UNIVERSITY OF CALGARY

The Development of a Reliable and Valid Instrument to Evaluate Consultation Letters in Internal Medicine

by

Martha Ainslie

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MEDICAL SCIENCE

CALGARY, ALBERTA

SEPTEMBER 2003

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "The Development of a Reliable and Valid Instrument to Evaluate Consultation Letters in Internal Medicine" submitted by Martha Ainslie in partial fulfillment of the requirements for the degree of Master of Science.

Dr P Harasym

Dr J-G DesCoteaux

Dr. E Keely

Dr R Cowie

Dr K Brownell

September 9, 2003

Abstract

The purpose of this master's thesis was to develop and test a tool for the evaluation of consultation letters in Internal Medicine.

The development of a consultation letter assessment tool (C.L.A.T.) followed several steps. The current evaluation tools were reviewed. The construct's attributes were defined. Items were chosen and weighed in order to assure content validity.

The reliability was determined in a pilot study. The initial CLAT had a Cronbach's coefficient of 0.91. The tool was analyzed to look for items that detracted from its reliability and was modified.

The validity of the modified CLAT was studied in 2 settings: the oral examination and the observed structured clinical examination (OSCE). Seventeen residents and six faculty members dictated a consultation letter as part of an OSCE. There was a significant difference in total score of the faculty compared to the residents. The faculty scored higher in the communication and education sections and on the global score. These findings support the construct validity of the CLAT.

The concurrent validity of the instrument was assessed by comparing the results from an oral examination to the CLAT scores within the same clinical case. There was no correlation between the total score on the oral exam and the CLAT. There was modest correlation for data gathering and management. The residents performed better on the oral examination than on the CLAT suggesting that factors other than knowledge may be important in what a resident chooses to include in a consultation letter. Thus the CLAT has very little concurrent validity to a traditional oral examination.

Overall, the modified CLAT scores are reliable and there was evidence of construct validity but not concurrent validity. The current version of the CLAT should be used informally in the outpatient setting to provide learner feedback. With further refinement, it could have a role to play in the formal evaluation of Internal Medicine trainees.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

When a medical student graduates from medical school, he/she is only part way to being a physician. During residency the undifferentiated medical student's knowledge, skills and attributes are further transformed to become a family practice physician or a consultant in a specific field of medicine. It is the mandate of the Royal College of Physicians and Surgeons of Canada (RCPSC) to oversee both the training and the certification of the medical and surgical specialties in Canada. As part of their mandate, the RCPSC has developed a framework of the attributes that a specialist must master in order to practice effectively in this millennium. The CanMeds 2000[1] document outlines 7 key attributes of a specialist. These attributes are medical expert, communicator, collaborator, health advocate, manager, scholar and professional.

Traditionally postgraduate programmes have focused heavily on the attribute of the medical expert with little or minimal training in communication. To this extent the postgraduate programmes lag significantly behind the undergraduate programmes where communication skills with patients and families are now taught and evaluated. However, this is changing. In 2002, the RCPSC examination in Internal Medicine had communication stations where the candidate's ability to communicate with patients, families or other physicians was tested.

At both the undergraduate and postgraduate levels, the teaching in communication skills has focused mainly on patient-physician interactions. The area of physician-physician communication has been the least taught, the least studied and the least tested form of communication despite its obvious importance[2;3]. A survey of 300 physicians in Eastern Ontario in 1996-97 showed that only 16% of specialists received formal

instruction in consultation letter writing in their training[3]. This same survey showed that only 49% of the specialists had received feedback on their dictated letters during their training.

Communication between the consultant and the referring medical doctor (MD) plays a pivotal role in the consultation process. The importance of communication between physicians has been recognised by the RCPSC in the CanMeds 2000 document[1]. As a medical expert/clinical decision maker the physician must demonstrate effective consultation skills including presenting "well documented assessment and recommendations in both written and verbal response to requests from other health care providers (p.4)" [1]. The specialist's role as a communicator requires the ability to discuss "appropriate information with other health care providers that facilitates optimal health care of patients (p.5)" [1]. This includes the ability to "maintain clear, accurate and appropriate records (p.5)"[1]. Finally in his/her role as a collaborator the specialist is expected to "effectively consult with other physicians and health care providers (p.6)"[1].

Communication from the specialist to the referring MD most often occurs through the consultation letter especially in the outpatient setting [3,5]. This mode of communication offers a flexible medium that can be adopted to cover referrals ranging from straightforward technical problems to complicated cases in which complex treatment and management regimes are implemented.

Deficiencies in the consultation letter may have an impact on patient care. A lack of effective communication between physicians may result in patient uncertainty and dissatisfaction[4-6]. Continuing care may not be effectively co-ordinated and tests may be duplicated. Poor communication has implications on the quality of care that "could cause

potential or actual inconvenience, discomfort or harm to patients (p.14)"[7]. The referring MD may become dissatisfied and change his referral practices.

In order to have a valid assessment of clinical performance it is important to measure skills and behaviours that are used in day-to-day practice. The evaluation of the consultation letter offers an excellent opportunity to assess an important aspect of the clinical practice of a specialist[8].Timely feedback is especially important during training. As technology changes, more patient care and teaching is occurring in the ambulatory setting. This setting is characterised by brief resident and preceptor interactions with little formal teaching[9]. There is even less feedback. In order to maximise patient flow and clinic efficiency there is reduced time for teaching. A systematic critique of a consultation letter offers an efficient way of giving feedback without affecting the flow of patients in the ambulatory care setting and can highlight knowledge gaps of the trainee.

Unfortunately, the evaluation of any construct is a difficult process. A construct is an unobservable, postulated variable, which is proposed to summarise and explain facts, data and behaviour. The RCPSC, professional bodies and postgraduate programmes are interested in evaluating the construct of 'a medical consultant'. The evaluation of a consultation letter is part of the evaluation of this construct.

Despite the acknowledged importance of the consultation letters as a form of communication between doctors, to date there are only a few reliable and valid generic tools [6;10]with which to evaluate it (Appendix A and B).

Unfortunately, there is no gold standard for the consultation letter. Crossley et al defined the gold standard as a letter that: "clearly conveys the information I would like to have about the patient if I were the next doctor to see him/her"[6]. This generic statement is

not helpful when trying to evaluate a letter as it does not clearly define the qualities that are needed in the letter. In order to assess the consultation letter there must be agreement on the qualities of a good consultation letter from the points of view of multiple users. The assessment instrument must evaluate these attributes. The instrument needs to be easy to use in both a formal evaluation setting as well as on a daily basis in the clinic setting. It must be versatile enough to be used in a variety of clinical settings ranging from single patient problems to a complex multisystem disease with multiple medical issues. It must be reliable and valid. It must offer precise feedback so as the trainee can improve their consultation skills.

Reliability is the extent to which the instrument measures consistently (what ever it measures). A reliable evaluation instrument should give the same result for a given letter when different examiners use it (inter-rater reliability). In addition, the instrument should give the same result when a single examiner uses it on a single letter at multiple points in time (intra-rater reliability). The instrument should have internal consistency. All these forms of reliability can be measured.

The instrument must also be valid. A valid instrument is one that measures what we think it measures. Validity refers to the correctness of the inference made from the performance of the measure. Reliability is a prerequisite for validity. The other components of validity include face validity, content validity, criterion-related validity and construct validity.

Face validity has to do with appearances. Does the instrument appear to measure what it is supposed to measure from the point of view of the examinee? From the trainees' point of view, does the instrument look fair? Is the instrument evaluating trivia

or a key component of clinical practice? As the consultation letter is a daily reality of the practice of a specialist one can argue that de facto the face validity is high if the instrument samples the key components of the letter (content validity).

Content validity refers to the extent to which the instrument samples the domain of measurement, in this case the consultation letter. In order to ensure content validity all the important attributes of the construct must be measured in the correct proportion that they contribute to the construct. Content validity cannot be calculated. Content validity can be assessed by analysing the degree to which the attributes of the construct are reflected in the instrument.

Criterion related validity has 2 subcategories: predictive validity and concurrent validity. Predictive validity refers to the extent to which performance measured with the current instrument can predict subsequent performances. Predictive validity of an instrument evaluating consultation letters would be expected to be poor because of the case specificity of medical knowledge (i.e. how well a physician manages headaches is not predictive of his/her ability to manage asthma). However a valid consultation letter may well demonstrate concurrent validity. That is to say, the ability to write an excellent consultation letter on asthma may well be correlated with an oral examination score on the same topic.

Construct validity refers to the degree to which the instrument evaluates the abstract psychological trait that the individual possesses. In the case of the consultation letter the construct refers to the ability of the individual to communicate as a consultant to another physician. The assessment of construct validity involves both a logical and empirical analysis. The construct needs to be analysed in order to identify its key

attributes. The evaluation instrument should contain questions that are relevant to the construct. The empirical analysis of construct validity tests predictions regarding the relationship between the construct and other empirical measures. Individuals that are felt to possess a greater degree of the construct (i.e. respected consultants) should score higher than individuals with less ability (i.e. first year residents). The ability to write a consultation letter should correlate with other markers of consulting skills. Every bit of information about a test has relevance for construct validity. There is no single criterion to establish construct validity. Content validity, face validity and criterion validity contribute to the analysis of the construct validity.

The purpose of this master's thesis is to develop a reliable and valid tool with which to evaluate the consultation letter in internal medicine. In order for such a tool to be developed the literature was reviewed to ascertain the role of the consultation letter, the current state of the consultation letter, and the desired attributes of the letter with respect to format and content. The literature examining the differences in information processing between beginners and experts was reviewed to see how this might affect the consultation letter. The current tools used in the evaluation of the consultation letter were reviewed in order to build on their success.

II. REVIEW OF THE LITERATURE

1. The Role of the Consultation Letter in Clinical Practice

The consultation letter can be viewed as the net result of a two-step process consisting of the acquisition of the information by the consultant and the processing of this information. Data unique to each patient encounter is acquired from three main sources. The first source is the referral letter from the physician explaining the reason for the consultation. This most often is in the form of a request for assistance (i.e.: please help manage this person's chronic obstructive lung disease (COPD)) or as a question (i.e.: What asthma drugs are safe to use in this patient while she is pregnant?). The second, and most important source of information, comes from the clinical encounter between the consultant and the patient[11]. The amount of information obtained is in part a function of both the physician's and patient's ability to communicate. The medical skills of the consultant, including the ability to ask key questions and to detect pertinent negative and positive physical examination findings are crucial to this process. A final source of information comes from ancillary tests (including X-rays, blood work, and functional testing). The value of these tests depends on the skill of the consultant to interpret the information in the light of the patient's problems.

The second step in the generation of the letter is the processing of the information by the consultant and the conversion of this information into the consultation letter. Several factors influence how the information is processed. These include the consultant's degree of understanding of the medical problem, the perceived reason by the consultant for the consultation, the role the letter plays as part of the consultant's medical records, the consultant's ability to communicate effectively, and time constraints.

In order to develop an instrument with which to evaluate the consultation letter it is important to define the attributes of the letter as precisely as possible. The components of the consultation letter are determined by the function that the serves[3]. Not surprisingly the letter has several functions. It serves as a medical record for the consultant documenting the pertinent historical, physical, and laboratory findings. It serves as a blueprint for ongoing patient management by outlining investigation and treatment plans. The letter is also seen as a way to provide continuing medical education (CME) to the family physicians. The letter is used as an aide memoir by the consultant during follow-up visits. For the family doctor, the letter provides feedback on his/her initial clinical impressions. It is a guide for future management of the patient's problems. It is also valued as an important educational tool. The letter is part of the patient's health care records and as such is used in court cases and in disciplinary hearings by the provincial colleges. Finally, the consultation letter may serve as an important education tool for the patient. There are currently several pilot projects in the U.K. looking at sending a copy of all consultation letters to the patients[12]. To date this initiative has been highly received by the public with a survey indicating that 98% of patients favoured receiving the letters.

2. The Current State of the Consultation Letter

Despite the acknowledged importance of the consult letter, several studies have shown large deficiencies in this area. McPhee[5] studied 373 referrals made by primary care internists to specialists in San Francisco. The referring MD was not made aware of the results of the consultation in 45% of the cases. This occurred predominantly with surgical specialties. No response was received from 45/65 referrals to ophthalmology, 14/23 referrals to obstetrics and gynaecology, 12/21 referrals to orthopaedics and 25/48

referrals to dermatology. This response rate is lower than previously reported. In Metcalfe's study[13], referring MD received a consultation report in 81% of referrals. Williams[14] found that 2/3 of physicians at a University Medical clinic sent a summary letter to referring doctors.

Westerman et al reviewed the quality of 144 referral and consult letters at the Free University in Amsterdam[15]. Four general practitioners (GP) and 4 specialists scored the letters. The consult letters were evaluated on the documentation of objective and subjective findings, assessment and management plans, documentation of the patient's psychosocial background, educational value, tone of the letter and whether the referring physicians questions were answered. Overall 78% of the specialists' letters were judged to be good or excellent. However the reason for referral was felt to be answered 'very well' in only 50-60% of the letters. Only 20% of the letters indicated very good awareness of the patient's psychosocial background. The assessment was deemed to be very poor to moderate in 30% of the cases. The medical plan for the patient was rated very poor to moderate in 22% of cases. With respect to the educational value of the letter, the communication was felt to have taught the GPs 'nothing' or ' a little' in 45% of the cases.

Pullen reported on a survey of both referral letters and consultation letters in psychiatry in the Edinburgh area[2]. A questionnaire was sent to 40 GPs and 40 trainees in family practice. They were asked to list the 5 most important items that they considered a psychiatrist should include in a report on one of their patients. Seventy questionnaires were returned. The 5 most important items were diagnosis, recommended treatment, follow up arrangements, prognosis, and a concise explanation of the condition.

Subsequently 60 consecutive consultation letters from January 1973 and again from 1983 were assessed on the 5 key items using a dichotomous scale. In 1983, 88% of the letters had a diagnosis, 92% suggested treatment, and 95% discussed follow up. However only 60% of the letters had a concise explanation of the symptoms and 27% discussed prognosis. The style of the letter was noted to have changed over the 10 years. Subheadings were used in 10% of the letters in 1973 compared to 35% of the time in 1983. Registrars' letters were twice as long as the consultants' although the number of key items present in the letters was the same in both groups.

Kentish[16] also looked at the quality of the psychiatry consultation letter. Seventy-nine GPs were sent a survey about the importance of certain items in a consultation letters and on their satisfaction with the consultation letter. Fifty-one responded. Of the 51 respondents, 12 felt that they always received a useful report. Thirty-three said they usually did whereas 6 said they sometimes received a useful report. Their main complaints were that they were rarely told what information the parents of the patients had been given or advised to their own role in relation to the family.

Tattersall et al examined the quality of the consultation letter in oncology[17]. This was part of a study looking at the use of audiotapes versus letters sent to the patients as part of the consultation process in oncology. One hundred and eighty-two patients were included in the trial. The oncologist wrote a traditional letter to the referring physicians and an individualised letter documenting the important aspects of the consultation to the patients. A total of 303 physicians received letters. Fifty-five GPs and 53 referring specialists were selected from the 303 physicians and surveyed about the preferred information content of letters from consultant physicians. The survey consisted of 14

content categories. The physicians were asked if the item was essential, useful, of little use and no use. Of the 14 items, 9 items were scored as essential by more than 50% of the responders. These items were: prognosis, benefits of treatment, side effects, diagnosis, further tests/investigations, test results, clinical findings, treatment options/recommendations, arrangements for follow up and what the patient had been told.

Subsequently 94 of the letters from the original cohort of 182 patients were analysed (Table 1). A significant discrepancy in what doctors wanted and what they received in the letters was noted in information categories pertaining to prognosis, benefits of treatment, side effects, further testing, and what the patient had been told. The letters to the patients contained more information about prognosis, side effects of treatment, and explanation of the symptoms than the letters to the physicians.

