

RESEARCH

Open Access



Development and use of a computerized system to track the competency development of family medicine residents: analysis of the convergence between system proposals and assessor decisions

Isabelle Savard^{1*}, Luc Côté², Abdelhamid Kadhi³, Caroline Simard⁴, Christian Rheault² and Miriam Lacasse^{2,4}

*Correspondence:

Isabelle.savard@teluq.ca

¹ Department of Education,
Téluq University, 455 Du
Parvis, Quebec, QC G1K-9H6,
Canada

Full list of author information
is available at the end of the
article

Abstract

In recent decades, a number of training environments have moved toward program approaches targeting the development of competencies. Because of their complexity, monitoring the development of those competencies is a considerable challenge. Our hypothesis is that a computerized system could help overcome this challenge if it is well accepted by its users. We first summarize the context surrounding the implementation of such approaches. Next, we present a computerized assessment system established in the Family Medicine Residency Program of Laval University (Québec, Canada) that we have developed for tracking the development of residents' competencies. We then present the analysis of interactions between the system and users and the various proposals that were made to improve the system and longitudinal tracking of the development of the targeted competencies. We consider that this research provides useful guidelines for the computerized monitoring of learners' competencies development and for the design of such systems.

Keywords: Competencies development, Computerized assessment system, Program approaches, User-system interactions, Competency development

Introduction: context

In recent decades, a number of training environments have revised their educational strategy and moved toward program approaches targeting the development of competencies. This is particularly the case in health sciences in Québec, where programs in medicine, physiotherapy, occupational therapy and speech therapy now have their own competency frameworks. The definition provided by Tardif (2006), whereby a competency is “a complex knowledge of what to do relying on effectively harnessing and combining a variety of internal and external resources within a family of situations,” has influenced most of these frameworks. The experience of recent years has shown that the

assessment and documentation of competency development has also proven to be very complex. Training programs can no longer consist of a sum of activities or courses that are juxtaposed or separate from each other; they must be part of a program approach (Basque, 2017; Prigent et al., 2009) within which the program becomes a cohesive whole that is greater than the sum of its parts. In fact, in a competency-based training curriculum, learning assessment must be done on an ongoing basis, to ensure assessment FOR learning (rather than assessment OF learning). Thus, assessment strategies should give priority to continuous, documented formative feedback, which fosters the progress of learners. However, competency development must also be guided by decisions using a summative approach.

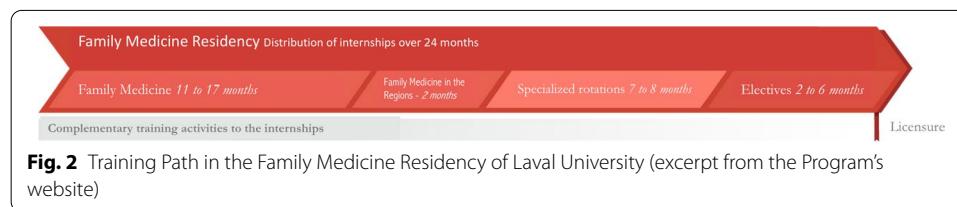
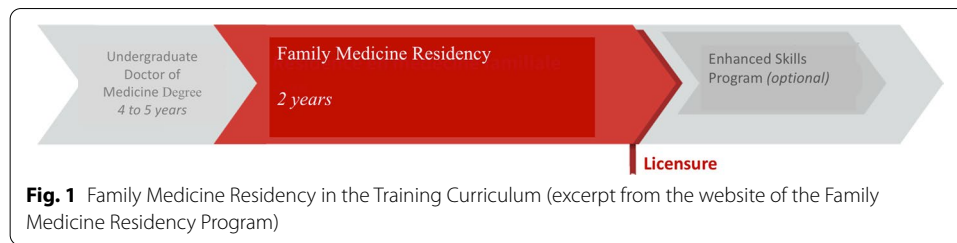
Residency programs in medicine, regardless of the specialty, use a variety of methods to assess residents, but they still rely heavily on global normative assessment scales at the end of a rotation (Chou et al., 2009), mainly because of their usability. These scales are not well adapted to a competency-based approach. Rather than interpreting the performance of residents in a norm-referenced manner, based on their placement within a group, the assessment should be conducted using a criterion-referenced approach in order to assess their performance level on a descriptive scale, through multiple measures based on authentic situations (Carraccio et al., 2002). Thus, residents' progress should be tracked using descriptive scales that include different performance levels. These scales, also known under the terms "developmental benchmarks," "milestones" or "rubrics," specify expectations at various important stages of training for a number of areas or contexts of practice (Tardif, 2006).

The fact that norm-referenced interpretation practices are so firmly established with teachers represents a considerable challenge for implementing a criterion-referenced approach. To this is added another daunting challenge, i.e., the longitudinal documentation of competency development, due precisely to the complexity of assessing competencies, using formative and summative approaches, and due to the fact that they must be observed in different situations and in varied contexts.

Can a computerized system promote such longitudinal tracking? If so, how can we ensure that such a system will be well accepted by its users? Is it possible to obtain a convergence¹ between the system proposals and the users' decisions?

The next section first describes the computerized competency assessment system implemented in the Family Medicine Residency Program of the Faculty of Medicine of Laval University (Québec, Canada). Subsequently, we present the results of the analysis of the convergence between the system proposals (made based on the program expectations) and the assessors' judgement. Such analysis is important, because it can improve the credibility and acceptability of the computerized system's suggestion. To this end, the qualitative analysis of the reasons provided by the assessors, when there was a discrepancy between the decision proposed by the system and their own decision, is also presented.

¹ Convergence is the action of reaching the same result.



Case description: computerized competency assessment system in family medicine

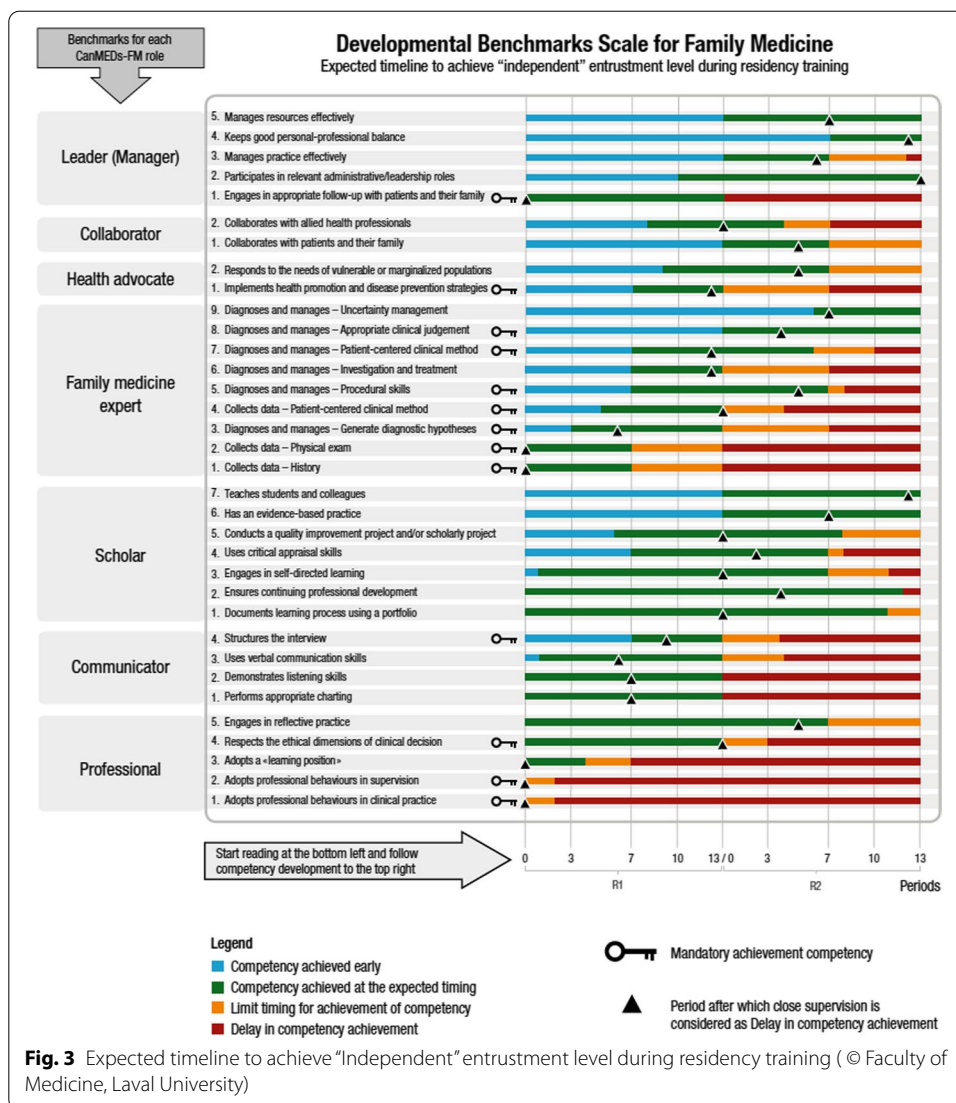
Residents' training path: a variety of contexts and assessors

As shown in Fig. 1, after completing the Undergraduate Doctor of Medicine Program in Family Medicine, which is four to five years in duration, depending on the pathway of the student, the latter must complete a residency program. Specialization in family medicine requires two years and leads to a certification from the College of Family Physicians of Canada, mandatory for licensure by the Collège des médecins du Québec [Quebec College of Physicians].

