporting of assessment data. **Error! Reference source not found.**

### 2. Background

*Each residency program will have a Program of Assessment*

**A program of assessment is a system which includes multiple types of assessments at multiple points in time.**

This is specifically described in the 2018 Accreditation Standards [1]:

- **Standard 3:** 3.4.1 The residency program has a planned, defined and implemented system of assessment.

There is a rich literature on programs of assessment [2-5]. The BPEA project included a background paper that reviewed the literature and issues around a program of assessment. (See BPEA Theme Paper 4 [6]).

*Programs of assessment will have multiple forms of data*

The 2018 Accreditation standards stipulate:

- **Standard 3:** 3.4.1.4: The system of assessment includes identification and use of appropriate assessment tools tailored to the residency program's educational experiences, with an emphasis on direct observation where appropriate.
- **Standard 3:** 3.4.1.6: The system of assessment is based on multiple assessments of residents' competencies during the various educational experiences and over time, by multiple assessors, in multiple contexts.

The *Standards for Educational and Psychological Testing* (SEPT) [7, page 198], supports the concept of a program of assessment. Standard 12.10 states: "In educational settings, a decision or characterization that will have a major impact on a student should take into consideration not just scores from a single test but other relevant information". This standard discusses the use of multiple measures or data sources to enhance the decision making.

Additionally, [The Guidelines for the Assessment of Postgraduate Residents](#) [8], Section 4, goes to some length in describing the expectations for a Program Assessment Plan. It includes that there will be “Designated Assessment Tools”<sup>1</sup> [8, Section 2.3], that there must be “regular longitudinal assessment and a written Summative Assessment<sup>2</sup> against established required competencies” [8, Section 4.1], and that that the assessment plan will be at the “level of training which are derived from the national training standards” [8, Section 4.2.1].

In this section it goes on to stipulate the purpose of the Program Assessment Plan, which includes:

- 4.2.1.1 to provide a framework for the assessment of the Resident's knowledge, skills and attitudes by a Supervisor;
- 4.2.2.2 to facilitate feedback to the Resident by a Supervisor or the Program Director;
- 4.2.1.3 to serve as a record of the strengths and weaknesses of the Resident for the Program Director;
- 4.2.1.4 to enable the Program Director to assist future Supervisors in ongoing supervision;
- 4.2.1.5 to assist the Program Director in providing a progress and/or Summative Assessment of the Resident for the Royal College, the CFPC and/or the CPSO; and
- 4.2.1.6 to establish the basis for confirmation of progress, identification of needs and promotion.

The College of Family Physicians of Canada (CFPC) implemented their approach to Competency Based Medical Education, called Triple C, in 2010. The cornerstone of Triple C is the use of frequent assessment forms called Field Notes. In addition, the Family Medicine residents have a wide range of other assessments including: progress written exams every 6 months, Simulated Office Orals, Quality Improvement Projects, and In Training Evaluation Reports (ITERS).

The Royal College of Physicians and Surgeons of Canada (Royal College) is implementing new accreditation standards for specialty education, called Competence by Design (CBD). As CBD is implemented, each program has new standards that includes the assessment of Entrustable Professional Activities (EPAs). In addition to EPAs, programs will have other assessments in their program of assessment such as in-training exams, oral exams, OSCEs, Workplace Based Assessments (e.g. procedural observations), multi-source feedback forms (MSFs), ITERS or In Training Assessment Reports (ITARs), research projects, etc. Each of these assessments provides data that will be used by the various stakeholders for various purposes.

### *There will be multiple forms and formats for assessment data*

---

<sup>1</sup> 2.3 “Designated Assessment Tools” is the specified assessment tools approved by the Residency Program Committee for inclusion in the Program Assessment Plan which are appropriately tailored to the specialty, level of training, and the national training standards”

<sup>2</sup> 2.13 “Summative Assessment” refers to a formal written summary of a resident's performance against established expectations which is carried out at specified intervals within each program”.