McConnell [18] evaluated 99 consultation letters from 6 medical oncologists from 2 hospitals in Sydney, Australia. Each letter was evaluated on items that a prior survey of referring doctors (surgeons and family practitioners) had highlighted as important information to be conveyed in the consultation letter. Their analysis revealed a significant discrepancy between what the referring physicians wanted and received (Table1). Their survey of the referring physicians about the key content items in a letter highlighted different expectations from different user groups. For example, while the surgeons were not interested in the history of the presenting problem the family doctors were. Likewise the family doctor was much more interested in side effects and treatment of the side effects than was the surgeon. This most likely reflects the fact that the patient follow up is normally with the family physicians.

Babington[4] et al looked at consultation letters of 204 consecutive referrals to the radiation oncology service at the Westmead Hospital in New South Wales, Australia. The letters were analysed using a checklist based on the prompt sheet developed by McConnell[18]. Twenty eight percent of new referrals did not have a letter written following their initial oncology assessment. The family doctor received a letter in only 58% of the total study cohort. Their analysis of the letters showed deficiencies in the information conveyed (Table1).

Myers et al[10] were the first to report on consultation letters dictated by trainees in core internal medicine and medical subspecialties. They reviewed 97 letters dictated by internal medicine and internal medicine subspecialty residents. They found that 1/3 of the letters did not contain relevant collateral history, ¼ of the letters had no rationale for the management plan and 1/5 gave no indication of who was responsible for follow-up. The average score on the overall quality of the letters was 2.75 (± 0.82) on a global rating scale ranging from 1 to 5. Though the residents included most of the important historical and physical examination content in their letters they did so at the expense of clarity and consistency.

Thus the literature on the quality of the consultation letter shows some common themes. Family physicians do not always receive the correspondence[4;5]. The letters contain excessive information on the history of presenting illness and past medical history from the point of view of the family physician[17;18]. However, if the letter is viewed as the sole document in the consultant's chart outlining the consultation findings, the letter may not contain all the key historical details[10].

This apparent contradiction speaks to the different uses for the letter. The letters do not contain enough information on prognosis, treatment including side effects of treatment, follow up details and communication with the patient[2;16-18]. Their educational value is felt to be poor[15].

There is evidence that consultants may not be aware of their deficiencies in communication. The Canadian task force that examined the relationship between family doctors and specialists found that only 7 % of specialists thought that inadequacy of communication between physicians was a problem[7].

Table 1: What Surgeons and GPs Want in Most/All Cases and What They Get in Reply Letters in Oncology [4;17;18]

	% of doctors that felt item was needed in most/all cases			% of letters containing item		
Content item	% of surgeon [18]	% of GPs [18]	% of GPs[17]	Tattersall[17]	McConnell[18]	Babington[4]
History of presenting problem	42.6	73.6	40	76	97	58
Past medical history	24.6	37.9	20	93	82.8	51
Reason for referral	31.5	75.7	35	0	6.1	32
Patient's understanding	64.9	84.3	52	0	6.1	20
Social problems	59.4	83.1	Not assessed	Not assessed	1	23
Patient compliance	68	88.8	Not assessed	Not assessed	2	Not assessed
Patient's wishes	66.7	86.1	Not assessed	Not assessed	26.3	25
Prognosis	81.5	95.4	71	39	31.3	20
Patient's coping	68.5	87.9	Not assessed	Not assessed	16.2	6
Tests/findings on investigations	92.6	98.1	68	77	41.4	74
Side effects of treatment	58.4	93.4	65	14	16.2	45
Management of side effects	43.6	91.5	Not assessed	Not assessed	5.1	2
Indicators for unscheduled review by oncologist	52.8	85.8	Not assessed	Not assessed	8.1	10
Aim/benefits of treatment	81.1	97.2	69	0	40.4	59
Involvement of other doctors	80.8	89.7	Not assessed	Not assessed	32.3	46
Arrangement made for treatment	77.4	85.9	Not assessed	Not assessed	33.3	89
Treatment options	84.9	94.4	91	84	31.3	85
What oncologist wants GP to do	92.4	99.1	Not assessed	Not assessed	14.1	Not assessed

3. Cognitive Psychology and the Consultation Letter

Knowledge structure changes with the acquisition of expertise. This in turn may influence what a clinician writes in the consultation letter. A novice initially has minimal, if any, knowledge. With instruction he/she acquires dispersed knowledge[19;20]. This knowledge is characterised by an elaborate and complex set of networks bridging one piece of information to another. With the acquisition of expertise, the knowledge becomes compiled. The expert develops a simplified causal model that explains many of the findings under diagnostic labels. The difference between the expert and non expert in terms of dispersed versus compiled knowledge is well demonstrated visually in the minimal elaborated networks (MEN) of novice versus expert fighter pilots rating a set of terms having to do with 'split plane concepts'[21]. The novices' MEN have multiple connecting links whereas the experts' MEN are simplified. This difference in the knowledge structure is also seen in internal medicine. It can be demonstrated by comparing the explanation of a patient presentation of a clerk compared to a specialist.

Problem solving strategies also changes with the acquisition of expertise. Expertise is associated with both pattern recognition [19;20]and forward thinking though this latter point is debated in the literature[22]. In routine cases the expert uses pattern recognition [23]. He/she is rapidly able to match the current patient to an internal representation or mental model of various diseases. These are the instances and illnesses scripts described by Bordage and Schmidt[19;20]. Illness scripts contain enabling conditions, faults (the disease) and the consequences for various diseases. The illness script represents an abstract blend of all the patients that the physician has seen with a given

disease whereas an instance script pertains to a particular patient. These scripts are idiosyncratic[19]. Different clinicians will use different scripts to arrive at a diagnosis. An individual clinician may also use different scripts at different times with patients presenting with an identical problem. Thus evaluation tools that are disease specific run the risk of penalising the expert if the key data features of the evaluation tool are not the same as the one used by the clinician[24]. This suggests that a generic tool may be more appropriate especially when assessing data collection.

With more diagnostically challenging cases the expert uses forward thinking. In this process the expert clumps diseases into clusters based on similarities [25]. The expert physician then uses a series of decision rules to reduce quickly the possibilities to a small family of diseases. At this point the expert is able to pattern recognise or compare and contrast the remaining possible diseases (hypothetical deductive reasoning) to arrive at the most likely diagnosis.

By contrast the novice is unable to pattern recognise successfully. The predominant method of problem solving for the novice is backwards reasoning. The novice thinks of all the possible causes for the problems, and goes back and forward from the patient to the problem list trying to eliminate or include various possibilities. The novice uses hypothetical deductive reasoning at the start. He/she is a risk of failing to make a diagnosis if the correct diagnosis is not in the initial list of possibilities. This differs from forward thinking where the expert uses rules (i.e. inductive thinking) to include or exclude large groups of diseases. Backward reasoning is exhaustive and inefficient whereas forward thinking is frugal and efficient.

How do these changes in knowledge structure and problem solving strategies effect what the consultant chooses to include in the letter? The differences between experts and non- experts may be seen in both the documentation of the history as well as in the assessment component of the letter. If the expert used pattern recognition to arrive at a diagnosis he/she may not have uncovered all the diagnostic clues in the history to arrive at the diagnosis. The same thing may also occur if the expert uses forward thinking. For example if the clinician approaches hemoptysis based on a scheme that uses chest x-rays, the history would not necessarily include questions about causes of hemoptysis that are associated with abnormal chest x-rays if the patient had a normal x-ray. Disease specific evaluation tools run the risk of penalising the expert if the key data features of the evaluation tool are not the same as the one used by the clinician at that time. This suggests that a generic tool may be more appropriate especially when assessing data collection.

This lack of an all inclusive data collection does not represent substandard care but rather is a marker of expertise and efficiency. Unfortunately to date none of the 2 generic tools has evaluated their instrument on experts[6;10]. However, there is indirect evidence from the literature that experts are briefer and less inclusive in their consultation letters. Boudreau et al [26] reported on the evaluation of consultation skills in respiratory medicine using an OSCE. The overall score of the consultants was statistically higher than that of the respirology fellows and the internal medicine residents. However, the residents completed their consultation letters with more database items than did the attending physicians. Data collection contributed to 21% of the overall score for the internal medicine residents whereas it contributed to 16% of the overall score for the

attending respirologists. The article implied that the respirologists scored higher than the residents did on the other aspects of the consultation letter including diagnosis and management issues.

In McCain's et al study[8] on the feasibility of the using the consultation letter as part of in-training evaluation, the Rheumatology consultant scored lower on the data base assessment category compared to the intern on the GI (gastrointestinal) service Despite this, the overall score of the consultant's letter was higher than any of the trainees' scores.

Thus any scoring system that has a heavy emphasis on data collection will penalise the expert and reward the novice. A concise letter may suggest expertise and scoring schemes should take this into account.

The problem solving strategies of the author of the consultation letter may also play a role in what is included in the assessment component of the letter. If the expert used pattern recognition, he/she may not have a differential diagnosis. Ridderikoff[23] analysed videotaped interviews of a standardised patient using practising general physicians. In 21% of the cases there was no differential diagnosis. The clinicians felt there was only 1 diagnosis possible, which Ridderikoff interpreted as the use of pattern recognition. Thus, any tool that assesses the differential diagnosis may penalise the experts in routine cases where pattern recognition is used as the problem solving strategy.

In summary, from our understanding of the changes in knowledge structure and problem solving with increasing expertise, we would expect the expert's letter to be briefer, more concise, have a shorter differential diagnosis and may contain more

semantic axes. Unfortunately, there is very little in the literature to support or refute these ideas.

4. Content of the Consultation Letter

Despite the multiple functions that a consultation letter must fulfil there is some consensus in the literature about the content of the letter. Though the medical content obviously varies from patient to patient there is some agreement as to what areas of information the letter should contain (Table 2).

Newton et al surveyed 115 GPs and 159 consultants in Newcastle upon Tyne about the content of consultation letters[27]. Responses from the GPs were compared to the responses from the consultants. There was a high degree of consensus between both groups of physicians as to the clinical content of the consultation letters. Items that were felt to be 'always or usually important' in the consultant's reply to the referring MD included: appraisal of the problem, management plan, findings on examination, who saw the patient, what the patient was told, findings on investigation, time of follow up appointment and a summary of the history. In this study the groups were not asked to rank the items in order of importance.

As discussed above Pullen's et al survey identified 5 key items: diagnosis, recommended treatment, follow up arrangements, prognosis, and a concise explanation of the condition[2]. The latter item most likely reflects the educational importance of the letter. The key items identified by Tattersall et al[17] were prognosis, benefits of treatment, side effects, diagnosis, further tests/investigations, test results, clinical findings, treatment options/recommendations, arrangements for follow up and what the patient had been told.

In 1995 Glaxo Wellcome Canada Inc, through the University of Toronto, sponsored a survey exploring the content of consultation letters with physicians attending a CME meeting[28]. Fifty five percent of participants wanted guidelines discussing the diagnosis and treatment in the letter. Sixty eight percent of the participants felt that the specialist's plan of action was highly needed. Seventy percent felt that information on newly initiated therapies was highly needed. Finally, the participant wanted feedback on their referral, diagnostic and medical management skills.

Dojeiji et al surveyed 347 doctors in Eastern Ontario in 1996-1997[29]. The physicians were asked to rank many different content items in order of importance. Consensus was obtained from the rank order on the essential content items. Content that was felt to be essential to the consultation letter included impression (diagnosis if possible), management plan (who will do what and when), investigations to be done and by whom and an indication of any medication changes. Items that were felt to be important to the consultation letter included: indication if a cross referral was made, who will provide ongoing, continuing care, what the patient was told (especially in complicated situations), a rationale explaining the recommendations, and educational information for the referring physician.

Table 2: Essential or Important Items in a Consultation Letter as Assessed by Study

Pullen[2]	Newton[27]	Dojeiji[29]	Babington[4]	Tatterstall[17]	McConnell[18]
			Reason For Referral		
Diagnosis	Appraisal of The Problem	Impression	Diagnosis	Diagnosis	Diagnosis
Recommended Treatment	Management Plans	Management Plans	Management Plans/ Treatment Options	Treatment Options and Recommendations	Treatment Options and Recommendations
	Findings on Examination/ Investigations	Investigations to be Done	Examination and Investigation Findings	Clinical Findings/Test Results	Clinical Findings/Test Results
	Who Saw the Patient	Indications for Medication Changes		Further Testing	
Concise Explanation of Recommenda-tions		Rationale for Recommend-ations			
				Likely Benefits of Treatment	Likely Benefits of Treatment
			Side Effects	Side Effects of Treatment	Side Effects of Treatment
Follow Up Arrangements	Follow Up Arrangements	Who Will Provide Ongoing Care		Follow Up	Follow Up
	What The Patient Was Told	What The Patient Was Told	What The Patient Was Told	What The Patient Was Told	What The Patient Was Told
Prognosis			Prognosis	Prognosis	Prognosis
		Indication if Cross Referral Made			Involvement of Other Doctors
		Educational Information			
	Summary of History		Current Medications		Psychosocial Concerns

As can be seen from Table 2 there seems to be agreement on the important and essential items in a consultation letter. These items include diagnoses, management plans, examination findings, investigation results, follow up plans, prognosis and what the patient was told. Other key items include rationale for recommendation, further testing

required, indication for medication changes and indication if a cross referral was made. The benefits and side effects of treatment as well as the psychosocial concerns of the patient were key items in letters pertaining to oncology patients.

5. Format/Style of the Consultation Letter

Consultation letters can be presented in a narrative or standardised format. This latter format relies heavily on the use of headings and point form (including numbering and bulleting items) without the use of sentences. The preference for the type of format has been studied for both discharge summaries and consultation letters. The different types of formatting has been studied for both the consultation letter and the hospital discharge summary.

Van Walraven et al [30] studied family doctors preferences for a standardised versus narrative discharge summary. Standardised summaries were preferred over narrative ones. The standardised format resulted in a shorter discharge summary compared to the narrative format. The preference for the standardised format increased as the quality rating of previously received summaries decreased and as the length of the narrative format increased. The physicians thought that the standardised summaries contained more relevant information that was more easily accessed.

Lloyd and Barnett[31] reported on a survey of 100 GPs in the catchment area of North Middlesex Hospital NHS Trust on different formatting of a consultation letter. The GPs received 2 fictitious letters, one containing a problem list, one without. Both letters were of the same length and dealt with the same fictitious patient. Ninety-three GPs returned the survey. Eighty-four preferred the letter with the problem list, 3 had no preference and 6 preferred the letters without the problem list. Subsequently 100

consecutive letters from 16 hospitals and 20 specialties were reviewed. Only 10 of the hundred letters used a format that had a problem list.

Rawal et al[32] reported on a follow up study to Lloyd and Barnett's study. Again 100 GPs in the catchment area of North Middlesex Hospital NHS Trust were surveyed on different formatting styles of a consultation letter. The GPs received 2 fictitious letters, one containing a problem list and the other containing a problem list and a list of management proposals. Both letters used headings, which were underlined. Ninety-two surveys were returned. Eighty-one GPs preferred the letter with both problem lists and management proposal lists. Six preferred the letter with only the problem list and 5 had no preference. Subsequently 100 consecutive letters from 16 hospitals and 20 specialties were reviewed. Only 5 of the hundred letters used a format that had both a problem list and a management proposal list.

Glaxo Wellcome has developed a guide to writing effective consultation letters. Keely et al. have recently revised this guide[28]. As part of the development of the guide, communication specialists from McLuhan and Davies Communications Inc reviewed consultation letters. Their writing experts identified several problems that tended to recur. These problems include: too little information, too much information, not answering the referring physicians questions, poor organisation, long sentences, long paragraphs, long non-technical words, too much padding, too much jargon, passive tone, difficult to scan, and little use of point form.