The Family Medicine Residency Program at Laval University's Faculty of Medicine welcomes about 125 new residents each year, for a two-year program: first-year residents (R1) and second-year residents (R2). A network of clinical teachers comprising more than 1,100 family physicians and many other specialists supervise and evaluate these residents using formative and summative approaches.

Along the Family Medicine Residency Program, residents must complete different rotations divided into 26 periods. A period equivalent to 14 months is devoted to family medicine rotations, two of which are completed in clinical settings located far from major urban centers. The other ten months include specialized rotations and elective rotations, including some opportunities with family physicians who have a focused practice. Residents thus benefit from a variety of rotations, which are assessed by many clinical teachers during the program (Fig. 2).

Academic half-days are held throughout the two years of training, allowing for clinical case discussions, basic courses, seminars, simulated medical interviews, reading clubs, etc. These rotations and academic curriculum activities are opportunities for formative and summative assessment of the residents, based on the seven roles of the CanMEDs-FM framework of the College of Family Physicians of Canada: Leader, Collaborator, Health Advocate, Family Medicine Expert, Scholar, Communicator and Professional (Shaw et al., 2017).



Competencies, and benchmarks and timelines for their achievement

The developmental benchmarks developed and validated by Laval University’s Family Medicine Program (Lacasse et al., 2014, 2017; M.-L. Simard et al., 2017) characterize the program expectations with respect to the development of thirty-four competencies during the two years of training. Figure 3 presents the expected timelines for developing each competency level during the residency program. Three levels of supervision were defined: close supervision, distant supervision, and independent. In addition, the timelines make it possible to determine, for each competency, whether the progression is achieved early, at the expected timing, or is delayed, based on the program expectations. The mandatory achievement competencies are also represented in this figure (indicated by a key), as is the period after which a level of “close supervision” is considered to be a “developmental delay” (represented by a triangle).

From this figure, it can be seen that some competencies should be demonstrated toward the end of the residency program (e.g., Scholar 7- *Teaches students and*

colleagues), whereas others should be shown at the beginning of the residency program (e.g., Professional 1- *Adopts professional behaviours in clinical practice*). The expectations were determined based on the Delphi method with a group of clinical teachers before the system was developed (Lacasse et al., 2017). During this previous study, content and convergence validities among assessors were verified.

Tracking the residents' competency development

Continuous assessment is performed throughout the residency program. Each resident is first paired with a faculty advisor/competency coach (family physician) who supports them in their training path. The latter should encourage residents to adopt a reflective approach facilitating the integration of learning, should identify difficulties and recommend ways to overcome them, and also periodically exchange formative and summative feedback with the residents in order to guide their progress.

A summative assessment of competencies achieved in each of the clinical rotations is carried out by different clinical teachers who supervised the resident during the rotation. This assessment is conducted using a computerized system, presented in the following section. The faculty advisor provides longitudinal tracking through progress reports, based on the overall assessment data.

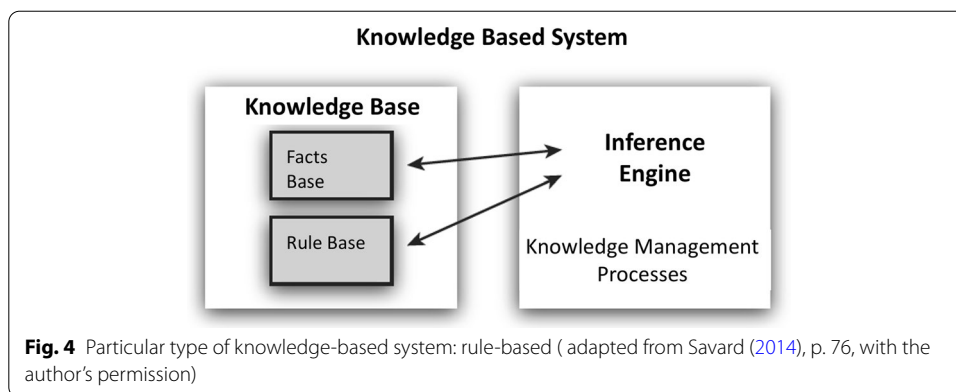
Description of the parameters of the computerized system

The system is based on explicitly represented knowledge (which can be machine-interpreted) of the developmental benchmarks, potential educational diagnoses, and educational prescriptions that may be made. In fact, the metaphors of diagnoses and prescriptions are used in this particular educational context, allowing clinical teachers to use a reasoning process similar to the one used in their medical practice, but by applying it to the assessment of residents. Educational diagnoses are at the root of the difficulties experienced by residents in achieving certain competencies (e.g., problems with knowledge, skills or attitudes, which can be influenced by personal considerations, issues related to the instructor, or environmental factors) (Lacasse et al., 2019). Educational prescriptions correspond to remedial interventions recommended to support the learner, representing “additional teaching going beyond the usual curriculum, personalized for each learner, and without which he could not succeed in developing the competencies that are necessary for the profession” (Guerrasio et al., 2014, p. 803).

The following section first describes the system mechanism and then presents the major steps in its use by users.

Knowledge-based system mechanism

The computerized system is a knowledge-based system (KBS). According to Houdé et al. (2003), the essential characteristic of a KBS is that it manipulates specific knowledge in the field of application, represented explicitly in the knowledge base (KB) and separately from the procedures designed for their use, which are themselves grouped together in the inference engine. A knowledge-based system (KBS) is thus comprised of a knowledge base and an inference engine. The criterion-referenced assessment tool (CAT) is part of a particular type of KBS: rule-based systems. In this type of system, the knowledge base contains a fact base and a rule base.



As illustrated in Fig. 4, this type of system includes a facts base (described in “[Facts base of the knowledge base](#)” section), a rule base (see “[Rule base of the knowledge base](#)” section) and an inference engine (see “[Inference engine of the knowledge-based system](#)” section), which includes the processes designed for using the system. The parts that follow provide a detailed description of the content of each of these knowledge-based system components, for the particular case of CAT.

Facts base of the knowledge base The facts base of the criterion-referenced assessment tool (CAT) includes two types of facts that are collected at two different times: first, the program parameters are defined, and then the assessment data are collected based on these parameters.

Program parameters defined in the management tool

A secure management tool makes it possible to construct the system’s knowledge base and to make different program-specific facts explicit. To do so, the system provides the main functionalities to the persons in charge, who have been previously authorized within the programs (often program directors and their designees), as follows (Table 1):

These facts compiled in the knowledge base (KB) establish all of the structure for receiving the facts of the second component of the facts base: those pertaining to the specific content of residents’ competency assessments for each period.

Data collected through assessment

The content of an assessment is primarily made up of levels of supervision (close supervision, distant supervision, and independent) selected by the assessor for each of the competencies. The development level achieved by the resident during the assessed rotation also represents one of the assessment-specific facts.

So that the system can suggest a result (early, expected, limit timing, or delayed) to the assessor, all of this knowledge must be viewed in association with the development timelines presented in Fig. 3. This is done by means of a table of correspondence, which is part of the rule base presented in the following part.

Rule base of the knowledge base The system’s rule base has three components: the tables of correspondence between levels of supervision and results; the mathematical formula for deducing the overall score; the tables of correspondence between educational prescriptions and competencies, based on the assessment results.