It is anticipated that the multiple forms of data will come from a variety of systems including:

- Elenra, the main location for CBME and CBD assessments
- POWER, the main location for ITERs and ITARs
- Other online systems (e.g. case logs)
- Paper (e.g. oral exams)
- Electronic documents Pdf formats (e.g. local or national specialty exam reports, quality project reports)

The users may have to gather assessment reports from different systems to have a cohesive look at performance.

***There will be multiple users or stakeholders of the assessment data***

Depending on the size and organization of the program, there will be a wide variety of users of assessment data. Sometimes the same person (e.g. Program Director) may serve in multiple roles (e.g. Assessor, Program Director, Learner Advisory)

Examples of key stakeholders of assessment data are outlined in Table 1 below.

**Table 1**

<b>Role</b> (Note: often people have > 1 role)	<b>Purpose(s) of assessment data</b>
<b>Assessee</b>  E.g. Residents	<ul style="list-style-type: none"> <li>- Document assessments completed</li> <li>- Monitor own progress of assessments</li> <li>- Document own progress across residency program</li> <li>- Comparison against program expectations</li> <li>- Comparison against peer/cohort group (i.e. current or historical)</li> </ul>
<b>Assessor</b>  E.g. Faculty, Co-residents	<ul style="list-style-type: none"> <li>- Document assessments completed</li> <li>- Monitor own progress of assessments completed, as compared to other assessors</li> <li>- Document own progress over time in terms of number of assessments completed</li> <li>- Comparison against departmental or divisional expectations</li> <li>- Comparison against peer/cohort group (i.e. current or historical)</li> </ul>
<b>Program Director (PD)</b>	For all residents, assessors, and by site and rotation: <ul style="list-style-type: none"> <li>- Monitor progress of individual residents</li> <li>- Identify residents who are struggling</li> <li>- Identify residents who need more challenges</li> <li>- Monitor progress of individuals against groups, program expectations</li> <li>- Monitor reasonableness of program expectations of residents, and assessors</li> <li>- Monitor assessors who are struggling or exceeding expectations</li> </ul>
<b>Site Director And/or Rotation Director</b>	For specific site, rotation and those residents and assessors: <ul style="list-style-type: none"> <li>- Monitor progress of individual residents</li> <li>- Identify residents who are struggling</li> <li>- Identify residents who need more challenges</li> <li>- Monitore progress of individuals against groups, program expectations</li> <li>- Monitor reasonableness of program expectations of residents, assessors</li> <li>- Monitor assessors who are struggling or exceeding expectations</li> </ul>



<b>Role</b> (Note: often people have > 1 role)	<b>Purpose(s) of assessment data</b>
<b>Resident Advisor</b> (may be separate person or PD)	<ul style="list-style-type: none"> <li>- Help individual resident understand their assessment reports and progress</li> <li>- Provide guidance and access to resources for improvement and progress.</li> <li>- Identify appropriate educational opportunities to either address gaps or enhance strengths.</li> </ul>
<b>Competence Committee Chair &amp; members</b>	<ul style="list-style-type: none"> <li>- Make individual resident decisions on progress and promotion</li> <li>- Identify residents in need of additional or different experience (e.g. remedial, enrichment)</li> <li>- Monitor for trends across assessment tools, sites, rotations, assessors</li> </ul>
<b>Residency Program Committees</b>	<ul style="list-style-type: none"> <li>- Monitor for trends across residents, assessment tools, sites, rotations, assessors</li> </ul>
<b>Remediation Coordinators</b>	<ul style="list-style-type: none"> <li>- Assist program in developing appropriate remediation plans</li> <li>- Provide coaches/tutors with information to assist them in working with the resident</li> </ul>

Each of these stakeholders will have slightly different perspectives and may need slightly different data and/or presentation of data.

**A score report is a form of communication with the particular stakeholder** [9]. As discussed in the SEPT [7, page 193] “Score reports for educational assessments should support the interpretations and decisions of their intended audiences...Different reports may be developed and produced for different audiences, and the score report layouts may differ accordingly.”

