Workplace-based assessments in the Singapore radiology residency programme – aiming for the next milestone

Martin Weng Chin H'ng[1], MBBS, FRCR, *Nur Ayudia Binte Kassim*[2], BBBM, *Daniel En-Shen Wong*[1], MBBS, FRCR

INTRODUCTION

Postgraduate training in the medical specialities in Singapore began transitioning from the British system to the American Accreditation Council for Graduate Medical Education (ACGME) residency programme (RP) at around 2010. Singapore became the first country outside the United States to embrace this training programme, and radiology training was one of the early adopters of this change.[1] Previously, training was likened to an apprenticeship where experience was gained 'on the job' and logbook documentation was akin to the keeping of a personal diary.[2-4] The learning journey was opportunistic and the contents of the logbook may not have been regularly reviewed to detect deficiencies in a timely manner.[5] The newer framework described six core clinical competencies for the trainee, now termed 'resident', and introduced multiple workplace-based assessment (WBA) tools to assess these competencies.[6-8] The WBA tools served as formative assessments to regularly evaluate the resident's performance in order to give feedback and identify gaps for improvement so that they can be remedied early.[2,3,9,10] The three tools that our programme utilises are: (a) the mini-clinical examination (CEX) applied to reading of radiographs or interpretation of computed tomography images; (b) directly observed procedure (DOP) for a fluoroscopic or interventional radiology task; and (c) the Global Performance Assessment (GPA) form at the end of a rotation.

In 2013, the ACGME rolled out Milestones as part of the Next Accreditation System (NAS).[11,12] It expanded upon the six core competencies, resulting in additional sub-competencies.[11,13] In line with this movement, radiology emerged, once again, as one of the early responders, and we are currently in the process of developing local Milestones. This makes it a good time to reflect on the performance of our raters thus far in using WBA tools. Presently, only a handful of articles have reported on the local experience in the RP and even fewer dealt with the specialty of radiology.[1,4,5]

This research aimed to determine if our WBAs have been undertaken appropriately, by using the GPA form (Appendix) as a proxy, for the following three reasons. Firstly, the GPA form is one of the most frequently used WBA tools and purports to assess all the six core competencies.[7-8] Secondly, it is a common form completed for every resident in each rotation, unlike a DOP that would be used for an intervention radiology posting or a mini-CEX that would be utilised for a general body posting. Thirdly, it is the one used by the widest range of raters and hence can be a good gauge of our specialists.

METHODS

The population of this study included all residents enrolled in the Radiology RP of one of the three accredited radiology programmes in Singapore. Our study involved the retrospective review of secondary data collected since its inception in July 2011 until July 2017. The training period for radiology spans five years and every resident undergoes three-month-long postings during this time. Each resident is assessed using the GPA form at the end of each three-month posting (or after six months if he/she stays for two consecutive postings in the same hospital). The programme director nominates one rater, a qualified radiologist, to undertake the assessment and complete the entire form. The radiologist may be part of the core or clinical faculty, the former being an official appointment, and is granted protected time for RP duties. All the core faculty and the majority of the clinical faculty have undergone WBA training, but in the utilisation of the mini-CEX and not specifically the GPA. The selection of the rater was based on Radiology Information System/Picture Archiving and Communications System data indicating which radiologist had the greatest number of interactions with the resident during that posting. After being assigned, raters are given up to two weeks to deliberate before submitting their assessment.

The GPA form for radiology was derived after making a few modifications to a generic form used by the sponsoring institution. Among all the question items, item Q22 epitomises the complexity of translating clinical criteria to make them applicable to radiology. As such, the equivalent targets could be interpreted as the ability of the resident to understand the gist of the clinical problem, his/her accuracy and speed of reporting, and timeliness in informing the clinician of a critical finding. The form consisted of 22 question items that assessed the six core competencies. Each item was marked on a 9-point Likert scale. There was an option for not giving a score and instead selecting 'Not applicable/not observed' (NA/NO), with no distinction made between the two. There was also a space below the 22 items for comments, which allowed the rater to give qualitative feedback.

We investigated our raters using descriptive statistics and the dimensionality of assessment by employing exploratory factor analysis (EFA). Analysis was performed using IBM SPSS Statistics

[1]Department of Diagnostic Imaging, Tan Tock Seng Hospital, [2]NHG Residency, National Healthcare Group, Singapore
Correspondence: Dr Martin H'ng Weng Chin, Senior Consultant, Department of Diagnostic Imaging, Tan Tock Seng Hospital, 11 Jalan Tan Tock Seng, Singapore 308433. martinhngwc@gmail.com

Table I. Demographics of the 42 residents and 58 raters.