Table 1 Management tool functionalities and facts made explicit

Functionality	Facts made explicit
Program parameters	Makes it possible to specify the general assessment parameters of the residency program, such as the wordings of the response options on the assessment forms, the type of assessment (in the case of the Family Medicine Program = by milestones), etc.
Management of competencies	Makes it possible to manage the data pertaining to the “Competencies” section of the assessment form. In fact, this functionality makes it possible to manage the competencies that must appear in the assessment form of the different assessment sheets of the program. It helps to render explicit the framework of competencies targeted by the program
Management of assessment sheets	Makes it possible to manage the assessment sheets of the residency program and their content. Thus, a program can create a new sheet, modify an existing sheet, or delete one. For each sheet, the program must specify the start and end dates, the content and, in particular, the competencies for which it is necessary to specify whether their assessment is optional or mandatory, the assessment context (residency level, period of residency, internship setting, residency activity)
Educational prescriptions	Makes it possible to manage the data of the different educational diagnoses and prescriptions that will be generated the assessment of a resident for a program activity. The educational diagnoses and prescriptions are organized as follows: One type of educational diagnosis contains one or more subtypes of educational diagnoses One subtype of educational diagnosis is associated with one or more educational prescriptions The educational prescriptions can then be associated with the competencies targeted by the program

Tables of correspondence between levels of supervision and results

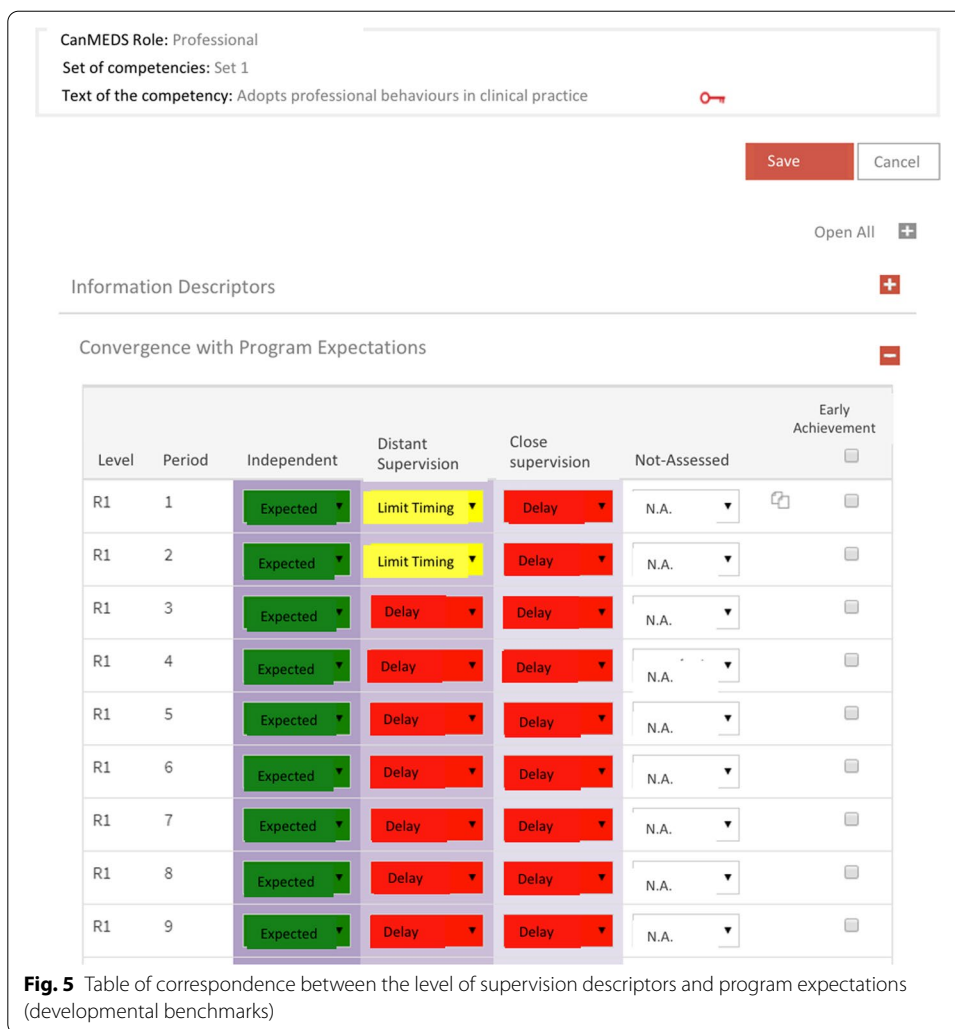
The table of correspondence of a competency is in the form of an interface connecting the levels of supervision (independent, distant supervision, and not assessed) to the results categories (early, expected, limit timing, delayed, not assessed), for each development period of a residency level. This interface is presented in Fig. 5.

This table of correspondence with expectations is organized as follows:

- The table columns display the level of supervision (independent, distant supervision, close supervision, and not assessed);
- The rows of the table display the combination of Level / Training Period;
- The table cells display a drop-down list containing the results categories (early, expected, limit timing, delay, not assessed). Thus, for a particular Level / Development Period combination, a result can be associated with each level of supervision, in order to represent the benchmarks presented in Fig. 3.

Formula for deducing the overall score

This is a mathematical formula for calculating a measurement, called a demerit score, based on the results for each competency in an assessment. This score allows the system to propose an overall score for the assessment. The formula takes into account five variables whose values are established by the Program (Lacasse et al., 2017). Figure 6 shows the formula and specifies the values chosen by the Family Medicine Residency Program for these variables.



Based on the score obtained, the overall score proposed for the assessment will be established according to the following timelines:

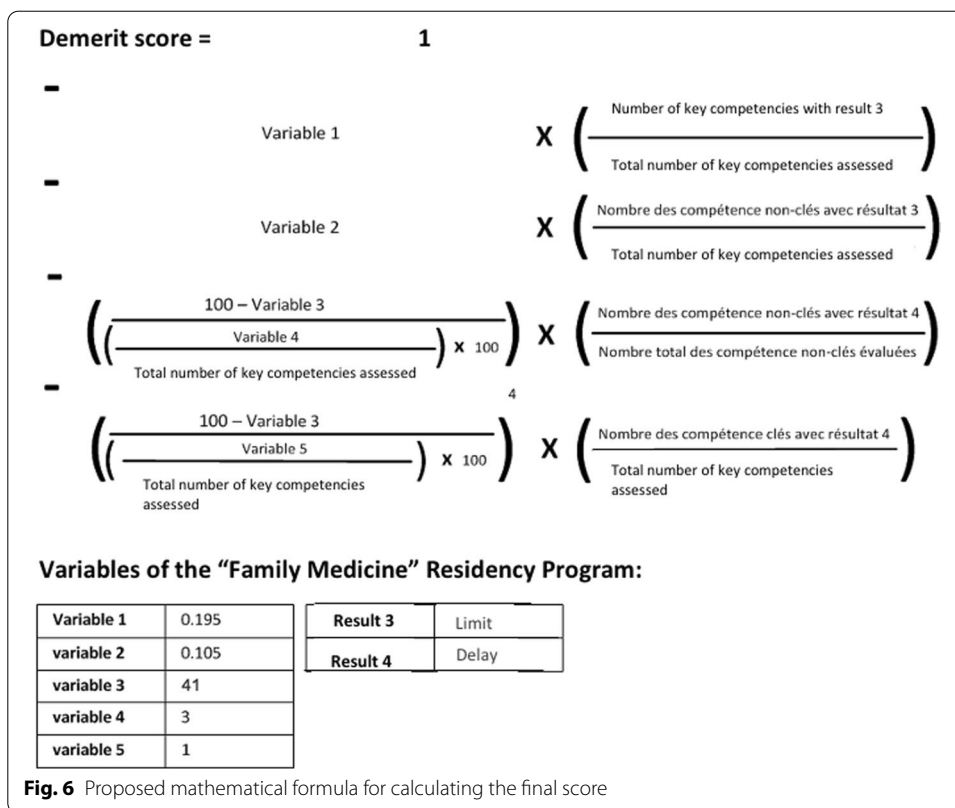
- 0% to 59% → Failure
- 60% to 79% → In difficulty
- 80% to 100% → Success

Tables of correspondence between educational prescriptions and competencies

This table of correspondence is in the form of an interface connecting each of the educational prescriptions with the program competencies. It is presented in Fig. 7.