***The data reports should meet the needs of the reader(s)***

The information presented in a score report should be understandable and match the needs of the user. For example, the resident may benefit from seeing their own results in comparison to others at their level. The Competence Committee may benefit from seeing an individual’s results as compared to a benchmark for that level.

A result on an assessment is simply that, a result. To interpret the results of the assessment some normative data is required to indicate whether the result is at level, above level, or below level. A standard from SEPT that is relevant at the program level is Standard 12.5: “Local norms should be developed when appropriate to support test users’ intended interpretations” (page 196). With the implementation of CBD this type of data will be gathered over time. Larger programs will be able to identify norms more rapidly than smaller programs, who may need a couple years to collate data from a larger number of residents to see the patterns more clearly.

Some best practices and recommendations on creating and using score reports have been created [9]. A summary of the five major areas is as follows:

- 1) The reports should be developed with consultation of the stakeholder groups. This will allow the groups to determine whether the report provides the appropriate information in a format that is useful for their particular needs.

- 2) The layout should be simple, with key results easy to find. Language should be clear and concise. Graphs are helpful.
- 3) The report should contain the information that is essential for accurate interpretation. It is helpful if the report contains actionable items that might lead to improvements in future performance. Benchmarking can provide context and assist in interpreting the results. Results should also be clearly linked to performance standards.
- 4) An annotated example of a score report is helpful to the reader in interpreting the reports. Guides may provide in-depth discussion of the various sections of the report.
- 5) Results should be communicated in a timely fashion as this will be more meaningful to the resident, as well as the other stakeholders.

The [\*Guidelines for the Assessment of Postgraduate Residents\*](#) [8, page 8] provides instruction that, at a minimum, twice a year a "...completed Summative Assessment must be submitted using all data collected with the Designated Assessment Tools" [8, Section 4.3.4].

### ***What data is included in reports***

Considering all the above, the task then becomes one of identifying what information needs to be presented to each stakeholder, and in what format, to be the most helpful in decision making.

With respect to the **Competence Committees'** use of the data, the *PGME Guidelines for Competence Committees* [10] state that these Committees will:

- 4a) "make judgements about a resident's progression through residency (e.g. competence stages, promotion from one year/level to next, identification of needed improvement or remediation, identification of needed enhancement, or enrichment, readiness for certification examinations)"
- 4b) "review assessment and performance data patterns and trends (e.g. across residents, stages, sites, rotations, or assessment tools/approaches) to identify areas of excellence and areas needing improvement for individual residents and the residency program"

The guidelines also provide a sample Terms of Reference, which recommends the material that the Committee may include in their review. The recommendations (Section 7) are:

- Resident performance for the period under consideration
- Resident pattern of performance over time
- Patient safety needs
- Service needs of rotations
- The need for different approaches to resident supervision
- Individual committee member experience regarding resident performance may be included if there is a request to clarify the available assessment documentation