In summary there is a strong preference for letters that rely on a standardised format which utilised problem lists, point form, bulleted items, and headings. This format results in a shorter letter, which is more functional with easier retrieval of items.

6. Current Tools for Evaluating the Consultation Letter

There has been a paucity of research evaluating the consultation letter. Stillman et al looked at the use of a written note, as part of an examination, to evaluate the candidate's ability to describe subjective and objective findings, formulate differential diagnoses, and propose an initial management plan after a standard patient encounter[33]. She used 1784 notes comprising 8 cases completed by 233 post graduate year one (PGY1) foreign trained medical graduates. For each case 20 items of importance were identified and weighed according to their importance. Transcription clerks marked the notes. Because of the use of non-MD as markers, no points were deducted for false positive findings, inappropriate tests ordered and misdiagnoses. There was a wide range in individual resident scores for different cases indicating variability in resident skills for different problems. The scores on the notes were correlated with various other components of the examination. The correlation coefficients between the residents' notes and the other components of the examination were as follows: data gathering: 0.53, interviewing skills: 0.26, tests of spoken English: 0.20, videodisc examination: 0.32, program director ratings: 0.25, FMGEM (foreign medical graduate examination) day 2: 0.14. Program directors felt the evaluation of the patient note had significant face validity. In her study the evaluation tool was case specific and not a generic instrument. This limits the applicability of her marking scheme to general use.

Boudreau et al reported on an OSCE designed to test the consultation skills in respiratory medicine[26]. Three of the 6 stations involved dictating a consultation letter, 1 station involved hand writing a consultation. Each case was marked using a scoring key that was designed by 4 senior respirologists. The scoring key was specific to each

problem and was designed to test the optimal management of the case. In addition, the letters were graded on a 5-point ordinal scale looking at organisation, clarity, succinctness, and educational value. Seven junior medical residents, 9 pulmonary fellows, and 6 consultant were evaluated. There was a systematic trend for better scores with increased training. The difference between the junior staff and the staff was significant. The reliability of the score on the overall quality of the consultation was 0.65. There was poor reliability for the scores pertaining to style and education. The authors noted that the residents completed their consultations with more data items than did the attending physician. The attending physicians' letters dealt more with diagnostic possibilities and recommendations for management.

McCain et al[8] reported on the use of outpatient consultation letters as part of assessment of in-training performance. Their measurement instrument had 4 sections looking at database (history, past medical history, physical examination), problem formulation, diagnostic tests and management. There was a dichotomously scaled checklist for each category and an analogue rating scale. This 165-mm long scale was used to assess the relative severity of any errors noted. Types of errors could include errors of omission, errors of commission, errors in sequence, accuracy and the presence or absence of key findings. The scale was anchored at the far end in terms of the performance of a certified specialist and on the other end indicating that there were major errors.

Trainees were provided with a portable recorder and dictated a letter after seeing a patient in the outpatient setting and before reviewing the case with the supervising consultant. The letter was evaluated by the supervising consultant and 2 other

consultants. Each of the six trainees who participated in the study (2 residents, 2 interns, and 2 clinical clerks) dictated 8 letters. Two clinics were used: a gastrointestinal (GI) clinic and a Rheumatology clinic. The letters were scored in terms of the difficulty of the consultation. For the six most difficult consultations the supervising consultant's letter was also evaluated.

The product-moment correlation between pairs of consultants using the analogue scale showed moderately strong agreement between raters for the categories of problem formulation and overall assessment (range 0.57 to 0.76). There was less agreement on the data base category with product moment correlation coefficients ranging from –0.06 to 0.28 .The product-moment correlation for letters that were assessed by the same consultant at 2 separate points in time were very high (range 0.63-0.96) except for one consultant (range 0.13-0.46). Reliability for the measurement was assessed using a split-half procedure which compared the data sets obtained from even numbered letters with data from odd number letters for each of the trainees. The reliability of each category based on the set of 8 letters per trainee is listed in table 3. This table shows a tendency for the coefficients to be higher for interns and residents compared to clerks. Using the Spearman-Brown formula the authors estimated that good reliability coefficients, in the range of 0.63-0.85, could be obtained for individual trainees based on a sample of 24 consultation letters. The difficulty of the consultation did not affect the inter- or intra-rater agreement. Rather the difficulty level produced shifts in the numerical value of the trainees' assessment with simpler letters receiving higher scores and difficult consultation receiving lower scores.

Table 3: Split Half Reliability Coefficients for the 5 Assessment Categories by Training Level and Type of Service

Training level	Data base	Problem formulation	Diagnostic tests	Management plan	Overall assessment
Rh clerk	0.14	0.14	0.10	0.50	0.19
GI clerk	0.26	0.47	0.22	0.05	0.23
Rh intern	0.64	0.42	0.23	0.78	0.39
GI intern	0.03	0.22	0.34	0.04	0.35
Rh resident	0.26	0.19	0.04	0.27	0.30
GI resident	0.60	0.34	0.25	0.62	0.59

Rh = Rheumatology, GI =Gastroenterology service

The composite means of the trainees were compared to the scores obtained by analysing the letters dictated by the consultants on the 6 consultations that were deemed to be the most difficult (Table 4). Consultants had higher overall scores in every category except for the Rheumatology consultant in the data base category.

Table 4: Mean Ratings of 5 Assessment Categories by Training Level and Type of Service

Training level	Data base	Problem formulation	Diagnostic tests	Management plan	Overall assessment
Rh* clerk	87.9	56.6	7309	61.9	58.3
Rh intern	133.1	124.2	118.3	119.6	122.1
Rh resident	115.5	113.4	118.8	111.5	109.6
GI clerk	125.3	92.8	89.8	89.9	92.4
GI intern	139.9	120.7	130.9	131.5	122.5
GI resident	135.2	132.0	136.7	141.1	132.6
Rh consultant	138.3	160.7	157.7	156.7	154.3
GI consultant	154.7	152.7	153.3	154.7	150.3

*Rh = Rheumatology

Myers et al. developed a generic rating scale to evaluate internal medicine consultation letters[10](Appendix B). The instrument was divided into sections looking at history, physical examination, impression/management, analysis of the writing style and an overall global rating for the letter. A combination of interval and dichotomous scales

was used. The instrument had 34 items. There was 1 overall global rating item. The history section contained 10 items (6 dichotomous, 3 interval and 1 global rating scale), the physical examination section contained 5 items (1 dichotomous, 3 interval, and 1 global rating scale) and the impression / management section contained 9 items (5 dichotomous, 3 interval and 1 global rating item). The writing style section had 9 dichotomous items (table 5).

The scale was pilot tested on 97 consultation letters dictated by 21 residents (the mean number of letters per resident was 4.6 with a range of 1 to 13). The sources of variance were analysed using generalisability theory. The overall quality of the letters using the global rating scale was 2.75 (± 0.82) on a scale of 1 to 5. The inter-rater reliability on each item was poor with a range of 0.15 using the dichotomous scale on the impression/management section to 0.43 for the physical examination dichotomous scale. However, when the items were averaged per resident (an average of 4.6 letters per resident), the reliability improved. The overall reliability per resident was 0.76 for the rating scale on the history and lowest for the dichotomous writing items (0.36). Analysis of the rating scales revealed that the internal consistency of the subscales varied widely. It was lowest for dichotomous items in the impression/ management sections (alpha=0.21) and highest for rating items on the history section (alpha = 0.69).

Keely et al[34] has subsequently used this evaluation instrument in an OSCE setting. Internal medicine residents, after being provided with all the pertinent clinical information, were given 20 minutes to dictate a consultation letter on a case of hyperthyroidism. Two markers scored the letters. Keely et al modified Myer's et al initial instrument in order to eliminate the dichotomous items when calculating the

consultation letter score. Three modified instruments were tested. One contained the 13 items that used Likert scales from the original instrument. The second modified scoring system used an overall rating scale for each section (history, physical exam, and impression/plan) and a global rating scale. The final score system was simply a global rating score.

Thirty-six internal medicine residents participated in the OSCE. In order to compare the results obtained on the consultation letter station to results on other stations, all the station scores were converted to a mark out of 10. All 3 methods of calculating the consult letter score had good inter-rater reliability with a interclass correlation score of 0.71 for all Likert items, 0.72 for overall rating for each section and global score and 0.65 for the global rating item only. The internal consistency of the overall rating scale per section including the global score was 0.79. A total of 2% of the variance was secondary to the rating scale items. Forty-nine percent of the variance was secondary to the resident and 20% was secondary to item x resident.

The correlation between the consultation letter and the total exam score was 0.59. There was a correlation between the verbal communication station and the consultation letter station (0.37) but not between the consultation letter station (addressing hyperthyroidism) and a physical exam station testing the exam of the thyroid. PGY1 residents scored the lowest on the consultation letter station (mean 5.1) and PGY4 residents scored the highest (mean 7.4). The PGY2 residents scored higher than did the PGY3 residents. This was attributed to the observation that the PGY2 residents were a stronger group of residents overall. Thus Keely and Myers instrument shows considerable

improvement in the inter-rater reliability in the OSCE setting where there is standardised patient information for a single patient problem.

The results suggest that the ability to dictate a good letter is only in part related to the ability to communicate with a patient. The correlation between the total score on the OSCE and the letter score suggests that the ability to dictate a good referral letter is an important component of the construct of a general internist. This suggests that the evaluation of the letter may be important for assessment of competence.

Crossley et al[6] developed a tool to assess outpatient letters (the Sheffield Assessment Instrument for Letters (SAIL)). A prototype instrument was developed using the consensus of 1 family physician, 1 paediatric consultant and 16 specialist registrars. The authors based their instrument on the statement that the perfect consultation letter would be one that "clearly conveys the information I would like to have about the patient if I were the next doctor to see him/her". The pilot instrument was piloted in 2 stages to assess clarity and completeness. The final instrument (Appendix A) had 18 dichotomous items and a 10 point global rating scale (Table 5).

The reliability of the SAIL was assessed on 260 letters submitted by paediatric specialist registrars. A general practice physician, a consultant paediatrician and a paediatric specialist registrar marked each letter. Judges also marked the letters on the complexity of the case. There was high agreement between the global rating of the letter and the sum of the checklist with a Pearson's coefficient of 0.91. The overall reliability of the instrument was 0.72. The overall reliability of the global rating scale was 0.74. Analysis of variance showed that only 3% of the variance was related to the clinical problem and 1% related to the difficulty of the case. Doctors contributed 17% to the

variance, judges' different views about a given letter accounted for 37% of the variance and the tendency of a judge to be a hawk or a dove accounted for 11% to the variance. The letter to letter variation within each doctor accounted for 37% of the variance. Using generalisability analysis the authors concluded that 8 markers marking 8 letters (written by a single physician) or 3 judges marking 40 letters (from a single physician) would yield a reliability coefficient of 0.80. Comparison between the pairs of judges over individual items showed that the consultant and registrar agreed closely about whether a letter contained a clear management plan (kappa = 0.59) but the general practitioner agreed with neither of them (kappa = 0.03 and 0.02).

Both generic instruments have the advantage of being very simple to use. The mean time to mark a letter was 5 minutes using Myers' et al instrument[10]. Both instruments rely heavily on dichotomously scored items (Table 5). Myers' et al instrument uses dichotomous scales in twenty-one of the 34 items whereas the SAIL used dichotomously scaled items except for the global scale. The use of dichotomous scales (in items dealing with attributes and behaviours) results in a loss of potential information, a loss of efficiency and decrease reliability (as different markers have different ideas about what constitutes a positive result)[35]. As a consequence, in Myers' et al study[10], the inter-rater reliability on each item was poor with a range of 0.15 using the dichotomous scale on the impression/management section to 0.43 for the physical examination dichotomous scale.

Table 5: Comparison of the Generic Tools for the Evaluation of the Consultation Letter

	Myers and Keely[10]			SAIL[6]		
Section	Number of items-ordinal scale	Number of items-dichotomous scale	Weight of section relative to total score (%) (Max score = 34)	Number of items-ordinal scale	Number of items-ordinal scale	Weight of section relative to total score (%) (max score 28)
Brevity	3	0	17	Not assessed	Not assessed	
Format	0	9	10	0	3	11
Clarity	3	0	17	0	3	11
History	2	6	19	0	2	7
Exam	2	1	13	0	1	4
Diagnosis	2	2	14	0	1	4
Management				0	3	11
Discussion with family/pt	0	1	1	0	1	4
Follow up	0	1	1	0	2	7
Explanation/ education	0	1	1	0	2	7
Global scale	1	0	6	10 point scale	0	36

With respect to the content validity, both instruments assess the key items that are required in a consultation letter. Myers' et al instrument heavily weighs style issues (table 5). Excluding the 4 global rating items, 15 of the remaining 30 items pertain to style (including clarity, organisation of each section, use of words, paragraphs, sentences, and headings). In addition, the instrument has relatively high weighting on the patient's history and physical exam with less weight on the diagnosis and management. This detracts from the content validity of the instrument since, from the point of view of the referring doctor, it is the management issues that form the crux of the letter.

The SAIL assesses management issues to a greater degree with very little weight placed on the history or physical examination. This weighting of management issues is very appropriate from the point of view of the referring physician. Unfortunately, the use of dichotomous items is a significant draw back for this instrument.

Neither instrument allows for the assessment of the management of individual medical problems. This is especially important in internal medicine patients with multiple medical issues that need addressing. This may result in a decreased ability of the instrument (and the marker) to give detailed feedback in the ambulatory care setting. In addition the inability to assess each medical problem individually may result in a decrease in the reliability of the instrument in view of the case specificity of knowledge. How does one mark a letter that manages COPD well but does not adequately address the issue of osteoporosis in a patient with both these problems?

Thus, though there are 2 good generic instrument available, both these instruments have limitations, which detracts from their reliability, validity and their ability to give detailed feedback in the ambulatory care setting.

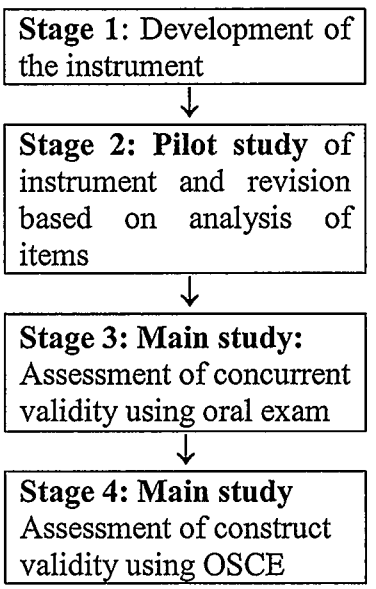
7. Research Question

Though 2 generic consultation letter evaluation tools are available neither of them have been used on practising consultants. Both instruments rely heavily on dichotomous items, which detract from their reliability and their ability to give precise feedback to the trainee. Neither instrument is designed to assess letters pertaining to patients with complex multi-system illnesses and multiple medical problems. These patients are becoming more common in the practice of internal medicine. The purpose of this thesis is to develop a reliable and valid consultation letter assessment instrument (the Consultation Letter Assessment Tool = C.L.A.T) that can be used on patients with multiple medical problems. The research question is: can an assessment tool be developed to assess consultation letters on patients with multi-system illness that is reliable and has content, concurrent, and construct validity?

III. METHODS

In order to assess the reliability and validity of the CLAT the project had 4 stages (figure 1). Stage 1 consisted of the development of the instrument. Stage 2 consisted of a pilot study. The purpose of the pilot study was to assess the reliability of the instrument as well as to analyse individual items. Based on the analysis, the instrument was modified to optimise reliability. Stage 3 and 4 formed the main parts of the study, and assessed the validity of the instrument. Stage 3 examined for evidence of concurrent validity by comparing the results obtained from the CLAT to results obtained from an oral exam. Stage 4 compared the scores from the CLAT for various level of expertise in order to obtain further evidence of construct validity.