Characteristic	No. (%)	
	Residents	Raters
Gender		
Male	28 (66.7%)	41 (70.7%)
Female	14 (33.3%)	17 (29.3%)
Ethnicity		
Chinese	34 (81.0%)	38 (65.5%)
Indian	7 (16.7%)	15 (25.9%)
Others	1 (2.4%)	5 (8.6%)

Table II. Total number of valid responses, score range, NA/NO responses and available comments of the 22 items.

Item	Valid total no.	Score		NA/NO	No. of items with comments
		Range	Mean ± SD		
1	297	3–9	6.77 ± 1.022	0	31
2	266	4–8	6.44 ± 0.943	31	20
3	284	3–9	6.64 ± 1.111	13	15
4	275	4–8	6.53 ± 1.012	22	20
5	269	4–9	6.44 ± 0.974	28	18
6	293	3–8	6.56 ± 0.987	4	22
7	296	3–9	6.47 ± 1.138	1	19
8	296	2–9	6.37 ± 1.157	1	18
9	289	2–9	6.18 ± 1.094	8	14
10	293	2–9	6.26 ± 1.091	4	13
11	293	2–9	6.43 ± 1.056	4	11
12	242	4–8	6.06 ± 1.066	55	15
13	248	2–8	6.02 ± 1.087	49	11
14	275	2–9	6.32 ± 1.136	22	14
15	250	3–8	6.11 ± 1.053	47	9
16	231	4–9	6.42 ± 0.942	66	7
17	259	4–9	6.25 ± 1.027	38	17
18	297	3–9	6.67 ± 1.046	0	10
19	235	4–8	6.14 ± 0.970	62	10
20	263	3–9	6.23 ± 1.013	34	10
21	223	3–8	6.14 ± 0.960	74	6
22	297	4–9	6.83 ± 1.007	0	0

NA/NO: not applicable/not observed; SD: standard deviation

version 24.0 (IBM Corp, Armonk, NY, USA). Ethics approval was granted by the National Healthcare Group Domain Specific Review Board.

RESULTS

A total of 297 valid GPA forms were collated from all 42 of our residents to ensure a good sample size in this niche speciality. Two batches totalling 15 residents had graduated from the five-year programme and had been assessed on 9–14 occasions each. There were 58 raters who conducted an average of 5.1 assessments each (Table I). Of these 58 raters, 15 (25.9%) belonged to the core faculty, while 43 (74.1%) were from the clinical faculty and comprised radiologists of varying grades

Table III. Exploratory factor analysis of the 22 items in the Global Performance Assessment form.

Component	Initial eigenvalues		
	Total	% of variance	Cumulative %
1	15.866	72.119	72.119
2	1.173	5.330	77.449
3	0.773	3.513	80.962
4	0.613	2.787	83.749
5	0.382	1.735	85.484
6	0.356	1.619	87.103
7	0.336	1.529	88.631
8	0.297	1.352	89.983
9	0.257	1.168	91.151
10	0.236	1.071	92.223
11	0.230	1.045	93.268
12	0.201	0.915	94.182
13	0.189	0.860	95.042
14	0.164	0.744	95.786
15	0.148	0.673	96.459
16	0.142	0.647	97.106
17	0.134	0.609	97.715
18	0.126	0.572	98.287
19	0.112	0.509	98.796
20	0.103	0.469	99.265
21	0.095	0.432	99.696
22	0.067	0.304	100.000

from Associate Consultant and Consultant to Senior Consultant.

There were no invalid or missing answers to the 22 items (Table II). The mean score by raters was 6.5217 (range 4.6818–8.0714). Only 2 (3.4%) raters gave scores of < 4, while 11 (19.0%) raters awarded the maximum score of 9. 14 (24.1%) raters used three Likert score categories, whereas another 10 (17.2%) raters used two Likert score categories in their assessments. 2 (3.4%) raters utilised only a single Likert score to answer all the items on eight occasions. The 22 items had means ranging from 6.02 to 6.77 with a normal distribution. The NA/NO response was used 563 (8.6%) out of 6,534 times. A high percentage (> 20%) was observed for items Q16 (under Patient Care) as well as Q19 and Q21 (under Systems-Based Practice). On the other hand, very few NA/NO checkboxes were ticked for items Q7–Q10, which were in the Medical Knowledge section. For calculations, NA/NO responses were omitted. In addition, for items that allowed for qualitative feedback, there was a total of 310 (4.7%) comments. EFA was used to examine the construct validity of the items. There were two components whose eigenvalues were > 1 (Table III). The first was significantly more than the second and accounted for 72.1% of the variance.