### **Table 2**

<b>Role</b> (Note: often people have > 1 role)	<b>Samples of needed assessment data in reports</b>
<b>Assessee</b>  E.g. Residents	<ul style="list-style-type: none"> <li>- Assessment scores, including narratives, for assessments completed about them</li> <li>- Completion of assessments compared to program expectations (e.g. target scores or frequencies; target variety of assessments)</li> <li>- Completions compared to peers (e.g. scores, frequency, variety)</li> </ul>
<b>Assessor</b> E.g. Faculty, Co-residents	<ul style="list-style-type: none"> <li>- Assessment scores, including narratives, for assessments they completed</li> <li>- Completion rates and scores compared to other assessors.</li> <li>- Outstanding requests for assessment</li> <li>- Comparisons to departmental or divisional expectations</li> </ul>
<b>Program Director (PD)</b>	<ul style="list-style-type: none"> <li>- Assessment scores, including narratives, for assessments completed about each resident</li> <li>- Assessment reports that compare and contrast residents for each stage/year</li> <li>- Quantity and variety of assessments for each resident as compared to program expectations (e.g. target scores or frequencies; target variety of assessments, narratives)</li> <li>- Quantity and variety of assessments for each resident as compared to peers (e.g. scores, frequency, variety, narratives)</li> <li>- Completion rate of assessment reports by individual faculty against benchmarks</li> </ul>
<b>Site Director And/or Rotation Director</b>	<ul style="list-style-type: none"> <li>- Assessment scores, including narratives, for assessments completed about each resident</li> <li>- Assessment reports that compare and contrast individual residents to their peers at that stage/year</li> <li>- Completion of required assessments for each resident for that rotation/educational experience</li> <li>- Completion rate of assessment reports by individual faculty</li> </ul>
<b>Resident Advisor</b> (may be separate person or PD)	<ul style="list-style-type: none"> <li>- Assessment scores, including narratives, for assessments completed about each resident they are advising</li> <li>- Assessment reports that compare and contrast each resident they are advising for each stage/year</li> <li>- Quantity and variety of assessments for each resident they are advising as compared to program expectations (e.g. target scores or frequencies; target variety of assessments, narratives)</li> <li>- Quantity and variety of assessments for each resident as compared to peers (e.g. scores, frequency, variety, narratives)</li> </ul>
<b>Competence Committee Chair &amp; members</b>	<ul style="list-style-type: none"> <li>- Assessment scores, including narratives, for assessments completed about each resident</li> <li>- Assessment reports that compare and contrast residents for each stage/year</li> <li>- Quantity and variety of assessments for each resident as compared to program expectations (e.g. target scores or frequencies; target variety of assessments, narratives)</li> <li>- Quantity and variety of assessments for each resident as compared to peers (e.g. scores, frequency, variety, narratives)</li> <li>- Completion rate of required assessments for each stage/year of education</li> <li>- Completion rate of assessment reports by individual faculty against benchmarks</li> </ul>
<b>Residency Program Committees</b>	<ul style="list-style-type: none"> <li>- Assessment scores, including narratives, for assessments completed about each resident</li> <li>- Assessment reports that compare and contrast residents for each stage/year</li> <li>- Quantity and variety of assessments for each resident, compared to program expectations (e.g. target scores or frequencies; target variety of assessments, narratives)</li> </ul>

<b>Role</b> (Note: often people have > 1 role)	<b>Samples of needed assessment data in reports</b>
	<ul style="list-style-type: none"> <li>- Quantity and variety of assessments for each resident as compared to peers (e.g. scores, frequency, variety, narratives)</li> <li>- Completion rate of assessment reports by individual faculty, as compared to peers</li> <li>- Quality of assessment reports by individual faculty</li> </ul>
<b>Remediation Coordinators</b>	<ul style="list-style-type: none"> <li>- Assessment scores, including narratives, for assessments completed about resident</li> <li>- Assessment reports and that compare and contrast the resident with other residents at that level</li> <li>- Quantity and variety of assessments for each resident, compared to Remediation Plan expectations (e.g. target scores or frequencies; target variety of assessments)</li> <li>- Quantity and variety of assessments for resident as compared to peers (e.g. scores, frequency, variety)</li> <li>- Areas of strength</li> <li>- Gaps noted</li> </ul>

Some of this information will not be available through score reports (e.g. service needs of rotations or clarifying information from faculty), however the other needs dictate that the Competence Committees receive as much information as possible, yet it must be presented in a format that makes it easy to understand.

When relevant, legends for terms and abbreviations as well as Scoring Rubrics<sup>3</sup> [8, Section 2.12] will need to be provided to assist in interpretation of results. This applies to all stakeholder groups, though the specifics may differ depending on the stakeholder needs.

Some general guidelines for reporting can be inferred from a recent study of reporting related to teacher evaluations, conducted by the University of Toronto [11].