Figure 1: Outline of the Assessment of the CLAT



1. Stage 1: Development of the Instrument

Prior to developing the questionnaire, a literature review was done to look at the attributes of a consultation letter in Internal Medicine. This included an estimate of the relative importance of each attribute with respect to the letter as a whole. The current

tools available for the analysis of the consultation letter were analysed to see how well they performed and how they weighed each attribute.

Interval scales of measurement were chosen for the items. The use of dichotomously scaled items was avoided in view of their impact on reliability. The literature supports 7 ± 2 as the ideal number of categories per item[35] Five categories were used. The use of 5 categories reduces the reliability by about 12% compared to items with 7 categories. However, when a large number of individual items are summed to create a global score, it is felt that the use of 5 categories does not result in a significant loss of information.

A pilot instrument was developed (table 6, Appendix C). The development of the tool included selection of the measuring scales as well as the items. The consultation letter assessment tool (CLAT) had 23 items. There were 6 main sections in the CLAT: format, educational value of the letter, data gathering (or data synthesis), patient management, communication and a global rating item. The section on formatting had six items, the section on educational value of the letter had three items, and the data section had seven items. Given case specificity, the section on medical management was expanded to allow each problem to be marked independently using four items. These four items assessed patient management including diagnosis, investigations, pharmacological management and prevention (one item for each of these attributes). The communication section assessed the communication of follow up plans with the referring doctor and had three items. There was one global rating item. The CLAT did not have a maximum score since it was dependent on the number of medical problems that needed addressing.

Table 6: Pilot Questionnaire: Items And Rationale

Section	Item Assessed	Rationale
1. Format	Assessment identified in the text	Increases the user-friendly nature of the letter[3;28;30-32]
	Plan identified in the text	
	Letter easy to scan	
	Letter is succinct with appropriate level of detail	Balance between the requirements for completeness especially from legal point of view and user friendliness[28]
	Tone of letter	Assesses professionalism in communication[3]-this is especially important if patients are reading the letters[12]
	one idea per paragraph	Increases the user-friendly nature of the letter[28]
2. Educational value of letter	Letter contributes to the reader's CME	Emphasises the educational importance that family doctors attribute to the letter[2;15;29]
	Letter provides explanation for recommendations	
	Letter highlights area of controversy or new developments	
3. Data gathering	Reason for referral identified	Important house keeping components of the letter as a legal document as well as for future reference for the physician for follow up visits[28;29]
	Relevance of the history for the patient's illness	
	Completeness of patient's present illness	
	Pertinent ancillary history	
	Medications and allergies	
	Completeness and relevance of physical exam	
	Interpretation of ancillary tests	

4.Management*	Diagnosis and differential diagnosis	The most essential component of the letter for further patient management, also important in CME of reader[2;4;16-18;27]
	Investigations	
	Pharmacological management	
	Non-pharmacological management including prevention	
5. Communication	Follow up plans clear	Essential for management and follow up of patient[3;28;29]
	Reporting on discussion with patient	Important for collaborative care[16-18;27]
	Answers questions posed in referral letter	Contributes to CME and ensures that the family doctor's reason for referral addressed[15;29]
6.Global rating	Overall global rating	

* Each medical problem marked individually

2. The Pilot Study (Stage 2): Assessment of Reliability and Item Analysis

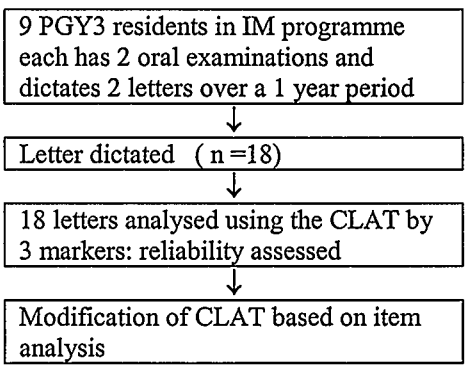
The purpose of the pilot study was the assessment of the reliability of the CLAT. This included both intraobserver and interobserver reliability. Item analysis was performed as part of the pilot study in order to enhance the reliability of the CLAT. (Figure 2).

Subjects: The reliability of the CLAT was assessed using 18 letters dictated by PGY3 residents in Internal Medicine at the University of Calgary. These letters were dictated as part of the regularly scheduled biannual oral examinations for PGY3 residents. Nine PGY3 residents each dictated 2 letters for a total of 18 letters. These letters covered a broad range of patient problems with several problems identified per patient. None of the

letters had an accompanying referral letter. Ancillary test data was not available to the resident at the time of dictation. An experienced transcriptionist transcribed the letters.

Raters: The 18 letters were marked independently by three faculty members from the Department of Internal Medicine at the University of Calgary: Dr D. Megran, Dr S. Coderre, and Dr M. Ainslie. All the letters were marked once. The markers were blinded as to the author of the letters.

Figure 2: Outline of the Assessment of Reliability



Statistical Analysis:

All the statistical analysis in the thesis was performed using Statistica 6™ software. Internal consistency of the CLAT was calculated using a Cronbach's alpha score. Inter-rater reliability was assessed using Pearson's correlation coefficients. Factor analysis was performed on each section of the CLAT in order to look for highly correlated items. Each item was analysed to see if it contributed or detracted from the overall reliability of the section. In order to examine the effect of case specificity on the CLAT score the results from each section of the CLAT on the resident's first letter were compared to the results from the resident's second letter using Pearson's correlation coefficients (table 7).

Table 7: Data from the Pilot Study

	Markers 1 to 3					
	Case 1			Case 2		
PGY3 N=9	V1	V2	Vn	V1	V2	Vn

V1, V2 and Vn represent items or items that cluster together
Case 1 and 2 represents the letters dictated by the residents

Sources of variance: There are multiple sources of variance with this study design. Sources of variance can be clumped into those pertaining to the residents and markers. Sources of variance for the residents include the resident's true ability to dictate a consultation letter, the resident's ability to manage each case (resident x case), the change in performance with time (resident x time). Sources of variance for the judges include the internal consistency for the judges with respect to the CLAT, the tendency of the judge to be a hawk or dove, and the interaction of the judges with the case.

3. Main study (Stage 3): Assessment of Concurrent Validity

The purpose of this part of the study was to look for evidence of concurrent validity. The results from an oral examination in Internal Medicine were compared to the scores obtained using the modified CLAT on letters dictated as part of the examination (figure 3).

Subjects: Eight PGY3 residents in Internal Medicine at the University of Calgary participated as part of their regularly scheduled biannual oral examination in Internal Medicine. Each resident dictated one letter. (Figure 3).

Exam format: Prior to meeting with the examiners each resident had an hour and 15 minutes to take a complete history, perform a physical examination and dictate a letter on a stable outpatient. The patients were volunteers with the medical school's clinical skills programme. The patients were instructed to pretend that their family doctor had referred

them to an internist for review of their medical problems. They were instructed to tell the resident everything that they knew about their medical problems, as they would do in real life when seeing a specialist. Each resident had a different patient. After dictating the letter the residents met with 2 examiners at the bedside. The resident had approximately 15 minutes to present the history and physical exam findings to the examiners. The examiners were allowed to clarify any questions that they had concerning the patient's history and verify any key physical exam findings. Following this the resident was asked to demonstrate certain physical exam manoeuvres (bedside scenarios). The examiners and resident then moved to a different room. There the resident was asked to summarise the case and discuss his/her assessment and management plans. Examiners were allowed to ask probing questions about the resident's clinical impression.

Scoring of the oral examination: Sixteen faculty members from the Department of Medicine at the University of Calgary volunteered to be examiners. A pair of faculty members examined each resident. The examiners independently marked the resident using a standardised scoring sheet. Table 8 compares the items assessed on the modified CLAT and the oral examination.. The standardised scoring sheet for the oral examination consisted of 14 items. Each item was scored using a 5-point interval scale with anchoring for each score. The items on the standardised scoring sheet were taken from the modified CLAT. These items were felt to be applicable to the oral examination setting. Items from the modified CLAT that pertained to format and style issues of the consultation letter were not scored on the oral examination. The wording of the oral exam items were modified to suit the oral examination format. For example item #4 in the modified CLAT which asks if the letter is succinct corresponded to an item which asked if the oral

presentation was succinct. Table 8 compares the items assessed on the CLAT to the items assessed on the oral examination.

Scoring of the letters: Two markers (S.C and M.A) marked each letter using the modified CLAT. The markers were blinded to the name of the resident and to the results of the oral examination.

Comparison: For each resident the results of the CLAT were compared to the standardised scoring sheet used by the oral examiners. Figure 3 outlines the process.

Figure 3: Assessment of Concurrent Validity

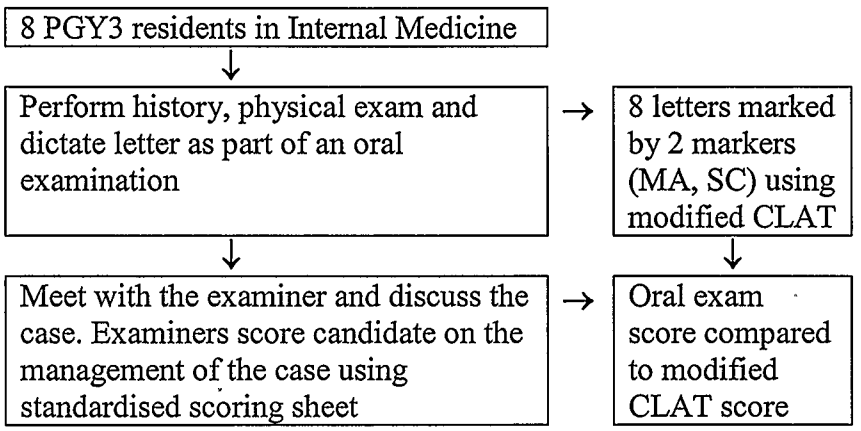


Table 8: Items Assessed by the Modified CLAT and in the Oral Examination

Attribute Assessed in the Modified CLAT	Assessed in Oral Examination
1. Assessment clearly identified in text	Not assessed
2.Plan clearly identified in text	Not assessed
3. Ability to scan text to obtain key data	Not assessed
4.Succinctness	Modified for oral presentation
5.tone of letter	Not assessed
6. Paragraphs dealing with 1 idea	Not assessed
7. CME value of letter	Not assessed
8. Explanations for recommendations	Not assessed
9. chief complaint clearly identified	Item modified for oral presentation
10. description of the patient's history	Identical item
11. completeness of history	Identical item

12.Pertinent ancillary history	Identical item
13. medications	Identical item
14.Allergies	Identical item
15. Physical examination	Identical item
16.ancillary tests interpretation-omitted	Not assessed
17.assessment	Identical item
18.investigations	Identical item
19. pharmacological management-dosing/duration	Identical item
20.pharmacological management-guidelines	Identical item
21.non pharmacological management	Identical item
22.follow up plans-ordering of tests	Not assessed
23. follow up plans-patient follow up	Not assessed
24. summary of what patient was told	Not assessed
25.answers to questions in referral letter-omitted	Not assessed
26. Global rating score	Identical item

Statistical analysis: Statistical analysis included the calculation of Cronbach's alpha coefficients to determine the reliability of the modified CLAT and the oral examination-scoring sheet. Descriptive analysis of the letters included the length of the letter, and number of problems per patient. Pearson's R correlation coefficients were used to compare the mean scores on each section of the modified CLAT to the mean score for the corresponding section of the oral examination-scoring sheet for each resident. Since each patient had a different number of problems and since many of the oral examiners did not give a separate mark for each medical problem in the management section, the scores for the items in the management section were averaged over all the medical problems. These average scores were compared.

Sources of variance: Sources of variance for the residents include the resident's skill in dictating the letter, the residents ability to manage the case (resident x case), the resident's ability to perform in an oral examination setting (including their interactions with the judges and possible halo effect). For the raters (both oral and written exam)

sources of variance include the internal consistency of the rater, the tendency to act as dove or hawk, the interaction between judges (oral x written), the interaction with the resident and their interaction with the case (judge x case).

4. Main Study (Stage 4): Assessment of Construct Validity of the Modified CLAT

In this part of the study letters dictated by residents were compared to letters dictated by faculty members. The effect of increasing expertise on the scores on the modified CLAT was assessed to look for evidence of construct validity.

Subjects: Eight PGY1 residents in Internal Medicine, 9 PGY2 IM residents and 6 faculty members of the Department of Internal Medicine.

Exam format: During the PGY1 and 2 years, residents participate in an OSCE as part of their ongoing assessment of training. At one of the OSCE stations PGY1 and PGY2 residents were asked to dictate a letter. At this 15-minute station the residents were provided with a referral letter and all pertinent information pertaining to history, physical examination findings and ancillary data. This information was given in writing in point form. The residents were instructed to dictate a consultation letter based on the information provided. This process was identical to that described by Keely et al [34]. Six faculty members also participated. They were given the identical instructions and information as the residents. They were told to dictate the letter in 15 minutes.

The station consisted of an imaginary case of shortness of breath in a hypertensive middle age male, with risk factors for both asthma and coronary artery disease. The values from a spirometry showing a normal FVC (forced vital capacity) and FeV_1 (forced expiratory volume in 1 second) but a borderline low FeV_1/FVC ratio were provided to the

resident. The scenario was chosen to be ambiguous with no single correct cause for the shortness of breath.

Raters: Two markers scored the letters (Dr M Ainslie and Dr K Rimmer). The markers were blinded to the name and training level of the author. One of the markers (KR) was not aware that faculty had dictated letters. It was decided a priori that only items 17 and 18 would be scored for the assessment of the patient's shortness of breath. Only item 17 was scored for the assessment of the patient's hypertension. Item 24 was omitted. This item deals with what the patient was told which was not applicable, as there was no patient in the OSCE station.

Statistical analysis: Cronbach's alpha correlation coefficient was calculated to look at the internal consistency of the modified CLAT. ANOVA was used to determine the differences in the mean scores of the modified CLAT (total overall scores and scores for each section) and the level of training (table 9).

Table 9: Data from the OSCE

Level of Expertise	Judge 1 + 2		
PGY 1-2	V1	V2	Vn
Internist	V1	V2	Vn

Level of expertise ranges from PGY1-2, internists
V1,V2 and Vn represent items or items that cluster together
Judge 1 and 2 is the mean of the scores from judge 1 and 2

Sources of variance: In this study the data collection was standardised and the same for all the participants. Sources of variance include level of training of the participants, the participant's ability to dictate a letter and participant's ability to manage the case. Sources of variance with respect to the judges include the judge's internal consistency for each

letter and case to case, tendency to act as hawk or dove, interaction with the case, and consistency judge to judge.

IV. RESULTS

The data from the pilot study were used to calculate the reliability of the instrument and to analyse individual items from the instrument. The instrument was modified based on the pilot study analysis. The validity of the modified CLAT was assessed in the main study. As construct validity is established by looking at evidence from multiple sources the main study had 2 components. The first part looked at the concurrent validity of the modified CLAT compared to an oral examination. The second part of the main study assessed the impact of expertise on the modified CLAT score.

PILOT STUDY:

1. Descriptive Characteristics of the Letters Used in the Pilot Study

Eighteen letters, dictated by 9 PGY3 residents (2 letters per resident), were studied in the pilot study. Three internists (DM, SC, MA) marked each letter. The mean time to mark each letter was 11 minutes. The mean number of pages per letter was 2.6. Each letter addressed a mean of 2.8 medical problems. (Table10)

Table 10: Descriptive Characteristics of the Letters

	Means +/- SD (range)
Time to mark each letter	11 +/- 3.5minutes (6-25)
Length of letter	2.6 +/- 0.7 pages (1.75-4.25)
Number of medical problems per letter	2.8 +/- 0.7 problems(2-4)

Item 16 was not scored in the pilot study, as the residents did not have access to ancillary tests in the oral examination setting. Likewise item 23 which pertains to answering the referring doctor's questions was not applicable, as the residents did not receive a referral letter.