DISCUSSION

Although the ACGME has given equal weightage to the six core clinical competencies, the literature has revealed that raters have been assessing based on a one- or two-dimensional perspective, with two broad components reflecting either the competence
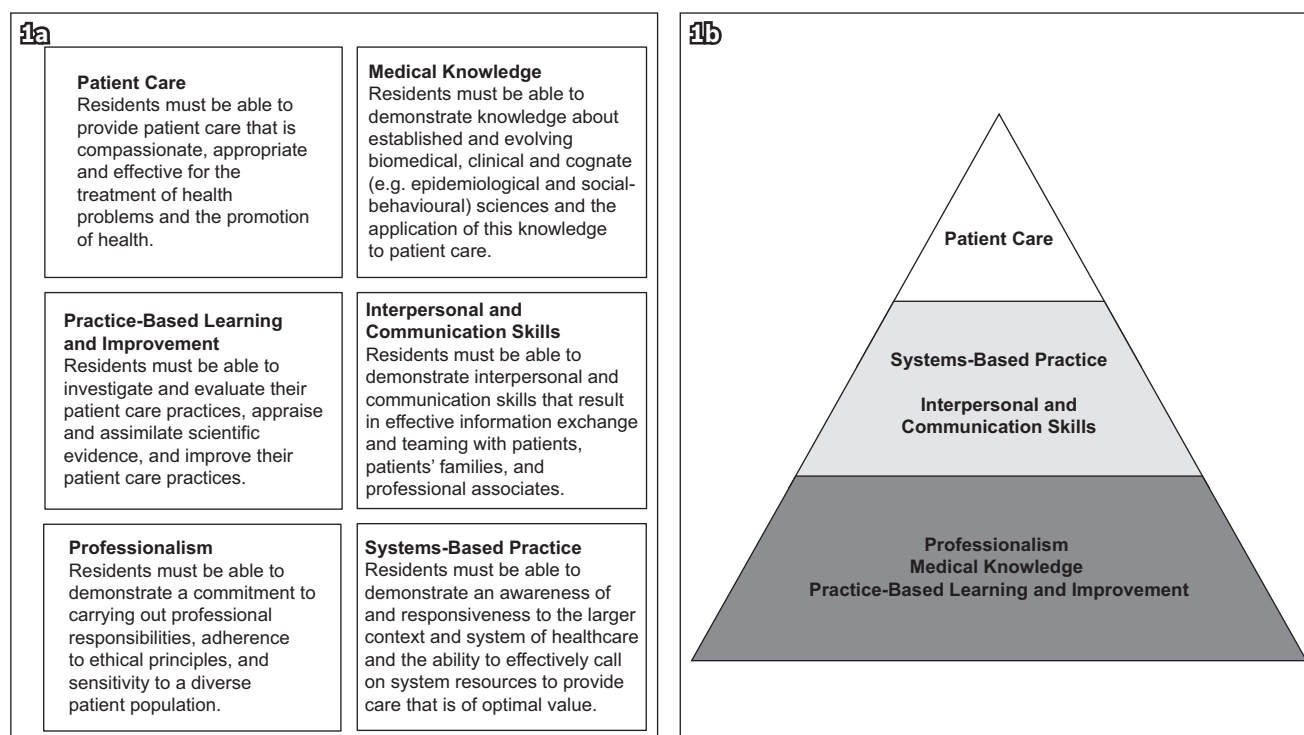
Fig. 1 Diagrams show (a) the six core clinical competencies as defined by the Accreditation Council for Graduate Medical Education, where each component is given equal weightage and (b) the hierarchical model proposed by Khoo et al,[1] which places the patient care component at the apex of a layered pyramid.

or friendliness of the resident.[6,8] The hierarchical layering described by Khoo et al, which places patient care as the most important competency, may explain the single dimension in our local assessment. It is based on the belief that this competency encompasses the knowledge and skills of the other five (Fig. 1), a concept not unlike that of total patient care in the traditional British system.[1] The raters in this study may still be holding on to that model and basing their assessment on that single dimension. Hence, once a resident gave the impression of being able to deliver total patient care, the six competency scores could then be moderated to produce that perceived average mark. The fact that only a single score was used to represent all 22 items on eight occasions may attest to this.