This figure must be interpreted as follows: the educational prescription “Discussion meeting with a mentor/an educational advisor” must be proposed in an assessment when the competency “Adopts professional behaviours in supervision” is assessed as “limit timing” or a “delay.” This same prescription must also be proposed when the competency “Engages in reflective practice” is assessed as “limit timing” or a “delay.”

This table of correspondence with educational prescriptions is organized as follows:



- The five table columns display the categories of the results (Early, Expected, etc.);
- The rows of the table display the competencies of the program;
- Each cell of the table, corresponding to the intersection between a competency and a result, displays a check box that makes it possible to associate the educational prescription with the competency and the result in question. Thus, when this result is selected for each competency in an assessment, an educational prescription is proposed.

Inference engine of the knowledge-based system The third component of the knowledge-based system is the inference engine, which uses the knowledge base to carry out logical reasoning and deductions in order to reach conclusions. In the case of the CAT, the inference engine includes three elements:

- Algorithm for proposing results for the assessed competencies uses the tables of correspondence for “Levels of Supervision—Results” and the content of an internship assessment, as well as certain program parameters to deduce the most accurate result for each of the assessed competencies;
- Algorithm for proposing overall assessment score uses the formula for deducing the overall score and the content of an internship assessment, as well as certain program parameters for the purpose of deducing the overall score (Pass, In Difficulty, Failure) that must be assigned in the assessment;

Correspondences (Educational Prescription – Competencies)

Text of the educational prescription: Discussion meeting with a mentor/educational advisor
Description of the educational prescription:

Liste des compétences (34)

	Early	Expected	Limit Timing	Delay	Not assessed
Professional	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Set 1					
Adopts professional behaviours in clinical practice	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Adopts professional behaviours in supervision	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Adopts a “learning position”	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Respects the ethical dimensions of clinical decision making	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Engages in reflective practice	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Communicator	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Set 2					
Performs appropriate charting	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Demonstrates listening skills	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Uses verbal communication skills	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

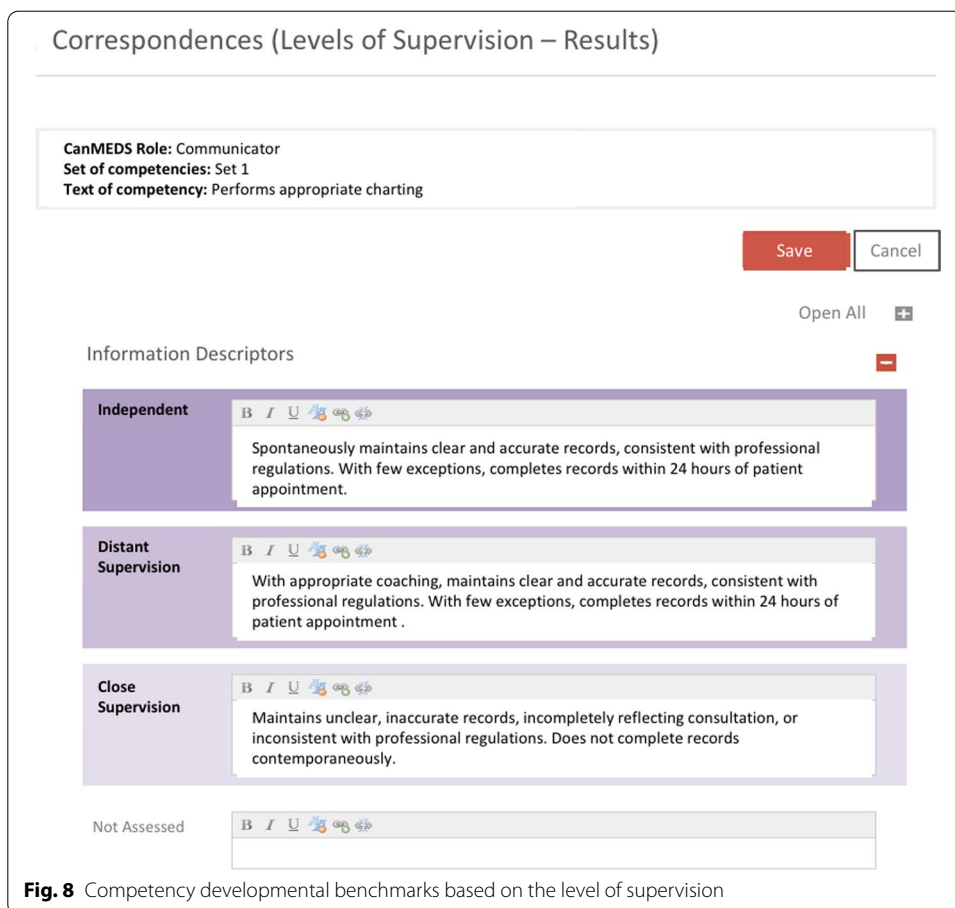
Fig. 7 Table of correspondences—educational prescriptions—competencies

- Algorithm for proposing educational diagnoses and prescriptions uses the tables of correspondence for “Educational Prescriptions—Competencies” and the content of an internship assessment, as well as certain program parameters to identify the list of educational diagnoses and prescriptions that are best suited to the learner’s situation.

One system, three steps

This mechanism and the information collected in the knowledge base enable assessors to use the system. This takes place in three steps: 1) the assessor selects the level of supervision, 2) the system deduces and displays the assessment result, and 3) the assessor decides whether to keep the result proposed by the system or whether to modify it.

Selection of the appropriate level of supervision by the clinical teacher For each competency, developmental benchmarks (descriptive/informative) are explained in detail in order to avoid having users interpret a particular wording in different ways. An example of this, for the role of Communicator, is presented in Fig. 8.

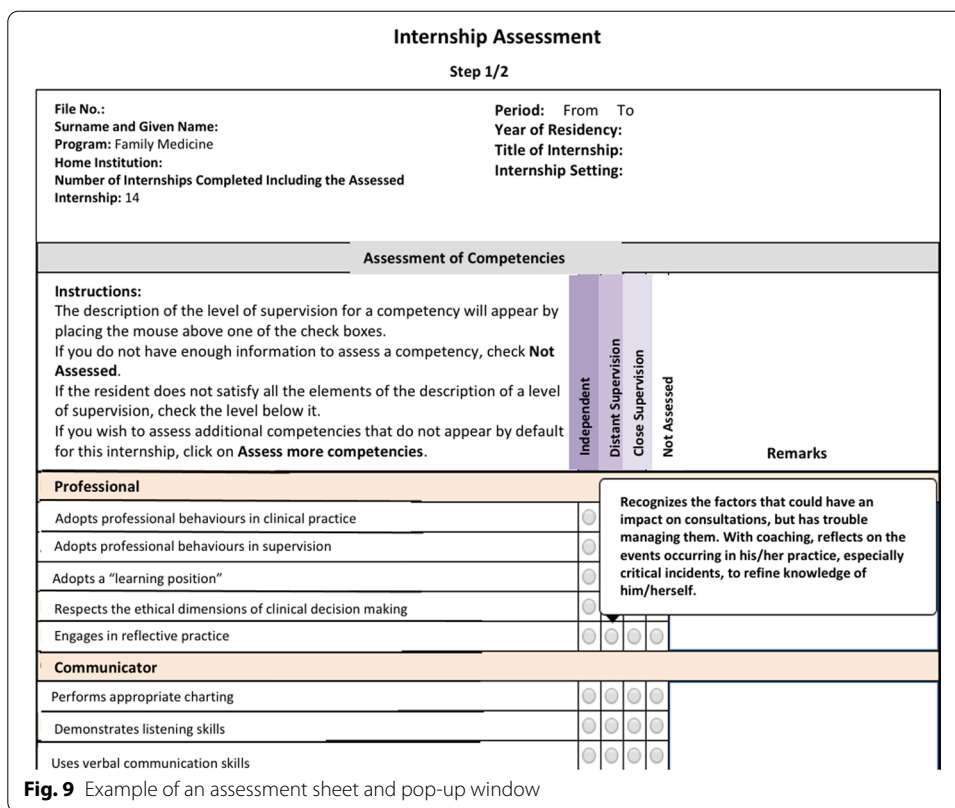


When the assessor places the cursor on the desired check box, these detailed benchmarks entered in the knowledge base (KB) are displayed in pop-up windows (as in Fig. 9).