The study utilised non-standardised patients for each resident. This resulted in a different number of medical problems being identified for each patient. As the CLAT allows for separate analysis of each medical problem, the absolute total score is dependent on the total number of medical problems. For the analysis, the score of each item in the management section averaged over the number of medical problems is reported.

The mean score for each section was as follows: format 3.63(standard deviation (SD) 1.07), data collection 3.46(SD 1.02), management 2.43(SD1.15), educational value 2.35(SD1.00), communication 1.62 (SD 1.13), and global scale 2.37(SD 0.71). For summary of results on the pilot study see table 11. Table 12 lists the mean scores per section per judge. The Pearson's correlation coefficient was calculated between judges for the totals of each section (table 13). For the complete instrument the Pearson's correlation coefficient between pairs of judges ranged from 0.50 to 0.75. This indicates moderate correlation. On individual sections the correlation coefficient between judges varied considerably. There was poor correlation between the judges on the sections pertaining to data collection (range 0.19-0.60), communication (0.10-0.35), and global score (-0.06-0.64). The correlation coefficient was highest on the format and management sections (range 0.55-0.87).

Table 11: Means of Scores and Standard Deviation per Sections of the CLAT

	Means of Scores per Section	Standard Deviation
Format	3.63	1.07
Educational value	2.35	1.00
Data collection	3.46	1.02
Management	2.43	1.15
Communication	1.62	1.13
Global scale	2.37	0.71

Table 12: Means of Scores per Sections Per Judge for the Pilot Study

	Total	Judge1	Judge2	Judge3
Format	3.63	3.44	3.61	3.86
Data collection	3.46	3.32	3.61	3.44
Management	2.43	2.25	2.66	2.33
Educational value	2.35	2.30	2.76	1.98
Global scale	2.37	2.27	2.50	2.33

Table 13: Correlation between Judges for Each Section of the Pilot Study

	Judge 1 Vs 2	Judge 2 Vs 3	Judge 1 Vs 3
Complete questionnaire	0.75	0.65	0.50
Format(Q1-6)	0.68	0.52	0.55
Educational value(Q7-9)	0.41	0.35	0.45
Data collection (Q10-15)	0.19	0.60	0.27
Management (Q17-20)	0.87	0.65	0.68
Communication (Q21-22)	0.35	0.10	0.26
Global scale	-0.06	0.64	0.23

Each resident dictated 2 letters as part of the pilot study. The total score and the average score per section were compared between the resident's first and second letter (table 14). There was moderate correlation between the scores on the communication section and educational value section between each letter. No significant correlation was found for the scores on format, data collection and management. Figure 4 shows the correlation between the total scores for each resident.

Table 14: Correlation between Each Letter Per Resident for Each Section

Section	Average for Sections on Letter 1	Average for Sections for Letter 2	Pearson's Correlation Coefficient Between First and Second Letter
Format	3.73	3.55	-0.05
Educational value	2.40	2.30	0.54
Data collection	3.50	3.42	0.25
Management	2.42	2.36	0.37
Communication	1.78	1.59	0.71
Total score	85.59	81.07	0.24

2. Pilot study: Evaluation of Reliability and Item Analysis of the CLAT

The reliability of the CLAT was assessed in the pilot study. The Cronbach's α coefficient was determined for the CLAT as a whole as well as for each section. The Cronbach's α coefficient was 0.91 calculated using the 3 markers each marking 18 letters. Table 15 lists the Cronbach's α coefficient for each section of the CLAT for the pilot study.

Table 15: Cronbach's α Coefficient for Each Section of the Pilot Study

Section	Cronbach's α Coefficient for Section	Variance of Scores per Section
Format	0.73	1.31
Education	0.78	0.99
Data collection	0.66	1.03
Management	0.70	0.86
Communication	0.66	0.97

Factor loading as well as reliability was calculated on the items for each section. Within the section pertaining to the format of the letter item 1 and 2 fell on to one factor and were tightly correlated, as was item 4 with respect to item 6 (table16). Item 5 (assessing whether the tone of the letter was professional) detracted from the reliability of the instrument. As a consequence, this item was changed in the modified CLAT to assess whether the tone of the letter was polite.

Table 16: Factor Loading for Items Pertaining to Format (Varimax Normalized)

Item on CLAT	Factor 1	Factor 2	Cronbach's α if item omitted (α=0.73 based in all items)
Item 1	-0.01	0.87	0.69
Item 2	0.04	0.86	0.70
Item 3	0.58	0.63	0.60

Item 4	0.91	0.15	0.67
Item 5	0.12	0.30	*0.75*
Item 6	0.91	0.02	0.70

All 3 items pertaining to the educational value of the letter fell onto one factor and were tightly related (table 17). Item 9 assessed whether the letter highlighted areas of controversy or areas of new development. This item detracted from the reliability of the CLAT and was omitted in the modified CLAT.

Table 17: Factor Loading for Items Pertaining To Education

Item on CLAT	Factor 1	Cronbach's α if item omitted (α=0.78 based in all items)
Item 7	-0.96	0.61
Item 8	-0.93	0.60
Item 9	-0.85	*0.83*

In the section pertaining to data collection items 11, 12 and 15 (items pertaining to the patient's history and physical examination) were tightly related (table 18). Item 14 detracted from the reliability of the section. This item assessed 2 distinct features: medications and drug allergies. As both of these pieces of information were felt to be important parts of the consultation letter the item was split into two items in the modified CLAT.

Table 18: Factor Loading for Items Pertaining To Data Collection

Item on CLAT	Factor 1	Factor 2	Cronbach's α if item omitted (α=0.66 based in all items)
Item 10	0.43	-0.34	0.66
Item 11	0.81	-0.33	0.56
Item 12	0.91	-0.16	0.52
Item 13	0.68	0.51	0.58
Item 14	-0.05	0.88	*0.77*
Item 15	0.75	0.16	0.54

Analysis of the factor loading for items pertaining to management showed that item 17 (pertaining to assessment) and item 19 (pertaining to pharmacological

management) fell onto one factor and were tightly related (table 19). Item 18 (pertaining to investigations) and item 20 (pertaining to non-pharmacological management) detracted from the reliability of the CLAT. In the modified CLAT pharmacological management was split into 2 items. Items 18 and 20 were felt to be important parts of the consultation letter and were left in the modified CLAT. The scoring key for item 18 was modified.

Table 19: Factor Loading for Items Pertaining To Management (Unrotated)

Item on CLAT	Factor 1	Cronbach's α if item omitted (α=0.70 based in all items)
Item 17	0.92	0.45
Item 18	0.54	0.74
Item 19	0.86	0.55
Item 20	0.59	0.75

Item 21 and 22 (assessing the communication of follow up plans and discussion with patient) were tightly related (table 20).

Table 20: Factor Analysis on Items Pertaining To Communication

Item on CLAT	Factor I
Item 21	0.80
Item 22	0.80

Based on the pilot study the instrument underwent modifications. The modified CLAT (appendix D) was used in the main study to assess the validity of the CLAT.

MAIN STUDY: Assessment of the Validity of the Modified CLAT

The main study assessed the reliability and the construct validity of the modified CLAT. The reliability of the modified CLAT was calculated using letters dictated as part of an OSCE and oral examination. The score of the CLAT were compared to scores from an oral examination to look for evidence of concurrent validity. Evidence of construct validity was obtained by comparing the CLAT of PGY1 and 2 residents to the scores obtained by faculty on an OSCE station.

Table 21 provides the means and the standard deviations for each section of the modified CLAT in the main study. The PGY1 and 2 residents scored highest on the sections pertaining to data collection and the format of the letter. The lowest average scores were for the sections pertaining to the management of the patient and the global score.

Table 21: Means of Scores and Standard Deviations for Each Section of the Modified CLAT

	Modified CLAT Used in Oral Examination (Standard Deviation)	Modified CLAT Used on OSCE – for PGY1 and 2 Residents (Standard Deviation)	Modified CLAT Used on OSCE – for Faculty (Standard Deviation)
Format	3.73 (0.99)	3.94 (0.92)	4.36 (0.70)
Educational value	3.16 (0.88)	3.29 (1.04)	4.63 (0.58)
Data collection	3.96 (1.17)	4.53 (0.84)	4.64 (0.62)
Management	2.84 (1.21)	2.90 (1.23)	3.86 (1.39)
Communication	1.98 (1.04)	3.23 (1.38)	4.56 (0.91)
Global scale	2.75 (0.68)	2.90 (0.97)	4.38 (0.71)

1. Evaluation of Reliability of the Modified CLAT and Item Analysis

The reliability of modified CLAT was evaluated using 23 letters dictated as part of the OSCE and 8 letters dictated as part of the PGY 3 oral examination. Table 21 shows the Cronbach's α coefficient for the modified CLAT for each examination. The reliability coefficient was 0.91 for the OSCE station, and 0.78 for the oral examination.

Table 22: Cronbach's α Coefficient for Each Examination

Study	Number of Letters	Cronbach's α Coefficient
Pilot study	18	0.91
Oral examination	8	0.85
Modified CLAT used in oral examination	8	0.78
Modified CLAT used in OSCE	23	0.91

The reliability was calculated for each section of the modified CLAT using the data from the OSCE (table 23). The reliability was the highest for the educational value section (0.96) and communication section (0.85) and lowest for the section on data collection (0.60). These reliability coefficients vary slightly from the reliability coefficients derived from the pilot study (table 15).

Table 23: Cronbach's α Coefficient for Each Section of the Modified CLAT

Section	Cronbach's α Coefficient	Variance of Scores for Each Section
Format	0.72	0.78
Educational Value	0.96	1.21
Data	0.60	0.62
Management	0.65	1.75
communication	0.85	2.24

Item analysis was performed to look for items that detracted from the overall reliability of the modified CLAT. Item 5 was modified from the initial CLAT to assess

whether the tone of the letter was polite. This item was found to be tightly related to item 4 (which assessed whether the letter was succinct) in the modified CLAT (table 24). In the modified CLAT item 5 did not contribute or detract from the reliability of the instrument. Item 14 in the original CLAT was separated into 2 items in the modified CLAT (items 13 and 14) assessing medications and drug allergies respectively. These items were tightly related and contributed significantly to the reliability of the data section in the modified CLAT. Item 18, which performed poorly in the pilot study, contributed significantly to the reliability of the modified CLAT.

Table 24: Reliability of Modified CLAT if Items Removed

Modified CLAT item	Section	Cronbach's α for Each Section	Cronbach's α for Each Section if Item Omitted
Item 4	Format	0.72	0.67
Item 5	Format	0.72	0.72
Item 13	Data collection	0.60	0.45
Item 14	Data collection	0.60	0.45
Item 16	Data collection	0.60	*0.71*
Item 18	Management	0.65	0.32

2. Main Study: Evaluation of Concurrent Validity

Concurrent validity was assessed using an oral examination. Eight PGY 3 residents participated in the third phase of the study that evaluated their consultation letter dictated as part of an oral exam using the modified CLAT. Two markers (MA and SC) evaluated eight letters using the CLAT. The average length (estimated to the quarter page) of the letters was 3.5 pages (range 2.5 to 6.25). An average of 3.6 problems was identified per patient. The average global score on the quality of the letter was 2.75 with 2 being borderline quality and 3 being satisfactory (table 11).

Each resident had two examiners who marked the oral examination independently. Table 25 compares the problems identified by each oral examiner to the problems identified by the resident in the consultation letter. There was not complete agreement between examiners on the key medical problems identified by the resident during the presentation of the case. Problems identified by residents in the consultation letter were not identical to the problems identified by the examiners during the oral examination. This undesired variability between examiners is illustrated in Table 24 below.

Table 25: Medical Problems of Patients Identified in the Letters Compared to that Identified by the Oral Examiners

Letter #	Problems in Letters	Problems –Oral Examiner 1	Problems –Oral Examiner2
1	stroke hypertension	stroke hypertension left ear infection	stroke hypertension left ear infection
2	Ischemic heart disease (IHD) airways disease(sarcoid and asthma) diabetes osteoporosis	IHD asthma sarcoid asbestos exposure	IHD asthma sarcoid asbestos exposure
3	COPD Angina thyroid disease nocturia	COPD Coronary artery disease(CAD) hypertension hypothyroidism	COPD CAD hypertension bipolar disorder
4	colonic cancer recurrent falls dysphagia pagets goitre	colonic cancer dysphagia falls goitre	colonic cancer dysphagia falls goitre
5	diabetes peripheral vascular disease (PVD) CAD Colonic cancer	PVD CAD Colonic cancer TB diabetes	hyperlipidemia CAD anticoagulation diabetes
6	syncope cough oral lesion polyuria	syncope aortic stenosis puritis oral lesions	syncope nocturnal cough puritis polyuria/polydypsia
7	diabetes renal disease rheumatoid arthritis CAD	diabetes nephrotic syndrome pulmonary embolism (PE) CAD	diabetes nephopathy PE neuropathy
8	gout asthma optic neuritis	gout asthma gastro-esophageal reflux diseaseoptic neuritis steroid use sinusitis	asthma hiatus hernia optic neuritis steroid use

The scores from the oral examination and the CLAT averaged over both examiners and for each section, were compared using a Pearson's correlation coefficient (table 26). This revealed moderate correlation between the scores on data collection (0.60) and management (0.53). There was poor correlation between the global score (0.37) and total score (0.20). Format was compared using one item only (succinctness of letter and presentation) and was moderately inversely correlated.

Table 26: Correlation between the Scores on the Oral Examination and the Letter

Section	Average for the Sections on the Oral Examination	Average for the Sections on the Letters	Pearson's Correlation Coefficient between Oral Exam and Letter (r)
Format (item 4 only)	3.8	3.7	-0.53
Data collection	3.9	4.0	0.60
Management	3.7	2.8	0.53
Global score	3.5	3.2	0.37
Total score	50.0	47.6	0.20

3. Evaluation of Construct Validity

A total of 23 letters were dictated as part of an OSCE examination. Eight PGY1 residents, 9 PGY2 residents and 6 Internal Medicine faculty members participated. A referral letter and laboratory tests were included as part of the information given to the participants. Item 16 (pertaining to interpretation of test results) and item 25 (pertaining to answering questions in the referral letter), which were not assessed in either the pilot study or the study looking at concurrent validity, were assessed in this study. The letters were marked in blinded fashion by 2 respirologists (MA and KR). Table 27 shows the average scores of the residents and faculty per section.

The scores on the CLAT showed significant differences with the level of expertise. Analysis of variance (ANOVA) between the averages per section revealed significant differences in the total score as well as the mean scores for sections on educational value of the letter, communication of follow up and global score. (Table 26). The differences in scores are depicted graphically in figures 5-11. The increasing scores with level of expertise is supportive of the construct validity of the CLAT.