A rather similar trend, termed straight-line scoring (SLS), has been reported when using Milestones in the ACGME system.[14] One reason cited for SLS is that raters generalise the ratings of one sub-competency to the remaining ones in the same section.[14] It may be due to lack of granularity in differentiating between the sub-competencies or an attempt by raters to simplify work, seeing that in Milestones, the initial six competencies are subdivided into multiple smaller components and the number of required assessments has increased. Given that we are in the development phase for local Milestones, we should aim to avoid this potential problem by crafting our sub-competencies so that they remain as distinct entities. Subsequently, training should emphasise to raters the importance of assessing each sub-competency on its own merits.[14]

Our results show that raters exhibited a central tendency effect, some with additional restriction of range. This could be an influence of the British model, which utilises a close marking system that follows a bell-shaped distribution and rarely gives marks at either extreme.[15,16] Only two raters used Likert scores of < 4, which may be another sequela of the close marking system that aims to avoid giving too low a score in one category so that the resident will have an opportunity to make amends in another category in order to pass.[15]

As our residents still sit for the Royal College fellowship examinations in their third or fourth year of training, the British influence endures, and more research is necessary to confirm or refute its secondary effects on the raters. Furthermore, with the electronic version of our GPA form, giving a failing mark requires raters to justify themselves by providing comments before the form can be successfully submitted. Artificially tweaking marks to ensure that a resident scrapes through with a pass would circumvent this time-consuming chore. Lastly, the absence of criterion standards or behavioural descriptors representing each Likert score in our GPA form may partly explain these rater effects.[17] Milestones attempts to overcome this by having detailed descriptive narratives for each competency and even sub-competency.[11] Furthermore, it includes a timeline that trends proficiency levels from novice to beginner, then to competent, followed by proficient, and finally expert.[11] Despite these modifications, exaggerated marking has still been reported when using Milestones.[12] Generalised inflated scoring may result in an inability to discriminate between residents in different levels of training and, thus, yields no useful information.[12] Hence, improvement in form design should occur parallel to rater training.[11,12] While we advance towards Milestones, concurrent rater training is necessary to ensure that they adhere to the marking rubrics, lest inappropriate usage of these forms clouds future Milestone data.

Another issue with the current GPA form is that no distinction is made between the NA/NO scores, resulting in a reduced number of valid responses for certain items. The 'NA' option can provide raters with an avenue to abstain from giving a committed response, especially if it meant failing the resident.[10,12] The 'NO' option is more specific and reserved for rare events that are not observed during the period being assessed.[12] Very few NA/NO checkboxes were ticked for items Q7–Q10, which is in the Medical Knowledge section, as these were more tangible and easily graded by raters. On the contrary, a high percentage (> 20%) of NA/NO responses were given for items Q16 (under Patient Care) as well as Q19 and Q21 (under Systems-Based Practice), as these items appear more abstract to evaluate but are nonetheless still applicable, as they represent core clinical competencies. Future form design should distinguish 'NA' from 'NO' in order to determine if it is a case of 'failure to fail' or a rare scenario not observed during that posting.[10] Simultaneously, rater training should be targeted at interpreting behavioural descriptors properly so that fewer 'NA' marks are awarded.

The privilege of providing qualitative feedback was underutilised at only 4.7%, compared to rates quoted in the literature of up to 50%.[18] Although this may be an indirect consequence of not wanting to fail a resident and avoiding extra commentary, there could be another reason. Assuming that raters had a traditional mindset, additional comments, especially in a summative assessment, were generally assumed to be negative and rarely given for fear of jeopardising the resident.[19] Although this WBA is undertaken at the end of the posting and may appear summative, it is nevertheless formative because there are more postings yet to come. Raters should be reminded to provide good qualitative feedback instead of one-sided criticism, as the intention is to guide the resident to improve for future rotations.