***It is important to recognize that each particular assessment provides only a partial view of the resident.***

Multiple sources of evidence (i.e., triangulation of data) is necessary for a more fulsome perspective of a resident’s progress. It is also important to recognize that there is a margin of error within all of our assessments; contextual variables change, assessors are different in their personal perspectives of competence, etc.

This plays into the interpretation of scores, particularly when there is no benchmark due to small sample sizes. In the first few years of Competence Committee, there will be no historical data to provide guidance.

***Decisions about individual residents “...should be drawn from clear trends and patterns (e.g. not from isolated comments or scores), after considering all available data, and in consideration of context(s)...”*** [11, Core Principles for Effective Interpretation].

---

<sup>3</sup> 2.12 “Scoring Rubrics” are the scoring guides used to assess performance for individual assessments and across assessment plans

It was noted in the University of Toronto study that it is generally best to consider the overall score, rather than one individual rating in isolation. Similar work in medical education, comparing global ratings and checklists, has indicated that global scores have validity and reliability above that of individual ratings on a scale [12-16].

This can be extrapolated here to suggest that the **overall score on the EPAs, the ITARs, MSFs, or other individual assessments can be used as a measure of resident performance**. If aggregated over multiple assessments (e.g. the intrinsic role ratings on the ITARs, or specific milestones that repeat over multiple EPAs), the individual ratings may also provide evidence of trends in performance.

When looking at scores, considering the distribution of scores is important. [17, as reported in Centre for Teaching Support and Innovation University of Toronto 2018]. Most commonly, the distribution is “skewed,” with the majority of responses ranging from 3-5 on the five-point scale. This is a common pattern that has been observed elsewhere in many types of ratings [18, 19, both as reported in Centre for Teaching Support and Innovation University of Toronto].

Any resident can receive a small number of low scores; however **it is the trend that is important**.

**Residents whose performance deviates from their peers may need to be considered more closely** [11, Core Principles for Effective Interpretation]. At the same time, **comparators must be considered cautiously if they are based on small numbers**.

**Qualitative comments are frequently particularly helpful in understanding resident performance** [20-22]. This is **especially true when there is a theme that runs through the comments on different assessment tools and from different faculty**.

At times positive comments may be shorter in length and negative comments more verbose. This may make the negative comments appear more meaningful, however this may not be the case [11 Core Principles for Effective Interpretation]. Again, **looking for themes will be important**.

The draft guidelines presented below will assist in ensuring consistency across the residency programs in how they present their data to the different stakeholders.

### **What learners need**

Residents need to be involved in the review and interpretation of their assessment reports. In addition to easy to read reports, residents usually need assistance to understand the meaning of the reports to their progress in residence and support to identify areas of strength, deficiencies that need improvement, solutions to support progress. Programs need to establish whose role it is to help resident understand their assessment reports and provide guidance for improvement and progress.