Table 27: Averages per Section for Level of Training for Modified CLAT in OSCE

Averages per Section	PGY1(n=8)	PGY2 (n=9)	Faculty (n=6)	P value
Format	3.99	3.95	4.36	0.20
Education	3.28	3.31	4.63	0.006
Data collection	4.44	4.53	4.64	0.45
Management	3.02	2.80	3.86	0.10
Communication	3.42	3.06	4.56	0.03
Global score	2.78	3.00	4.38	0.002
Total score	88	87	102	0.03

Figure 4: Correlation Between the First and Second Letter Scores for Each Resident in the Pilot Study

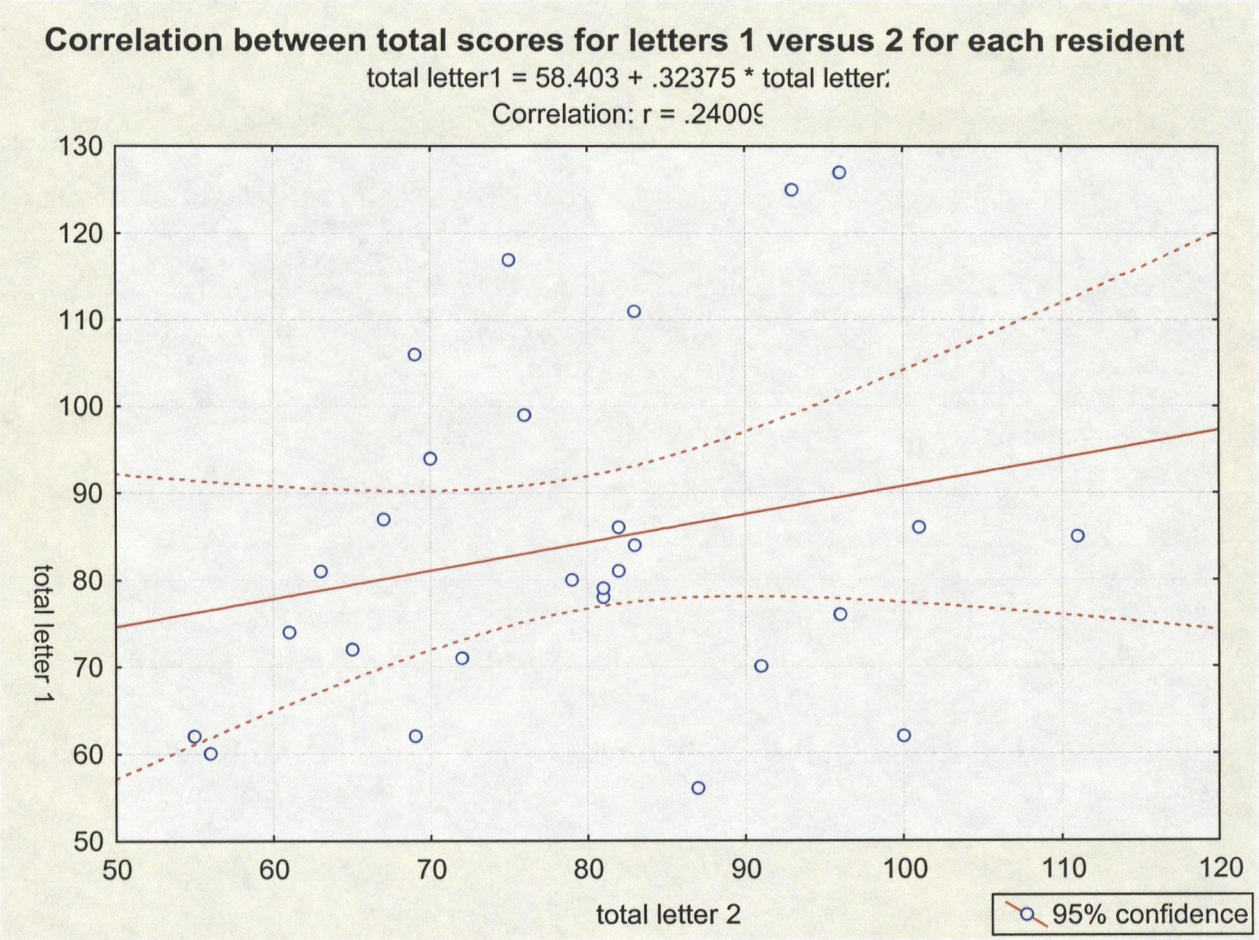


Figure 5: Average Score on Format Items for Level of Training in the OSCE

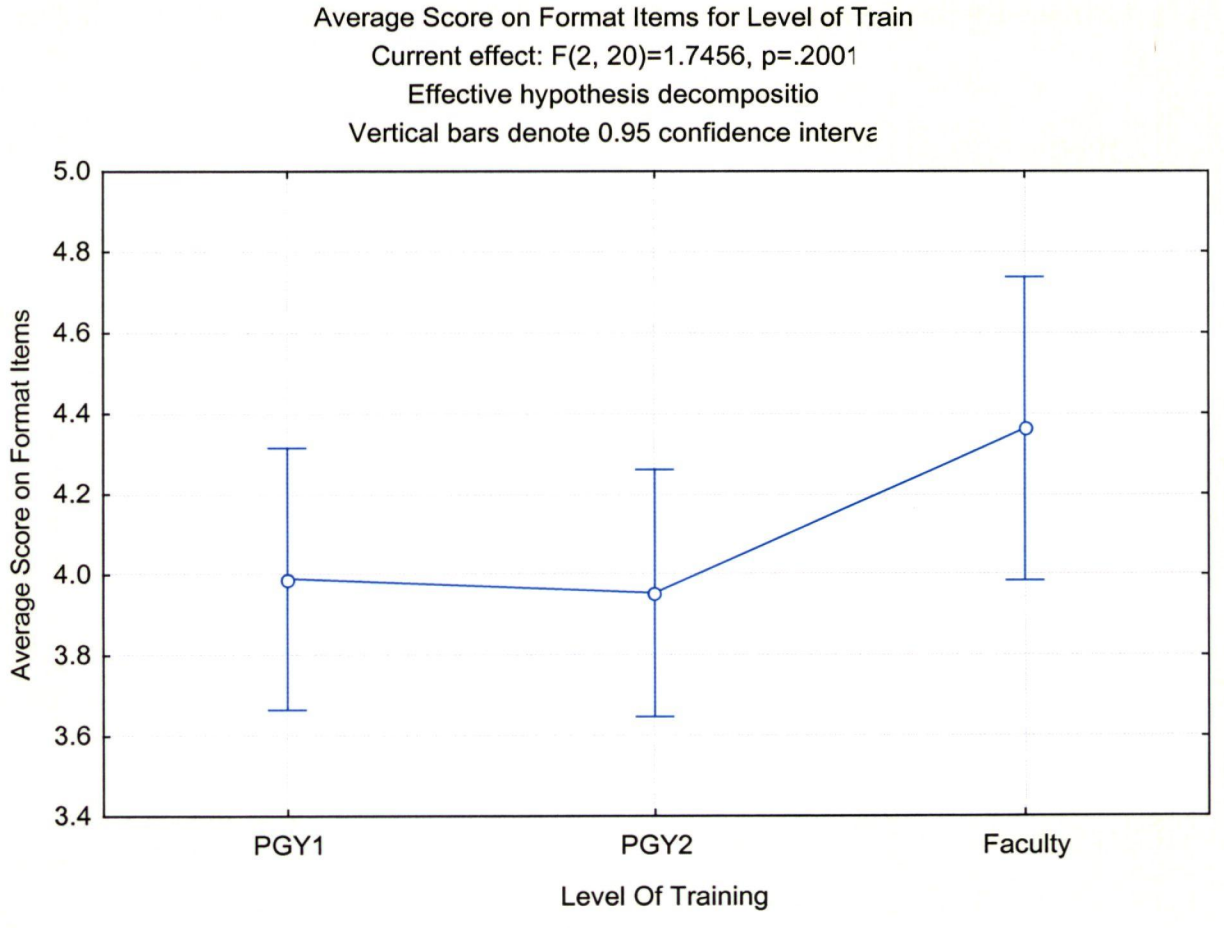


Figure 6: Average Score on Data Collection Items for Level of Training in the OSCE

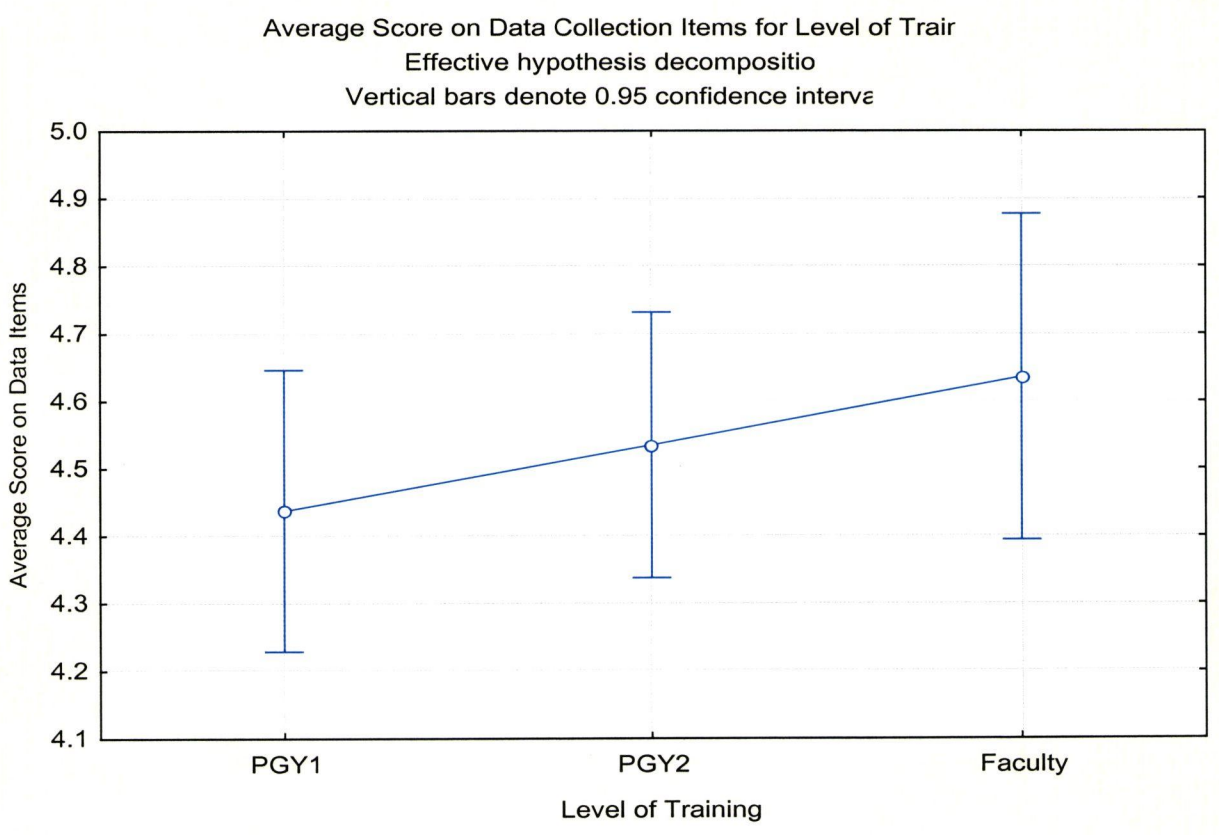


Figure 7: Average Scores on Educational Items for Level of Training in the OSCE

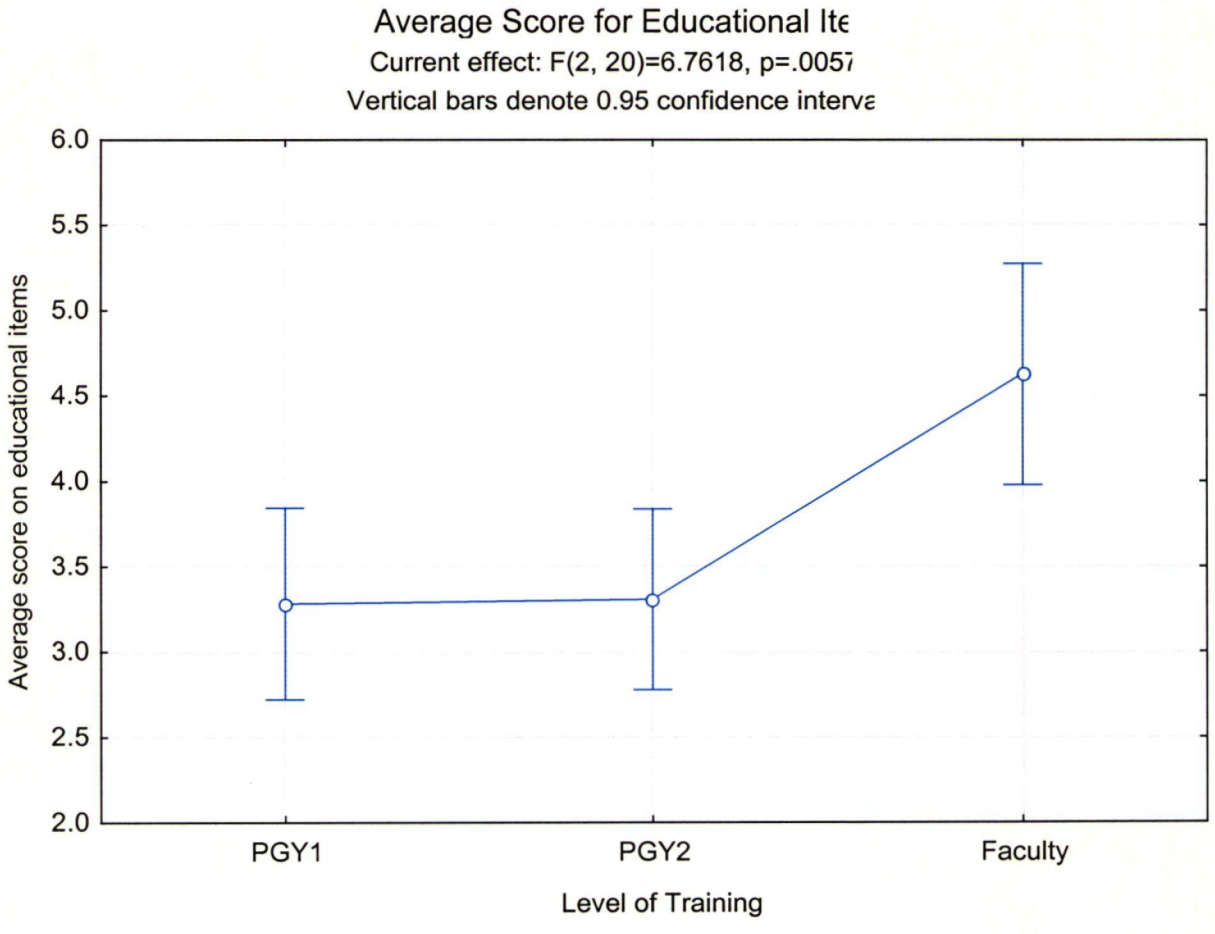


Figure 8: Average Scores on Management Items for Level of Training in the OSCE

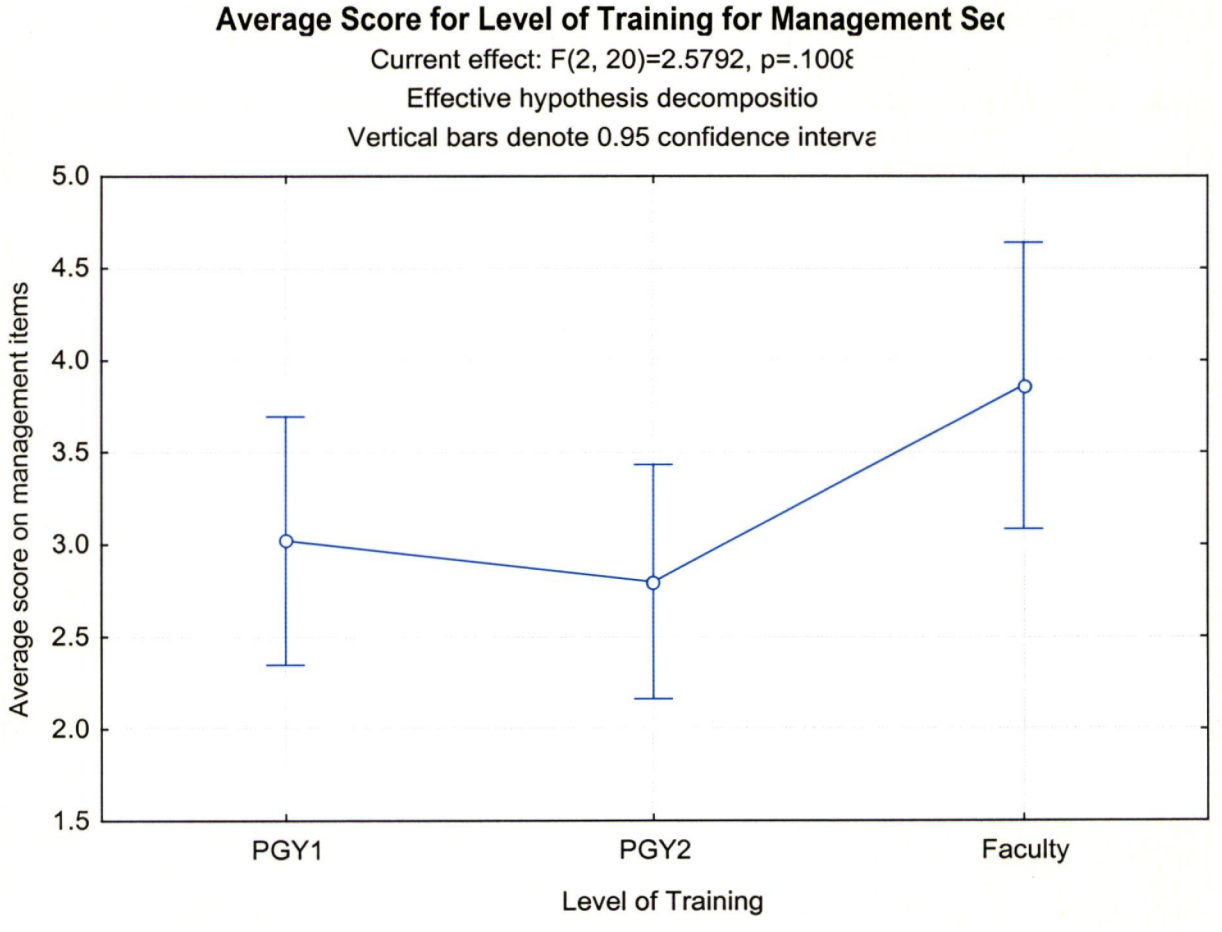


Figure 9: Average Scores for Communication Items for Level of Training in the OSCE