Deduction and display of assessment results by the system The computerized system compares the resident’s assessment data with the “normal curve” data of the developmental benchmarks (Fig. 5) and determines which of the following results applies to his/her competency development: early, expected, limit timing, or delayed. It displays the result for the assessor to see, as illustrated in Fig. 10.

For competencies whose development is identified as having “limit timing” or a “delay,” the system proposes educational diagnoses and prescriptions based on the table of correspondences presented in Fig. 7.

The assessors can then rely on a list of learning strategies or methods, inspired by an exhaustive review of the literature and of expert opinions (Lacasse, 2009; Lacasse et al., 2019), in order to recommend the best ways for learners to further develop their competencies.



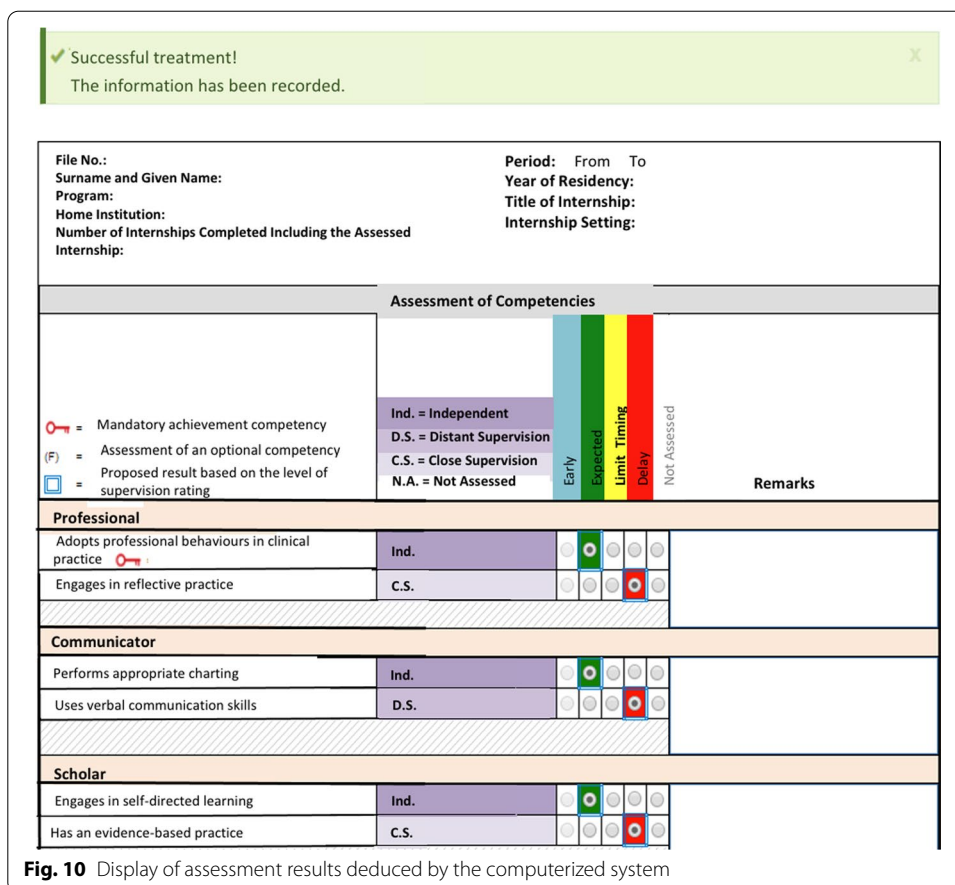
Decision of the assessor regarding the final result For each of the competencies assessed, the assessor decides whether to keep or to modify the result proposed by the system. When he decides to modify the result, he must explain his decision. The reasons for changes in ratings are therefore documented and are the subject of a quantitative analysis presented in the following section.

Method of analysis of rating changes

The competency assessment system of family medicine residents in the 2016–2018 cohort underwent different validations in the context of a Medical Council of Canada (MCC) Research in Clinical Assessment grant.² Various analyses have been carried out. For example, Renaud et al. (2020) performed a psychometric validation of the Laval developmental benchmarks scale for Family medicine. The analyses presented in this article are part of the convergence analysis between the program expectations (the results proposed by the system) and the assessors’ judgement.

Between September 2016 and May 2018, assessors completed 1,432 assessment sheets. They modified at least one rating out of 20.1% (n = 288) of these sheets. These sheets vary depending on the particular features of the different internships, and they each include on average 19.6 competencies to be assessed. In total, 27,891 competencies were assessed during the period analyzed, and 2.4% (n = 657) of them were modified by the

² Principal Investigator: Miriam Lacasse.



assessors. As part of this study, these 1,432 assessment sheets were analyzed, hence the 657 changes in ratings pertaining to them.

The codification and thematic content of the written comments (explanation of the changes in ratings) were analyzed inductively (without predetermined themes or categories) (Thomas, 2006), following the principle of triangulation of researchers (Shenton, 2004): two researchers (IS and LC) first codified the data separately, and then jointly carried out iterative analyses. Frequency calculations were also made.

Results

Number of changes, per role

Table 2 presents the distribution of changes in ratings based on the CanMEDS-FM roles. From this table, it can be seen that the Expert role underwent the greatest number of changes ($n = 269$), most of which are upward changes. The role of Professional follows ($n = 116$) with all of the upward changes, whereas the roles of Collaborator and Health Advocate show primarily downward changes. However, as the roles do not all target the same number of competencies, the average number per role ends up being different.

Table 2 Changes in ratings, increasing or decreasing, according to the roles

CanMEDS-FM role	Number of competencies	Number of changes in ratings			Average number per role
		Increasing ^a	Decreasing ^b	Total	
Leader/manager	5	28	43	71	14.2
Collaborator	2	16	44	60	30.0
Health advocate	2	15	35	50	25.0
Family medicine expert	9	171	98	269	29.9
Scholar	7	20	22	42	6.0
Communicator	4	41	7	48	12.0
Professional	5	116	0	116	23.2
	34	407	249	Grand total: 656 ^c	

^a The assessor is less strict than the system

^b The assessor is stricter than the system

^c There is a change for which the competency is not identified, which explains why the total is 656 and not 657

Table 3 Explanations associated with the appropriation period

Technical or organizational problem, or difficulties with system appropriation		Other references of the assessor			
Technical error or inattention	Lack of experience with the CAT	N/A—competency impossible to evaluate	Different interpretation of the wording (issue of vocabulary)	Reference to another assessment system	Reference to norm-referenced interpretation (and not criterion-referenced interpretation)
76	27	5	2	62	40
Total: 212					

Considering the average number of changes per role, the Collaborator is the one that had the most changes ($n = 30.0$), of which 73% were downward changes. The role of Family Medicine Expert, which is almost identical ($n = 29.9$), posts 64% downward changes. This is followed by the Health Advocate role, with changes that are primarily decreasing (70%), and the Professional role, for which all of the changes (100%) are upward. The Leader/Manager and Communicator roles experienced slightly fewer changes, whereas the Scholar role is the one for which a smaller number was recorded ($n = 42$); the changes associated with this role are almost equally distributed between the two categories. In total, 407 upward changes were made and 249 downward changes.

By conducting an inductive qualitative analysis of explanations provided by the assessors who made these changes, it was possible to identify categories of explanations.

Table 4 Explanations expressing disagreement between the system's proposition and the assessor's perception

Disagreement with the system—scores adjusted upwards		Disagreement with the system—scores adjusted downwards	
Overemphasis on assessing weaknesses compared to strengths	Disagreement with the system that "distant supervision" or "limit timing" results in a rating of inadequate or failure	Real or anticipated improvement in the resident's performance	"My judgement is better"
88	79	26	55
Total upward adjustments: 290		Total downward adjustments: 172	
		Resident who reaches the expected performance level, but not early (contrary to the system's interpretation)	Did not reflect the resident's actual performance
		122	29
		Total downward adjustments: 172	
		Criteria or expectations are not adapted to the specialty (other than family medicine) or to the specific nature of the internship	Criteria or expectations are not adapted to the specialty (other than family medicine) or to the specific nature of the internship
		42	21

Categories of explanations provided by the assessors

Two general categories of explanations of the changes in ratings were identified: those associated with a period of appropriation to a new system ($n=212$) and those representing a difference between the system proposal and the perception of the assessor ($n=462$).³ Table 3 presents the subcategories of the first category, while Table 4 presents those of the second one.