## References

1. Royal College of Physicians and Surgeons of Canada, College of Family Physicians of Canada, and Collège des médecins du Québec. *General Standards of Accreditation for Residency Programs*. 2018 July 1, 2018 [cited 2018 December 2]; V. 1.2:[Available from: <http://www.canrac.ca/canrac/canrac/documents/general-standards-accreditation-for-residency-programs-e.pdf>].
2. Dijkstra, J., C.P. van der Vleuten, and L.W.T. Schuwirth, *A new framework for designing programmes of assessment*. *Advances in Health Science Education*, 2010. **15**(3): p. 379–393.
3. Eva, K.W., et al., *Towards a program of assessment for health professionals: from training into practice*. *Advances in Health Science Education*, 2016. **21**(4): p. 897-913.
4. Hauer, K.E., et al., *Translating Theory Into Practice: Implementing a Program of Assessment*. *Academic Medicine*, 2018. **93**(3): p. 444-450.
5. van der Vleuten, C., et al., *A model for programmatic assessment fit for purpose*. *Medical Teacher*, 2012. **34**(3): p. 205-214.
6. Steele, D., C. Li, and G.A. Babcock. *Programs of Assessment. Best Practices in Evaluation and Assessment (BPEA) 2017* [cited 2018 December 2]; Available from: <http://cbme.postmd.utoronto.ca/wp-content/uploads/2017/09/Theme-4-BPEA-Programs-of-Assessment.pdf>.
7. American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, *Standards for educational and psychological tests*. 1999, Washington, DC: American Educational Research Association.
8. Postgraduate Medical Education University of Toronto. *Guidelines for the Assessment of Postgraduate Residents of the Faculty of Medicine at the University of Toronto*. 2017 [cited 2018 December 2]; Available from: [https://pg.postmd.utoronto.ca/wp-content/uploads/2017/04/Guidelines\\_ASSESSMENT-OF-PG-RESIDENTS\\_Approved\\_17mar09SEND.pdf](https://pg.postmd.utoronto.ca/wp-content/uploads/2017/04/Guidelines_ASSESSMENT-OF-PG-RESIDENTS_Approved_17mar09SEND.pdf).
9. Zenisky, A.L. and R.K. Hambleton, *Developing Test Score Reports That Work: The Process and Best Practices for Effective Communication*. *Educational Measurement: Issues and Practice*, 2012. **31**(2): p. 21-26.
10. Post MD Education. *PGME Minimum Standards for Competence Committees*. 2018 [cited 2018 December 6]; Available from: <http://cbme.postmd.utoronto.ca/?download=677>
11. Centre for Teaching Support and Innovation University of Toronto. *University of Toronto Course Evaluation Interpretation Guidelines for Academic Administrators*. 2018 [cited 2018 December 5]; Available from: [https://teaching.utoronto.ca/wp-content/uploads/2018/10/Interpretation-Guidelines\\_Final\\_Oct.1.2018.pdf](https://teaching.utoronto.ca/wp-content/uploads/2018/10/Interpretation-Guidelines_Final_Oct.1.2018.pdf).
12. Hodges, B., et al., *OSCE checklists do not capture increasing levels of expertise*. *Academic Medicine*, 1999. **74**(10): p. 1129-1134.
13. Hodges, B. and J.H. McIlroy, *Analytic global OSCE ratings are sensitive to level of training*. *Medical Education*, 2003. **37**(11): p. 1012-1016.
14. Bunmi, S.M.-A., et al., *Inter-Rater Reliability: Comparison of Checklist and Global Scoring for OSCEs*. *Creative Education*, 2012. **3 (special issue)**: p. 937-942.
15. Tavares, W., et al., *Global rating scale for the assessment of paramedic clinical competence*. *Prehospital Emergency Care*, 2012. **17**(1): p. 57-67.
16. Ilgen, J., et al., *A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment*. *Medical Education*, 2015. **49**: p. 161-173.

17. Linse, A.R., *Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees*. Studies in Educational Evaluation, 2017. **54**: p. 94-106.
18. Nulty, D.D., *The adequacy of response rates to online and paper surveys: what can be done?* Assessment & evaluation in higher education, 2008. **33**(3): p. 301-314.
19. Theall, M. and J. Franklin, *Using student ratings for teaching improvement*. . New Directions for Teaching and Learning, 1991. **48**: p. 83-96.
20. Ginsburg, S., et al., *Cracking the code: residents' interpretations of written assessment comments*. Medical Education, 2017. **51**(4): p. 401-410.
21. Ginsburg, S., K.W. Eva, and G. Regehr, *Do In-Training Evaluation Reports Deserve Their Bad Reputations? A Study of the Reliability and Predictive Ability of ITER Scores and Narrative Comments*. Academic Medicine, 2013. **88**(10): p. 1539-1544.
22. Regehr, G., et al., *Using "Standardized Narratives" to Explore New Ways to Represent Faculty Opinions of Resident Performance*. Academic Medicine, 2012. **87**(4): p. 419-427.