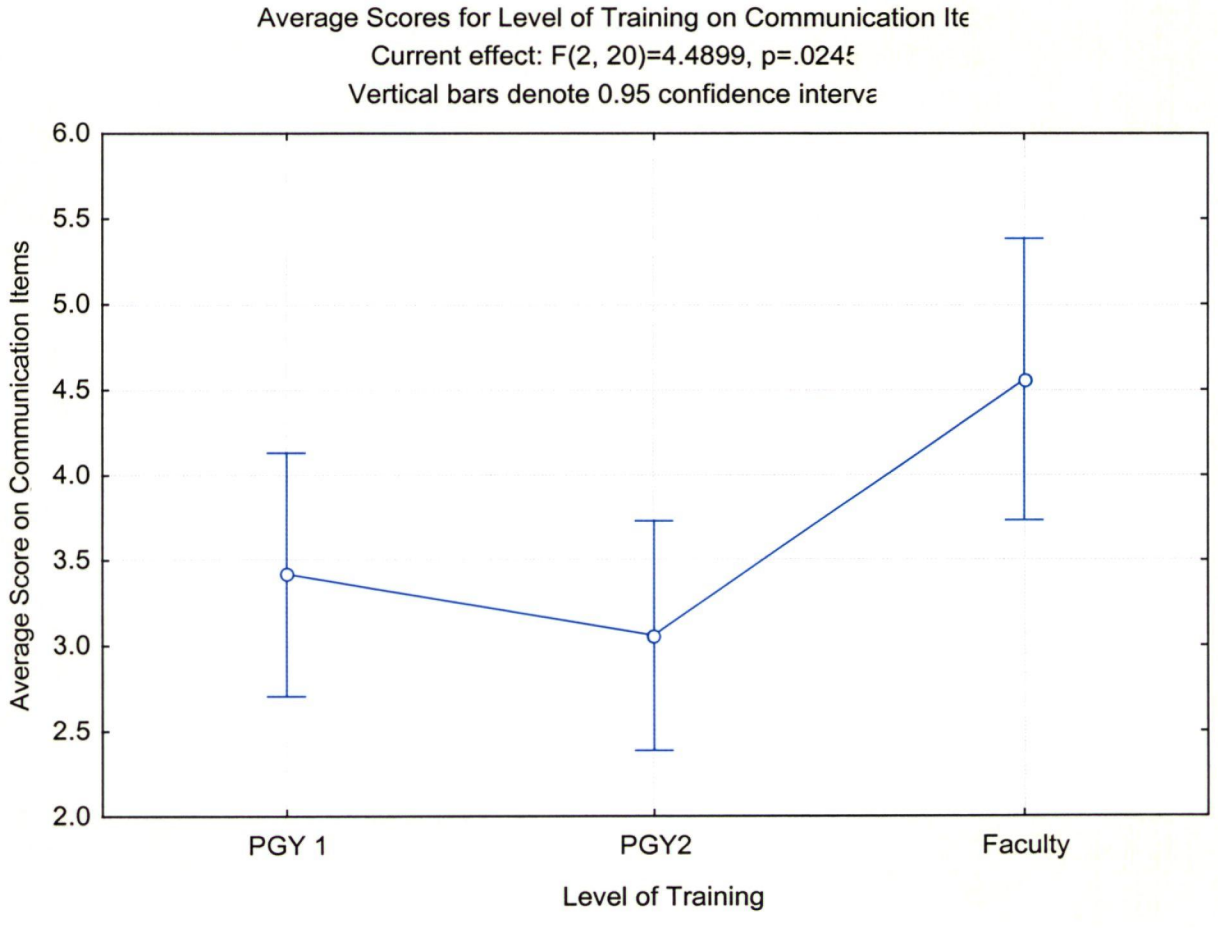


Figure 10: Average Global Score for Level of Training in the OSCE

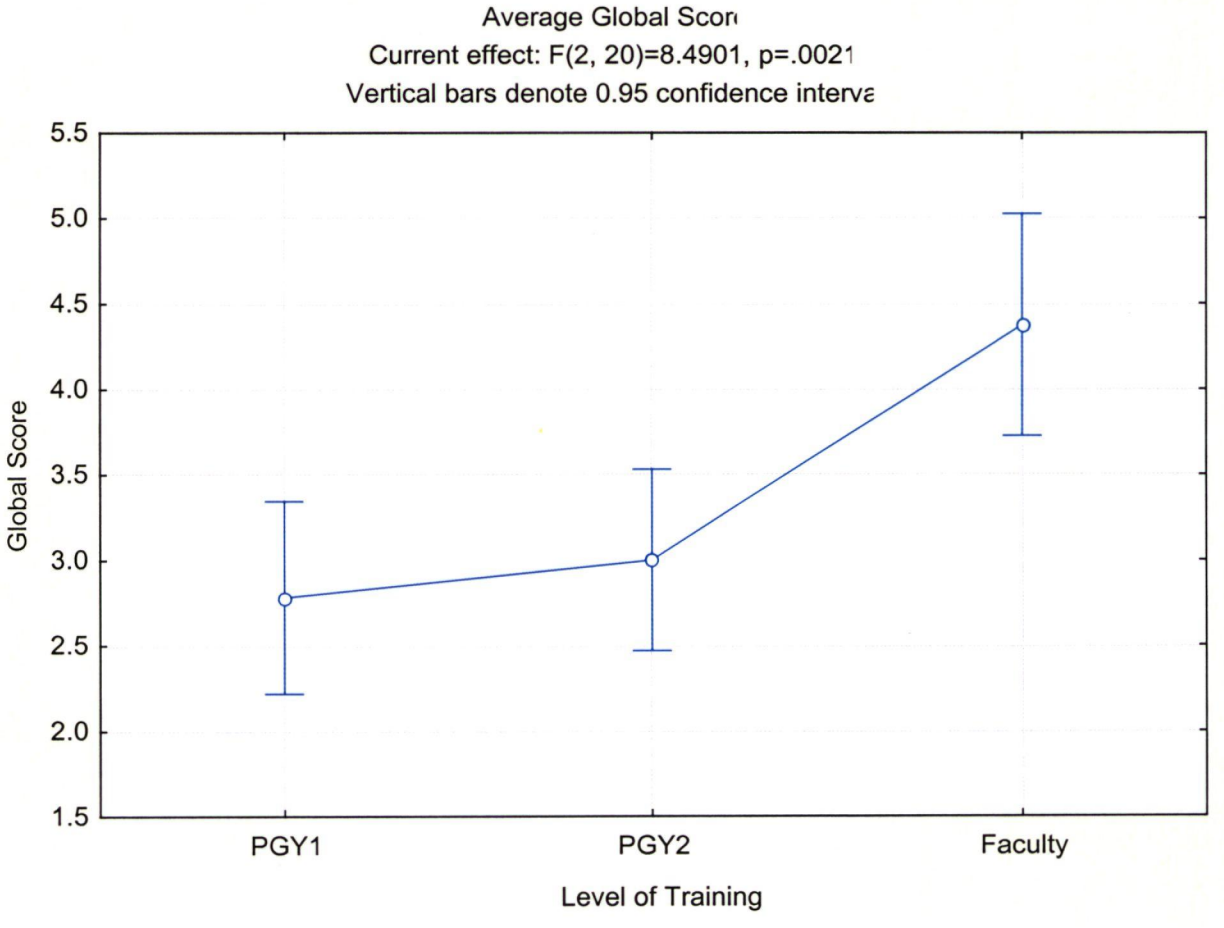
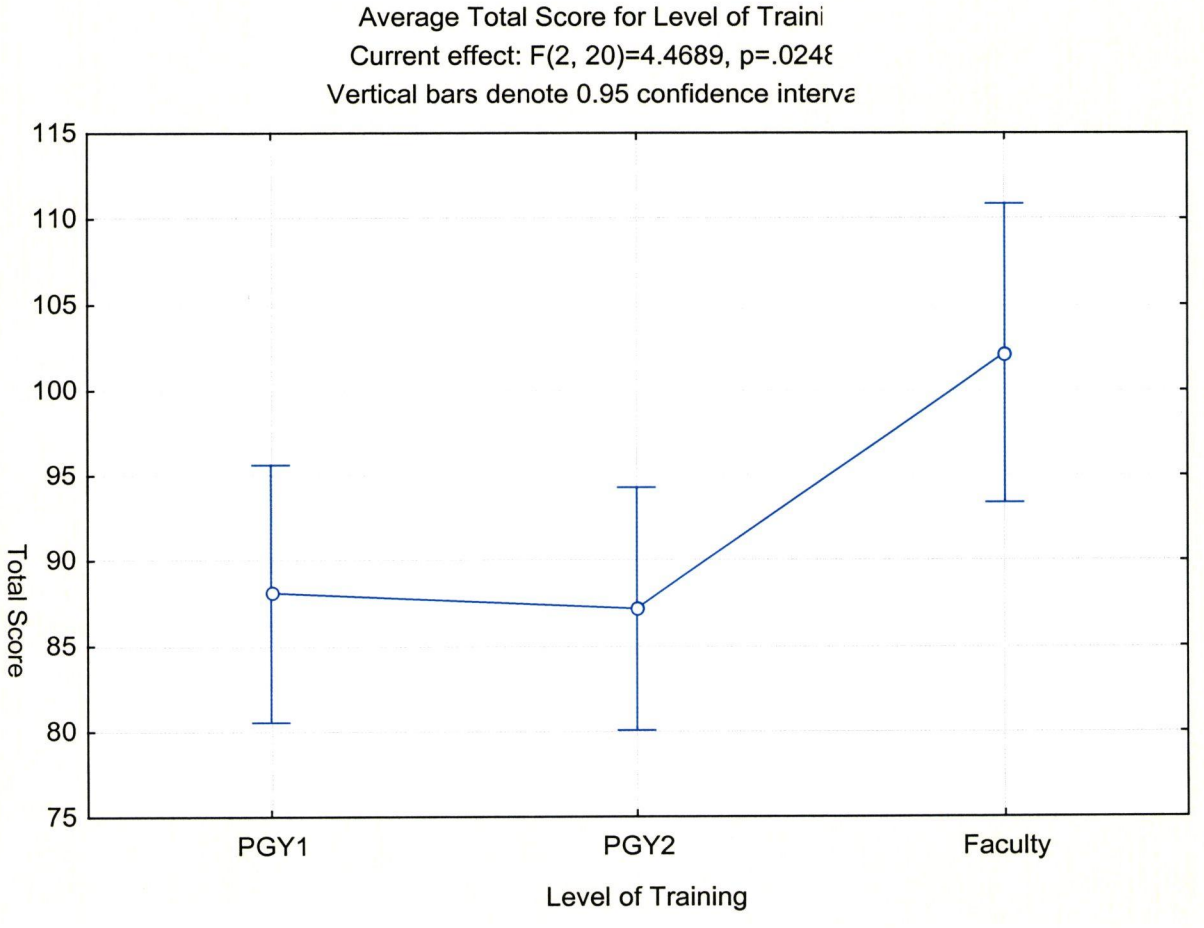


Figure 11: Average Total Score for Level of Training in the OSCE



V. DISCUSSION

The modified CLAT is a reliable and valid instrument for the evaluation of consultation letters in Internal Medicine. However, case specificity requires that multiple letters be scored using modified CLAT in order to obtain a stable measure of a consultant's ability. In addition, the data from the study suggest that the modified CLAT evaluates only one facet of the construct of a consulting internist and that multiple different evaluation formats should be used in order to sample the construct. Despite this limitation, the modified CLAT has an important role to play in both informal feedback and evaluation of trainees.

1. Reliability of the Modified CLAT and Item Analysis

Both the original CLAT and the modified CLAT were highly reliable instruments with Cronbach's α coefficients of 0.91 for both instruments. Though the reliability of the modified CLAT is identical to that of the original CLAT this does not mean that the modification of the original CLAT was for naught. Since the reliability of an instrument is dependent on the variance of scores, it will vary with the population sampled. The only way to show that the modified CLAT has improved reliability over the original CLAT would be to use both these instruments on the same population of letters.

The overall reliability of the modified CLAT is higher than the reliability both the SAIL instrument[6], which had an overall reliability of 0.72 and Myers et al[10]instrument, which had a Cronbach's α coefficient of 0.79. This high degree of reliability is one of the prerequisites for the use of the CLAT in summative examinations.

The reliability of an instrument is not a static number. Reliability is changed by the variance of the scores (which is related to the number of items and candidates

sampled as well as their range of ability). An instrument can have a high reliability coefficient in examination setting where there is a broad range of aptitude (high variance of scores). The same instrument will be less reliable in a situation where all the candidates are of equal aptitude (and the variance of the test is decreased). The discrepancy between the reliability of the original and modified CLAT is related to the sample size that affects the variance of the scores. Using the Spearman Brown equation a doubling of the sample size of judges (two to four) in the oral examination would give an estimated reliability of 0.88. This number is comparable to the reliability of 0.91 derived from the OSCE station.

The reliability of individual sections was decreased compared to the total reliability of the instrument (table 15, table 23). This is in part a function of the number of items assessed in each section compared to the total number of items assessed in the modified CLAT as well as the variance of the scores. Using data derived from the OSCE station, the sections on education and communication had the highest reliability despite small number of items in each section (table 23). The scores on these two sections had a large variance and were significantly different between the faculty and the residents.

The reliability of the data and management sections were the lowest with Cronbach's α coefficients of 0.60 and 0.65 respectively. The variance on the data section was the smallest contributing to its decreased reliability. The reliability coefficients for the data and management sections are acceptable for an individual section. However, they are low enough that items cannot be removed from these sections without significantly decreasing the reliability.