The users (in this case, the assessors) dealt with technical or organizational problems, and reported difficulties with system appropriation. Three subcategories of explanations thus emerged from the analyses and are grouped together in this category. For example, the subcategory “lack of experience with the CAT” emerged from statements such as:

- “I’m not used to the assessment scale. I made changes accordingly”;
- “It’s the first time that we’re doing this type of assessment. It’s not immediately clear, but it’s more interesting than the old assessments! We were authorized to use this site this afternoon in order to assess the new residents in our teaching unit.”

Other types of explanations appear under the heading “Other References of the Assessor”. In particular, some assessors made reference to another assessment system:

- “We modified the assessment based on the daily assessments made by the people in charge of the Emergency Department of Hospital Centre X”;
- “The tool developed for the final assessment is superb, BUT the computerized assessment tool to be completed daily does not correspond to this at all. This makes the linking between the two far less than optimal. It is absolutely essential to develop a daily assessment sheet that corresponds to the summative [assessment] sheet so that the final assessment will be more reliable.”

Some assessors referred to norm-referenced interpretation of the results:

- “... Your scale of “distant supervision” and “independent” in this section should be revised; otherwise, assessments in family medicine will always be associated with a particular “level of supervision” column. This does not allow for differentiating the strengths between residents.”
- “The difficulty observed does not lead to a delay compared with other residents of his level...”

The second general category encompasses the subcategories illustrating a difference between the system proposal and the assessor’s perception (Table 4), regarding the competency of the learner.

It can be seen that, out of a total of 462 reasons related to this category, most, i.e., 290 reasons (63%), are associated with upward changes.

³ By adding up the totals, one can see a difference of 17 in the number of changes made (674 instead of 657). This can be explained by the fact that the assessors provided one explanation per sheet and that, in some cases where there were several changes in ratings on the same sheet, there were contradictory explanations (e.g., for a rating revised downwards and another revised upwards on the same sheet). Some statements of explanation were split into more than one statement.

In 88 instances of upward changes, the assessor considered that the system over-emphasized the assessment of weaknesses compared to strengths. Responses such as those below led to this category being identified:

- “did not reflect strengths to the same degree as weaknesses... while there is knowledge that she needs to develop further, or areas where she needs advice or is less comfortable, like many new bosses, she also has outstanding knowledge/skills in other areas.”
- “I am surprised that he received a “limit timing” rating, given that he was quite adequate under supervision. He is independent, in my opinion, for his level.”

In 79 cases, assessors disagreed with the system in that “distant supervision” or “limit timing” results in a rating of inadequate or failure through statements such as the following:

- “the resident who exhibited some fluctuation in his learning stance during feedback. Therefore, I did not want to put him as being totally independent, but this was not a problem that would warrant giving him a rating of “limit timing” or even “delay” as suggested by the form...”
- “The proposed rating of “delay” is overly strict and implied a failure for the internship, which to me does not seem adequate as an assessment.”

In other instances of scores adjusted upwards ($n = 55$), the assessors explained the change by alluding to the superiority of their professional judgement compared to that of the system:

- “... Elements marked as “limit timing” rather than “delay” since the team of supervisors considers that the stage should be given a passing grade”;
- “I found him to be very satisfactory in conducting interviews and carrying out investigations and treatment.”

In addition, the assessors mentioned that the criteria or expectations were not adapted to the specialty (other than family medicine) or to the specific nature of the internship:

- “In our view, these competencies were expected in the context of intensive care expertise.”
- “Because your criteria are not adapted to a 2nd- or 3rd-line internship, it is impossible for the people in charge supervising a resident in the context of an intensive care unit to 1) consider him to be fully independent in all of the required tasks, and 2) even less to assess whether he is capable of being independent in connection with 1st-line patients.”

Table 5 Distribution of changes in ratings according to competencies

Roles and Competencies	Number of Changes										Total
	Upwards						Downwards				
	A	B1	B2	B3	C	D	E	F	G		
	A	R	R	L	R/L	NA	P	A	L		
Leader/Manager											71
Lead1- Engages in appropriate follow-up with patients and their family	2	6	9	-	-	2	-	1	-		20
Lead2- Assume administrative roles and act as a leader	-	-	-	-	-	-	2	-	-		2
Lead3- Manages practice effectively	5			2	-	-	12	4	-		23
Lead4- Keeps good personal-professional balance	-				-	-	11	-	-		11
Lead5- Manages resources effectively	1				-	1	13	-	-		15
Collaborator											60
Coll1- Collaborates with patients and their family	7			1	-	-	22	-	-		30
Coll2- Collaborates with allied health professionals	2	3		3	-	-	22	-	-		30
Health Advocate											50
Prom1- Implements health promotion and disease prevention strategies	0	-	-	7	-	1	19	-	-		27
Prom2- Responds to the needs of vulnerable or marginalized populations	4	-	-	1	1	1	16	-	-		23
Family Medicine Expert											269
Expe1- Collects data - History	5	9	4	15	-	-	-	3	-		36
Expe2- Collects data - Physical exam	4	8	4	22	-	-	-	-	-		38
Expe3- Diagnoses and manages-Generates diagnostic hypotheses	1	2	8	19	-	-	11	9	-		50
Expe4- Collects data -Patient-centered clinical method	-	-	-	6	-	-	6	-	-		12
Expe5- Diagnoses and manages -Procedural skills	-	9	-	2	2	4	6	-	-		23
Expe6 - Diagnoses and manages -investigation and treatment	4	3	1	30	-	-	10	6	2		56
Expe7 - Diagnoses and manages - Patient-centered clinical method	1	-	-	1	-	-	1	-	-		3
Expe8 - Diagnoses and manages - Appropriate clinical judgement	6	-	-	-	-	-	18	4	-		28
Expe9 - Diagnoses and manages -Uncertainty management	1	-	-	-	-	-	22	-	-		23
Scholar											42
Schol1- Documents learning process using a portfolio	-	-	-	-	-	-	-	-	-		1 ^a
Schol3- Engages in self-directed learning	2	1	-	4	-	-	-	-	-		7
Schol4- Uses critical appraisal skills	1	4	-	1	-	1	3	-	1		11
Schol5- Conducts a quality improvement project and/or scholarly project	-	-	-	1	-	2	-	-	-		3
Schol6- Has an evidence-based practice	3	-	-	-	-	-	14	1	1		19
Schol7- Teaches students and colleagues	-	-	-	-	-	-	1	-	-		1
Communicator											48
Comm1- Performs appropriate charting	-	8	7	-	-	-	-	-	-		15
Comm2- Demonstrates listening skills	-	4	3	-	-	-	-	-	-		7
Comm3- Uses verbal communication skills	-	1	2	6	-	-	2	-	-		11
Comm4- Structures the interview	-	2	-	8	-	-	5	-	-		15
Professional											116
Prof1- Adopts professional behaviours in clinical practice	3	22	10	5	-	-	-	-	-		40
Prof2- Adopts professional behaviours in supervision	-	16	13	5	-	1	-	-	-		35
Prof3- Adopts a "learning position"	1	14	7	8	-	-	-	-	-		30
Prof4- Respects the ethical dimensions of clinical decision making	1	4	2	4	-	-	-	-	-		11

Table 5 (continued)

The letters indicate the initial result proposed by the system: *P* early; *A* expected; *L* limit timing; *R* delay; *NA* not assessed

The colours reflect the rating that was changed by the assessor: ■ Early; ■ Expected; ■ Limit timing; ■ Delay; ■ Not assessed

Red font in bold = competencies for which there were 15 or more changes

^a From expected to not assessed

The latter subcategory (“The criteria or expectations are not adapted to the specialty ...”) also includes explanations of the ratings that were revised downwards ($n = 21$). However, it can be observed that 71% of the downward changes made ($n = 122$) are explained by the fact that, contrary to the system’s interpretation, the assessors do not consider the expected performance level to have been achieved early by the resident:

- “In the end, I found that it was overly generous to rate personal–professional balance as being achieved early because of the time that he needs to invest in order to achieve his goals. Thank you.”
- “I consider that this resident merits a rating of “Expected” for all of the criteria, and not the “Early” rating that was suggested to me for certain criteria. She meets the requirements, but I do not believe that she exceeds them.”

An analysis of the distribution of the changes in ratings based on competencies is presented in the section that follows.