The main limitation of this study is that it is descriptive. Hence, complementary qualitative research would help to elucidate whether our raters were still influenced by British practice and/or were merely inexperienced in navigating the new ACGME forms. The other limitation lies with employing the GPA form as a proxy for our other WBAs. It is less specific when compared to the DOP or min-CEX.[13] This form is usually completed at the end of the rotation and records one's impression of the resident over a period of time and not for a specific task. WBAs are better graded at or around the time of assessment, and this retrospective completion introduces lapses during recollection of events.[20] Some items may be given unfair weightage due to the 'recency effect', where the most current performance of the resident will have a higher bearing in the eyes of the rater, while responses to other items may be limited to selective recall or secondhand information.[6,7,9]

In conclusion, radiology training in Singapore is unique, given its a long association with the Royal College and the recent move to pursue ACGME's RP and NAS. The results of this study reveal that our radiology training is still traditional and tends to focus more on tangible competencies such as medical knowledge and how the resident communicates with fellow healthcare professionals. However, it is not impossible to straddle both new and old systems as long as we keep the criteria of each one separate and adopt distinct mindsets when performing their respective assessments. As we transit towards using Milestones, improvements in form design and concurrent rater training are even more crucial than ever.

ACKNOWLEDGEMENTS

SUPPLEMENTARY MATERIAL

The Appendix is available online at https://doi.org/10.11622/smedj.2021028.

REFERENCES

1. Khoo SM, Lahiri M, Huggan PJ, et al. When traditional model meets competencies in Singapore: beyond conflict resolution. Ann Acad Med Singapore 2014; 43:544-9.
2. Bari V. Direct observation of procedural skills in radiology. AJR AM J Roentgenol 2010; 195:W14-8.
3. Miller A, Archer J. Impact of workplace based assessment on doctors' education and performance: a systematic review. BMJ 2010; 341:c5064.
4. Yang H, Tan CJ, Lau DA, et al. Competency-based radiology residency: a survey of expectations from Singapore's perspective. Ann Acad Med Singap 2015; 44:98-108.
5. Peng LL, Ooi SB. Emergency medicine residency programme in Singapore – where are we at since inception? Ann Acad Med Singap 2015; 44:77-8.
6. Williams RG, Dunnington GL, Mellinger JD, Klamen DL. Placing constraints on the use of the ACGME milestones: a commentary on the limitations of global performance ratings. Acad Med 2015; 90:404-7.
7. Haurani MJ, Rubinfeld I, Rao S, et al. Are the communication and professionalism competencies the new critical values in a resident's global evaluation process? J Surg Educ 2007; 64:351-6.
8. Silber CG, Nasca TJ, Paskin DL, et al. Do global rating forms enable program directors to assess the ACGME competencies. Acad Med 2004; 79:549-56.
9. Brazil V, Ratcliffe L, Zhang J, Davin L. Mini-CEX as a workplace-based assessment tool for interns in an emergency department – does cost outweigh value? Med Teach 2012; 34:1017-23.
10. Barrett A, Galvin R, Steinert Y, et al. Profiling postgraduate workplace-based assessment implementation in Ireland: a retrospective cohort study. Springerplus 2016; 5:133.
11. Sullivan G, Simpson D, Cooney T, Beresin E. A milestone in the milestones movement: the JGME milestones supplement. J Grad Med Educ 2013; 5:1-4.
12. Dehon E, Jones J, Puskarich M, Sandifer JP, Sikes K. Use of emergency medicine milestones as items on end-of-shift evaluations results in overestimates of residents' proficiency level. J Grad Med Educ 2015; 7:192-6.
13. Dougherty PJ. What the ACGME's next accreditation system means to you. Clin Orthop Relat Res 2013; 471:2746-50.
14. Beeson MS, Hamstra SJ, Barton MA, et al. Straight line scoring by clinical competency committees using emergency medicine milestones. J Grad Med Educ 2017; 9:716-20.
15. Beasley SW, Wannan C, Hardware N. Justification and implications of the introduction of an expanded close marking system for the fellowship examination. ANZ J Surg 2013; 83:444-7.
16. Hobsley M. Primary FRCS (Eng). I. Evaluation of assessment techniques. Ann R Coll Surg Engl 1974; 55:161-8.
17. Tomisato S, Venter J, Weller J, Drachman D. Evaluating the utility, reliability, and validity of a resident performance evaluation instrument. Acad Psychiatry 2014; 38:458-63.
18. Barrett A, Galvin R, Steinert Y, et al. A BEME (Best Evidence in Medical Education) review of the use of workplace-based assessment in identifying and remediating underperformance among postgraduate medical trainees: BEME Guide No. 43. Med Teach 2016; 38:1188-98.
19. Saedon H, Salleh S, Balakrishnan A, Imray CH, Saedon M. The role of feedback in improving the effectiveness of workplace based assessments: a systematic review. BMC Med Educ 2012; 12:25.
20. Barrett A, Galvin R, Scherpbier AJ, et al. Is the learning value of workplace-based assessment being realised? A qualitative study of trainer and trainee perceptions and experiences. Postgrad Med J 2017; 93:138-42.