Changes per role and competencies

Table 5 presents the distribution of the changes in ratings based on competencies. From this table, it can be seen that two competencies did not undergo any changes: Scholar 2: Ensures continuing professional development, and Professional 5: Engages in reflective practice. However, 12 competencies underwent 15 or more changes (in red in the table): both competencies of the Collaborator role (2/2), both competencies of the Health Advocate role (2/2), six competencies (6/9) of the Family Medicine Expert role, and two competencies (2/5) of the Professional role.

This representation also makes it possible to note that instances of differences regarding early achievement of the performance level can be observed primarily in the roles of Collaborator and Health Advocate, and in two competencies of the role of Family Medicine Expert, as this is primarily where changes are observed from “Early” to “Expected” (column E in Table 5). Disagreements with the system were also observed in that “distant supervision” or “limit timing” results in a rating of inadequate or a failure are mainly in connection with competencies of the roles of Family Medicine Expert and Professional (columns B1, B2, B3 in Table 5).

Discussion

Overall, the results indicate that one sheet out of five (20.1%) was modified, which represents 2.4% of all of the competencies assessed. This low rate of changes seems to imply that most assessors agree with the system proposals. It would be important to validate

this hypothesis through discussion groups with assessors, particularly since modifications were made to one out of five assessment sheets.

As is often the case when a new computerized system is implemented, the users (here the assessors) dealt with technical and organizational problems, and reported difficulties with system appropriation. Such problems generally tend to diminish during the system appropriation period. Other problems that can be associated with an adjustment period are those underlying the explanations about “other references of the assessor”, and these can prove more complex to resolve. Indeed, these mismatches in the reference frameworks (the use of norm-referenced interpretation, for instance) are more time-consuming and painstaking to correct, as they refer to changes in habits and culture. Thus, they represent a significant challenge for the Family Medicine Program.

For the time being, an analysis of the convergence between the program expectations (results proposed by the system) and the assessors’ judgement makes it possible to identify the roles and competencies that lead to the most changes, but also to better understand what motivates these changes and see if there are any improvements to be made, so as to improve user confidence.

Number of changes, per role and per competency

The Expert role is the one that underwent the greatest number of changes, followed by the Collaborator role. The Expert role is often the one that assessors consider to be the most important, since clinical expertise is at the heart of the family physician’s work. Furthermore, different studies and writings on clinical supervision highlight the importance that supervisors give to this role (Côté & Laughrea, 2014; Côté et al., 2018; Ramani & Leinster, 2008). The Collaborator role has become increasingly important in recent years, particularly in light of the research underscoring the importance of collaboration between health science professionals (Careau et al., 2014; D’amour & Oandasan, 2005). It is appropriate to wonder about the connection between the importance given to a particular role and the number of changes made on assessment sheets. Indeed, one might think that assessors who consider a role to be more important will give it more attention, and that they will therefore be more likely to have doubts and then to make changes. This hypothesis could also be validated through discussion groups with assessors.

Table 5 shows that the assessors were less strict than the system for the Professional role, since nearly all of the changes were made upwards: from “delay” to “expected”, or from “limit timing” to “expected”. This occurred whereas, at the same time, the Family Medicine Program considers that competencies associated with this role should be developed before entering the program (during clerkship of previous degree). In our view, the fact that a resident can be identified as presenting a “developmental delay” immediately upon entering the program is difficult to accept for some assessors, particularly since they are not the ones who evaluated the competencies associated with this role. They would then be “bearers of bad news” and would run the risk of undermining the resident’s trust in them. We also wonder whether a connection can be made with barriers identified by Guerrasio et al. (2014), that lead clinical teachers to avoid placing a resident in a situation of failure and “failure to fail”. These authors identified four barriers: 1) a lack of documentation, 2) a lack of knowledge about what specifically needs to

be documented, 3) anticipation of the appeal process, and 4) a lack of options for remediation. It would be important to verify this with the assessors.

We also observed that the assessors were generally stricter than the system (score adjusted downwards) regarding the roles of Collaborator and Health Advocate. The changes for the role of Collaborator are exclusively from “early” to “expected.” This leads us to believe that the assessors perhaps did not read the description in the pop-up window and that, as a result, they intuitively evaluated the residents’ attitude instead of their ability. They would then have evaluated them as being independent too early (which would have caused the system proposal to consider that the competency was achieved “early”), given that a longer residency time is required to develop ability.

Regarding the role of Family Medicine Expert, opinions are more divided. For the first steps in the clinical approach, assessors tend to be less strict than the system (scores adjusted upwards). Here also, it would have been appropriate to check whether a connection can be made with “*Failure to fail*” (Guerrasio et al., 2014). However, when the time comes to “show appropriate clinical judgement” (Expe8) or to “manage uncertainty” (Expe9), they are stricter than the system, and they tend instead to adjust the scores downwards. It should be noted that these two competencies stand out because of their complexity and, for this very fact, because of the complexity of their assessment.

Categories of explanations provided by assessors

The categories of explanations provided by assessors help to better understand the reasons that motivate them to make changes to the results proposed by the system. An analysis of these explanations will result in the persons in charge of the program reviewing some of the expected timelines for achieving an independent entrustment level during the training process. It could also lead them to improve training activities aimed at helping assessors to better understand the nature and functioning of the new system, and to better standardize procedures. For example, such training activities could be an opportunity to draw a parallel between the criterion-referenced assessment tool (CAT) and other assessment systems mentioned in the explanations. Differences in the nature of the wordings used by assessors were observed according to the membership group. For example, some assessors referred to the daily assessments in emergency medicine, which are different from those proposed by the CAT. The latter is currently being modified by adjusting the daily feedback form.

We also observed that some assessors always refer to norm-referenced interpretation of results, whereas the system that was implemented is based specifically on criterion-referenced interpretation (precisely in order to avoid norm-referenced interpretation). A more in-depth analysis of these explanations over time would enable us to ascertain whether these mismatches in the reference frameworks are gradually diminishing as the culture shifts and the system is appropriated. The discussion groups held with the assessors also allowed us to verify whether specialist physicians (other than family physicians) tend to evaluate the level of supervision through the eyes of their specialty and by making norm-referenced comparisons with their own residents. If this is indeed the case, this could explain why, for example, they are less likely to consider residents as independent for physical examinations since, in their specialty, examinations involve subtleties acquired later by their own residents.

The dual purpose of the system as a whole also raises questions. On the one hand, it invites assessors to conduct formative evaluations by providing residents with possible ways to make improvements (educational prescriptions). On the other hand, assessors must determine, during the summative evaluation, whether the residents demonstrate, at different stages of their training, the expected competencies for practising family medicine. This is an onerous educational responsibility that, in actuality, is more complex than it appears. In fact, many assessors have modified the rating given that, initially (by selecting the level of supervision), they wanted to provide residents with suggestions for improvement, but the results proposed by the system led them to realize that this had resulted in the system considering them to present a “developmental delay” or be “in difficulty,” which the assessors did not want. Dividing the system into two components with different purposes—summative and formative evaluations—might be a solution to this problem.

Emotional reasons can also motivate assessors to change ratings. The emerging category entitled “my judgement is better” brings together 55 explanations along these lines. In fact, some assessors have underscored how uncomfortable they feel about the computerized system “makes decisions in their place”, whereas, in their opinion, it is the assessor who should have the best judgement. This unease could probably be dissipated by training activities, which would highlight that the system’s reasoning is based on the reasoning of the many clinical teachers who participated in a Delphi study (Lacasse et al., 2014). In such training sessions, it would be essential to emphasize to the participants, on different occasions, that the assessor has the predominant role with respect to the final result. The participants would then better understand that their judgement is central to the assessment. In accordance with the criteria for successful integration of technologies proposed by Bates and Sangra (2011), these training sessions could provide an opportunity to identify ambassadors who recognize the importance and utility of the system, and who are comfortable using it. The latter could become resource persons in their training environment and could ensure that a good level of trust in this knowledge base is maintained, which is very important (Shibl et al., 2013).

Finally, we observed that several of the changes explained by technical errors or lack of attention had been upward changes. We wondered if these explanations were not in fact disguised disagreements with the system. Indeed, it is quicker and less confrontational to use technical errors as an explanation, rather than openly setting out their disagreement with the system’s proposal. The “MUM effect” (Scarff et al., 2019), which refers to the difficulties that clinician assessors have in sharing so-called negative feedback with residents, could also explain some of these upward changes.

Conclusion

One of the challenges of training programs aimed at the achievement of competencies is the longitudinal tracking of this competency development. The analysis of a computerized system in family medicine has shown that it could facilitate such tracking. We note that the development of a credible computerized system requires a rigorous approach designed to properly identify the program expectations and targeted competencies, as well as the benchmarks for tracking their achievement. Furthermore, as emphasized by Bates (2018), it is essential to involve the users in developing such a system. In addition,

it is critical for the system's implementation to be accompanied by user training, and also by a willingness to have discussions that will facilitate the appropriation period and reduce resistance to system use. This central involvement of the user is at the heart of user-centered design practices as described by Abras et al. (2004). This research provides a better understanding of the reasons behind the rating changes. This has allowed the program to better orient its actions and to organize training and information sessions to facilitate the appropriation to the new computerized system. Moreover, the most recent probes conducted in the program between 2018 et 2020 demonstrate that the vast majority of proposed ratings or overall assessment results proposed by the system are retained by supervisor, whether family physicians or other specialist.

In conclusion, further research is needed on the use of the Advisor system, to ascertain whether it helps to improve the quality of the educational diagnoses and prescriptions (Simard et al., 2021). It would also be necessary to determine the impact of using such a system on the competency level of residents at the time when they complete the program and to conduct design-based research to better inform participatory instructional design practices.

Acknowledgements

The authors thank Mary Eady for the linguistic revision of the manuscript.

Authors' contributions

IS wrote the article in collaboration with LC and AK. LC, CS and CR participated in the data collection and analysis for the development and validation of the computerized system and the revision of the manuscript. AK, ML, IS et CR participated in the design and development of the computerized system. ML was the leader of the overall project funded by The Medical Council of Canada. All authors read and approved the final manuscript.

Funding

Medical Council of Canada.

Availability of data and materials

Not applicable.

Declaration

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Education, T luq University, 455 Du Parvis, Quebec, QC G1K-9H6, Canada. ²Department of Family Medicine and Emergency Medicine, Faculty of Medicine, Laval University, Quebec, Canada. ³Faculty of Medicine, Laval University, Quebec, Canada. ⁴QMA-CMA-MD Educational Leadership Chair in Health Professions Education, Faculty of Medicine, Laval University, Quebec, Canada.

Received: 21 February 2021 Accepted: 25 November 2021

Published online: 10 January 2022

References

- Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction*. Thousand Oaks: Sage Publications, 37(4), 445–456.
- Basque, J. (2017). *L'approche-programme-Les multiples connaissances mobilis es dans un projet d'approche-programme en enseignement sup rieur*.
- Bates, A. W. T., & Sangra, A. (2011). *Managing Technology in Higher Education : Strategies for Transforming Teaching and Learning*. John Wiley & Sons.
- Bates, A. T. (2018). *Teaching in a digital age : Guidelines for designing teaching and learning*.
- Careau, E., Bri re, N., Houle, N., Dumont, S., Maziade, J., Par , L., Desaulniers, M., & Museux, A. C. (2014). Continuum des pratiques de collaboration interprofessionnelle en sant  et services sociaux : Guide explicatif. *R seau de collaboration sur les pratiques interprofessionnelles en sant  et services sociaux (RCPI)*.
- Carraccio, C., Wolfsthal, S. D., Englander, R., Ferentz, K., & Martin, C. (2002). Shifting paradigms: From Flexner to competencies. *Academic Medicine*, 77(5), 361–367.
- Chou, S., Lockyer, J., Cole, G., & McLaughlin, K. (2009). Assessing postgraduate trainees in Canada: Are we achieving diversity in methods? *Medical Teacher*, 31(2), e58–e63.

- Côté, L., Laurin, S., & Sanche, G. (2018). Échanger de la rétroaction avec les étudiants. Dans *Comment mieux superviser les étudiants en sciences de la santé dans leurs stages et leurs activités de recherche* (DeBoeck Supérieur, pp 81–109).
- Côté, L., & Laughrea, P.-A. (2014). Preceptors' understanding and use of role modeling to develop the CanMEDS competencies in residents. *Academic Medicine*, 89(6), 934–939.
- D'amour, D., & Oandasan, I. (2005). Interprofessionalism as the field of interprofessional practice and interprofessional education: An emerging concept. *Journal of Interprofessional Care*, 19(1), 8–20.
- Guerrasio, J., Furfari, K. A., Rosenthal, L. D., Nogar, C. L., Wray, K. W., & Aagaard, E. M. (2014). Failure to fail: The institutional perspective. *Medical Teacher*, 36(9), 799–803.
- Houdé, O., Kayser, D., Koenig, O., Proust, J., & Rastier, F. (2003). Vocabulaire de sciences cognitives (neuroscience, psychologie, intelligence artificielle, linguistique et philosophie). *Psychologie et sciences de la pensée*.
- Lacasse, M. (2009). *Diagnostic et prise en charge des situations d'apprentissage problématiques en éducation médicale*. Université Laval, Faculté de médecine, Département de médecine familiale et de médecine d'urgence.
- Lacasse, M., Audétat, M.-C., Boileau, É., Caire Fon, N., Dufour, M.-H., Lafferrière, M.-C., Lafleur, A., La Rue, È., Lee, S., Nendaz, M., Paquette-Raynard, E., Simard, C., Steinert, Y., & Théorêt, J. (2019). *Remediation interventions for undergraduate and postgraduate medical learners with academic difficulties : A BEME systematic review*. BEME Best Evidence Medical Education.
- Lacasse, M., Rheault, C., Tremblay, I., Renaud, J.-S., Coché, F., St-Pierre, A., Théorêt, J., Tessier, S., Arsenault, L., & Simard, M.-L. (2017). Développement, validation et implantation d'un outil novateur critérié d'évaluation de la progression des compétences des résidents en médecine familiale. *Pédagogie Médicale*, 18(2), 83–100.
- Lacasse, M., Théorêt, J., Tessier, S., & Arsenault, L. (2014). Expectations of clinical teachers and faculty regarding development of the CanMEDS-family medicine competencies: Laval developmental benchmarks scale for family medicine residency training. *Teaching and Learning in Medicine*, 26(3), 244–251.
- Prégent, R., Bernard, H., & Kozanitis, A. (2009). *Enseigner à l'université dans une approche-programme : Guide à l'intention des nouveaux professeurs et chargés de cours*. Presses inter Polytechnique.
- Ramani, S., & Leinster, S. (2008). AMEE Guide no. 34 : Teaching in the clinical environment. *Medical teacher*, 30(4), 347–364.
- Renaud, J.-S., Lacasse, M., Cote, L., Théorêt, J., Rheault, C., & Simard, C. (2020). *Psychometric validation of the Laval Developmental Benchmarks Scale for Family Medicine*. <https://doi.org/10.21203/rs.3.rs-51118/v1>
- Savard, I. (2014). *Modélisation des connaissances pour un design pédagogique intégrant les variables culturelles*. Télé-université.
- Scarff, C. E., Bearman, M., Chiavaroli, N., & Trumble, S. (2019). Keeping mum in clinical supervision: Private thoughts and public judgements. *Medical Education*, 53(2), 133–142.
- Shaw, E., Onadasan, I., & Fowler, N. (2017). *CanMEDS-Médecine familiale 2017 : Un Référentiel de compétences pour les médecins de famille dans tout le continuum de formation*. Collège des médecins de famille du Canada. <https://www.cfpc.ca/ProjectAssets/Templates/Resource.aspx?id=3031&langType=3084>
- Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22(2), 63–75.
- Shibl, R., Lawley, M., & Debuse, J. (2013). Factors influencing decision support system acceptance. *Decision Support Systems*, 54(2), 953–961. <https://doi.org/10.1016/j.dss.2012.09.018>
- Simard, M.-L., Lacasse, M., Simard, C., Renaud, J.-S., Rheault, C., Tremblay, I., & Côté, L. (2017). Validation d'un outil critérié d'évaluation des compétences des résidents en médecine familiale : Étude qualitative du processus de réponse. *Pédagogie Médicale*, 18(1).
- Simard, C., Côté, L., de Bruyn, L., & Lacasse, M. (2021). CanMEDS Competencies in Family Medicine Residents : Can Criterion-Based Assessment Improve the Quality of Teacher Feedback? *MedEdPublish*. <https://doi.org/10.15694/mep.2021.000016.1>
- Tardif, J. (2006). *L'évaluation des compétences. Documenter le parcours de développement*. Chenelière Éducation.
- Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2), 237–246.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.