Appendix

The Global Performance Assessment form

Items Q1 to Q22 use a 9-point Likert scale with a space for comments below.

1–3	Unsatisfactory
4–6	Satisfactory
7–9	Superior
NA/NO	Not applicable/not observed

A. COMPETENCY

Professionalism *(passing mark is 4)*
Q1 Accepts responsibility and follows through on tasks. Does so willingly; industrious; completes tasks carefully and thoroughly.

Q2 Responds to patient's unique characteristics and needs equitably. Provides equitable care regardless of patient culture, disability or socioeconomic status.

Q3 Demonstrates integrity and ethical behaviour. Patient before self, addresses ethical dilemmas; takes responsibility for actions.

Interpersonal & Communication Skills *(passing mark is 4)*
Q4 Demonstrates care and concern for the patient/family. Establishes rapport; respectful and compassionate.

Q5 Communicates effectively with patient/family. Good verbal and non-verbal skills; involves patient or family in decision-making.

Q6 Communicates and works effectively with other healthcare professionals. Good medical records, summaries and referrals; considerate to other healthcare professionals.

Medical Knowledge *(passing mark is 4)*
Q7 Demonstrates good basic science knowledge. Intelligently discusses pathophysiology and basic sciences within his/her level.

Q8 Ability to apply knowledge in the clinical context. Intelligently discusses diagnosis, evaluation and treatments within his/her level.

Q9 Demonstrates up-to-date knowledge. Cites recent literature when appropriate, asks well-informed and knowledgeable questions.

Q10 Demonstrates good analytical thinking and problem-solving techniques. Demonstrates good analytical approach and problem-solving techniques in a medical setting.

Practice-Based Learning & Improvement *(passing mark is 3)*
Q11 Engages in ongoing learning. Does extra reading and surgical/procedural practice when needed; uses IT to add learning.

Q12 Facilitates the learning of others. Teaches/coaches junior colleagues and students; directs learners to relevant resources.

Q13 Understands and integrate concepts of quality improvement into practice. Systematically reviews outcomes; reflects to identify strengths and weaknesses; improves.

Patient Care *(passing mark is 4)*
Q14 Demonstrates comprehensive assessment to reach appropriate diagnosis. Thorough history, physical exams, investigations and appropriate diagnosis.

Q15 Provides the appropriate ongoing management based on best clinical practice. Synthesises and implements treatment plans using evidence-based medicine, protocols and specialist inputs.

Q16 Responds appropriately to emergency clinical problems. Initiates appropriate care and procedures in emergencies as part of team.

Q17 Demonstrates procedural skills appropriate to level of training. Demonstrates knowledge of indications and risks; technical ability, minimises patient discomfort.

Q18 Practises within the scope of his/her abilities. Makes correct judgement to consult and/or ask for help when needed.

Systems-Based Practice *(passing mark is 3)*
Q19 Provides cost-conscious medical care. Considers costs/benefits of care; adheres to pathways; does not order unnecessary tests.

Q20 Works to promote patient safety. Identifies system causes of medical error; adheres to patient safety protocols.

Q21 Coordinates care with providers in the larger healthcare community. Provides care options; makes appropriate referrals; assists with arrangement and follow-up.

B. **WORK OUTPUT EVALUATION** *(passing mark is 3)*
Q22 Able to organise, prioritise and finish work on time; includes summarising old notes, acknowledging and tabulating results, writing accurate and concise summaries, good medical record keeping and documentation. Suggested activities to base the evaluation on:
- CCA – timeliness and regular updates
- Old notes summary – timeliness
- Result acknowledgement
- Discharge summary – timeliness and audit (inpatient work)
- Number of days of medical leave (obtain from department AA)