The initial CLAT was modified based on item analysis to improve the reliability of the instrument (tables 16 to 20). Analysis of the individual items in the data section revealed that item 16 detracted from the reliability of the data section (table 24). This item assessed the interpretation of ancillary data. It was only tested in the OSCE station, as ancillary data was not available for the oral examination. Its poor performance may be related to the set up of the station since the participants were given the EKG interpretation and the FVC and FEV_1 from the patient's spirometry. They were not given the EKG nor the lab print out for the spirometry as it was felt that there would not be enough time for the station if the additional tasks of interpretation were added. One could make a reasonable argument that the ability to interpret data was not adequately assessed in the OSCE station and that this contributed to the poor performance of the item.

In the data section, the original item 14 was split into 2 items, one assessing medications and another assessing drug allergies. These modified items contributed significantly to the overall reliability of the section in the data section in the OSCE.

In the management section, item 18 performed poorly in the pilot study but contributed significantly to the reliability of the section in OSCE. The pilot study used stable outpatients with well-defined stable medical problems. These patients were not diagnostic dilemmas. Thus, investigations were less relevant in the overall management of the patient. Item 18 performed well in the OSCE. The OSCE station dealt with a patient with shortness of breath on exertion and risks for both cardiac and pulmonary disease. The scenario was purposefully set up to assess the candidate's management of ambiguous data that required further investigations. In this setting, item 18 contributed significantly to the reliability of the section.

Item 20 in the original CLAT dealt with non-pharmacological management of medical problems. In the pilot study it detracted from the overall reliability of the section. This item assesses an important and often forgotten aspect of the medical management of a patient. The residents performed poorly on this item, which resulted in decrease item variance. The item was not tested in the OSCE since the station dealt with diagnosis and investigations and did not focus on management. Thus this item should remain, as it has not been adequately tested to confidently exclude it as an unreliable item.

Though the original CLAT is a highly reliable instrument, there is still significant variation judge to judge (table12). A Pearson's correlation coefficient was calculated to look at the correlation in the scores between the judges for each section. The Pearson's correlation coefficient is an index of the linear relationship between the two variables and will underestimate the relationship if the variables are curvilinearly related. In the case of the CLAT, there is no reason to suspect that the relationship between the scores of the judges would be non-linear. The Pearson's correlation coefficient does not give us information about the tendency of the judge to act as either a hawk or dove. Thus, there may be a strong correlation between judges even if one of the judge is a dove and the other is a hawk. This effect can be determined using generalisability theory. This statistical method was not applied due to the small sample size (see limitations of the study).

The homogeneity of the group and the variance of the scores also influence the correlation coefficient. A judge that marks to the mean will not use the full range of scores, which will result in decreased variance. Likewise if a group is homogeneous with little variance in scores the correlation coefficient will be decreased. In the case of the

pilot study, only letters by PGY3 residents were marked. This is a relatively homogeneous group, which would result in decrease in the correlation coefficient.

The Pearson's correlation coefficient for the total instrument ranged from 0.50 to 0.75. This is indicative of moderate positive correlation. There was considerable range in the correlation coefficients between sections. Certain sections such as the management sections had high degree of correlation. This may reflect a consensus on the part of the physicians as to the key components in the management of medical patients. For example, in the case of a stable patient with congestive heart failure, internists may disagree as to what is important with respect to data collection. However, all would agree that angiotensin converting enzyme inhibitors, spironolactone and beta-blockers play a pivotal role in the management of these patients. In this case one would expect a higher correlation between internists for the management section compared to the data section. This postulate could be tested using two OSCE stations: one station dealing with a medical problem where there is clear consensus on the best medical practice for management (i.e. congestive heart failure) and another station where there is no consensus as to the best management (i.e. idiopathic pulmonary fibrosis). In the latter case, one would expect less correlation between markers, as there is no gold standard that they can agree upon.

The sections dealing with the educational value of the letter, communication and the global scale had poor correlation coefficients between judges. The poor correlation on the global rating item is most likely a function of a small degree of variance (i.e. a homogeneous population of scores) of this single item section. One would predict that the correlation would improve if letters representing a broad range of ability were used. The

poor correlation on the communication section may also be a function of narrow range of scores with the majority of residents performing poorly on this section (mean score of the section 1.62). It is less clear why there was poor correlation between the judges on the section pertaining to the educational value of the letter. One of the items in this section in the original CLAT performed poorly and detracted from the overall reliability of the section. This would contribute to the poor degree of correlation between judges. In addition, the educational value of the 'ideal letter' is poorly defined and standardised. This would contribute to the poor correlation between judges, as each judge may be using different gold standards.

Based on the data and the above discussion should the modified CLAT be changed to improve its performance? It already has excellent reliability and item analysis suggests that all the items contribute to the reliability of the instrument. Factor analysis revealed that many of the items were tightly correlated. This raises the possibility of simplifying the instrument by eliminating some of the items. Though this could be done from a reliability point of view, this would result in a change of the relative weighting of each section and hence the content validity (vide infra). For these reasons, the modified CLAT should remain unchanged.

2. Evaluation of Face, Content and Construct Validity of the CLAT

Validity refers to the extent that the instrument is measuring what it is supposed to measure. There are four types of validity: face validity, content validity, concurrent validity and construct validity. Analysis of the studies using the modified CLAT shows that it is valid instrument for the assessment of the consultation letter in Internal

Medicine. In order to be valid, the scores of an instrument must first be reliable. The scores from the modified CLAT were found to be reliable.

Face validity relates to the overall appearance of the instrument and the importance of what it is trying to assess. The consultation letter is the main form of communication between the consultant and the referring physician in the outpatient setting. Its documented importance in physician to physician communication gives it face validity.

Content validity refers to the selection and the weighting of items relative to the total instrument. In the setting of the consultation letter the selection of items is dependent on the perceived role of the letter. During the development of the CLAT, items were selected based on the premise that the consultation letter plays a major role in the ongoing management of patients. This point of view is very similar to the one used in the SAIL as evidenced by their global rating item ("This letter clearly conveys the information I would like to have about the patient if I were the next doctor to see him/her"[6]).

The CLAT does not try to assess the consultation letter from a medical-legal point of view, which would favour an all inclusive data collection. Though important from a legal point of view, this can result in a cumbersome and lengthy letter, which makes the letter less user friendly in a busy clinical practice.

There are strong psychological reasons for minimising the weight of the data collection relative to the other components of the letter. Extensive weighting of data collection penalises experts who often arrive at the diagnosis through pattern recognition or schemes[19;20]. This allows the clinician to make an accurate diagnosis without obtaining

all the data. In the OSCE, though there was a significant difference in the overall scores (figure 11) between faculty and residents, there was no difference in the scores of the data section (figure 6). The lack of difference between residents and faculty in this section suggests that the weighting of this section does not penalise the expert by putting too much weight on the data section.

In contrast to data collection, management is a very important component of the construct and needs to be weighed heavily[4;17;29]. Because of case-specificity of medical knowledge it is important that the management of each problem be assessed separately as in the case of the CLAT. This feature of the CLAT makes it unique compared to the other generic evaluation tools. It enhances the utility of the CLAT as a teaching tool by allowing accurate feedback to an individual candidate on different medical problems. The relative weighting of the management section increases for letters pertaining to patients with multiple medical problems. This contributes to the content and construct validity of the CLAT as patients with multisystem disease are becoming more common and their care is an important aspect of the role of a general internist.

In the OSCE station there was no difference between the scores of the faculty and the residents in the management section (figure 8). This does not detract from the validity of the instrument. The OSCE station represented a common diagnostic problem of shortness of breath. All of the residents would have had considerable experience in this area by the time they took the OSCE. Thus, the residents were not novices in dealing with this type of problem. This may explain why there was no difference in the management scores for level of training.

The CLAT assesses the educational value of the letter. Family physicians feel that this is an important role of the letter[28]. It should be assessed in order to give the instrument content validity. The educational aspect of the consultation letter was not specifically addressed in the SAIL[6] or in Myer's et al[10] instrument. Both contained one item, which assessed whether the rationale for the management changes was provided. The modified CLAT has 2 items pertaining to the educational value of the letter. The faculty scored significantly higher than the residents did on this section (figure 7). There are several explanations for this difference in scores. One is that the residents may know what to do but may not understand why they are doing a specific procedure. This is in keeping with the hierarchy of knowledge. More junior physicians are able to apply their knowledge to the case whereas the expert is able to able to analyse, synthesise and evaluate.

Another explanation is that the faculty is aware of the importance of the educational value of the letter whereas the residents are not. An educational and informative letter from a consultant could have the added benefit of encouraging the referring physician to send more patients to the consultant. Thus, an educational letter has the potential to improve referral practices. Whereas this would be an important aspect of the letter for the consultant, the resident may not be aware of its importance.

Faculty also scored significantly higher than the residents did on the section pertaining to communication of follow up details to the referring physician (figure 9). This difference most likely reflects a better understanding, by the faculty, of the role of the letter in patient management. Residents rarely follow patients longitudinally and thus do not gain an understanding of the importance of co-ordinated follow up between

physicians. This aspect of care becomes quite apparent in practice when the consultant sees patients in follow up and realises that his/her suggestions for the patient's care were not done at all or something else was done instead. The results of the pilot study reinforce the need to teach residents the importance of documenting in the letter who does what when.

Thus, the results of the OSCE station support the modified CLAT as a valid tool for the assessment of the consultation letter in Internal Medicine (table 27). If an instrument is valid then experts (who are experts because they possess a greater amount of the construct) should score higher than non-experts and novices. On the OSCE station the faculty scored significantly higher than did the residents on the overall score. This lends support to the construct validity of the CLAT.

However, because of case specificity of medical knowledge and the multiple sources of variance in the CLAT score, it should not be used as the sole tool for the evaluation of the construct of an Internist.

3. Sources of Variance

There are multiple sources of variance in the evaluation of the consultation letter. These sources include: the ability to act as a consultant, the ability to express one's opinion in a written form, the interaction between the consultant and the case, the interaction between the judge and the case, the tendency of the judge to mark as a hawk or a dove, and the internal consistency of the judge case to case. The ability to express a consultant's opinion is the construct that the CLAT is trying to assess. However, the other sources of variance also contribute to the score on the CLAT and have an impact on the stability of the measurement.

The interaction between the consultant and the case is related to case specificity. Case specificity makes evaluation in internal medicine difficult and time consuming. It requires that multiple facets of internal medicine be tested in order to obtain a stable measurement of a candidate's ability. This applies to the consultation letter. In the pilot study each resident dictated 2 letters. The letters pertained to different patients with different medical problems. As one would predict, there was little correlation between the scores on the first letter and the second one. This poor correlation between the letters was seen in the sections pertaining to educational value of the letter, data collection and management. This poor correlation can be explained by case specificity.

There was no correlation between the scores pertaining to format from the first letter to the second. One would have postulated that the ability to format a letter is a generic skill and would be stable from letter to letter. The lack of correlation in the format section may in part be related to the 6 months that elapsed between the dictation of the first and second letter. During this time the residents received feedback on their initial consultation letter including the formatting of the letter. The feedback may have influenced their dictation style during the second oral exam and may explain the poor correlation. However, if this were the case, then one would expect an improvement in the mean score for the format which did not occur.

The scores on the section pertaining to communication of follow up information were highly correlated between the letters. The residents performed poorly on the communication section in both letters. In the pilot study, the ability to explicitly outline follow up issues appears to be a generic skill and not related to the case.

The interactions of the judges can be overcome to some extent by using multiple judges. The downside of using multiple markers is the manpower requirements. The use of a single marker is not a concern when the CLAT is used for informal feedback in clinic setting. However, multiple judges would be required in a high stakes examination.

The effect of these sources of variance can be determined using generalisability theory. Unfortunately, the current study did not have enough letters per resident to calculate the effect of the sources of variance and to determine the number of letters that must be sampled in order to obtain a stable measurement. Crossely et al[6] used generalisability theory to predict the number of letter required to obtain a stable measurement using the SAIL. They calculated that 6 judges correcting 10 letters would give a stable and reliable (G>0.8) assessment of a consultant's ability. Using generalisability theory, Myers et al[10] estimated that 5 letters should be assessed in order to obtain a reliability of 0.6. (It is not clear in the article whether two markers per letter were required in order to obtain this degree of reliability). Thus, no matter how valid and reliable the instrument and how stable over time the ability of the consultant, multiple samples are required in order to overcome case specificity in Internal Medicine.

4. Concurrent Validity

Concurrent validity compares results from one type of evaluation to another format of evaluation. A valid instrument should show concurrent validity with other forms of validated assessments of the construct. From a medical educational point of view, the ultimate goal of post-graduate evaluation is to assess the ability to act as a consulting internist. The ability to write a consultation letter is part of this bigger construct.

The scores between an oral examination and the CLAT were compared to look for evidence of concurrent validity. Unfortunately, the oral examination format used has multiple sources of variance including non-standardised patients and different judges for the letters and the oral examination per candidate. Though the use of standardised patients eliminates some of the sources of variance, it does so at the expense of face validity. The ability to examine and detect abnormal physical examination findings is an important aspect of being an internist. The use of actors as standardised patients eliminates this aspect and detracts from the face validity of the examination. One possible solution to this is to have a stable patient with a real disease and physical examination findings interviewed by all the candidates. This solution would be quite burdensome for the patient even with the small number of residents that participated in the current study. It was not felt to be a practical option.

Table 25 demonstrates some of the difficulties in using non-standardised patients. There was not complete agreement between both oral examination judges and the residents as to the medical problems that were felt to be important. This discrepancy was made worse by the use of patients with multiple medical problems and would most likely be eliminated to a great degree by using patients with a single well-defined medical problem (i.e. a patient with bicuspid aortic valve). In addition, in real life the patient would ideally have a referral letter that would explicitly outline the medical issues for which that the referring doctor wanted the consultant's opinion. This would obviously guide the consulting physician's discussion and management of the case.

Overall, there was poor correlation between the oral examination and the CLAT. This is despite the high reliability of both the oral examination and the CLAT. Based on

test statistics one would expect a maximum correlation of 0.81 (the maximum correlation equals the square root of the product of the reliabilities of the individual tests). The poor correlation on the section pertaining to the format (brevity of the letter compared to brevity of the oral presentation) and the global score is in part a function of low number of items associated with little variance. The correlation coefficients were highest for the sections on data collection and management and indicative of moderate correlation.

The poor degree of correlation suggests that the oral examination is measuring something different than the CLAT. Part of the difficulty in assessing a candidate's ability based on the consultation letter is that the marker is unsure whether the resident's editing skills or his/her knowledge base is determining the content of the letter. In the oral examination the examiners has the ability to question and probe the candidate's knowledge. This allows the examiner to test for the depth of understanding of the candidate especially if the candidate omits items in the case presentation. This may explain the higher global scores on the oral examination compared to the letter. It suggests that, in letters dictated by senior residents, editing may play a more important role than knowledge in determining the content of the consultation letter. If this is indeed the case it reinforces the importance of teaching the residents what to include in a letter. This observation may not be true in the case of letters dictated by true novices (such as clinical clerks or PGY1 residents) where knowledge may become a factor in determining the content of the letter.

These observations are similar to observations from Keely'set al[34] study looking at the correlation between various scores on an OSCE. In their study, there was no correlation between the letter writing score and a station assessing the physical

examination in the same domain. There was a modest correlation between the consultation letter score and a station assessing communication skills (r =0.37) and a moderate correlation between the overall OSCE score and the letter score (r = 0.57). They postulated that letter writing required a higher degree of knowledge (i.e. the ability to synthesise, evaluate) than did stations that simply assessed content.

The results from the oral examination compared to the CLAT reinforce the concept that multiple facets of internal medicine must be evaluated in order to obtain a fair and stable assessment of the construct. It suggests that resident's knowledge pertaining to a medical problem may be underestimated if only the CLAT is used. Further study is required to assess the relative importance of knowledge in the formulation of the consultation letter.

5. Limitations of Current Study

The major limitations of this study are the small sample size and the lack of inclusion of letters dictated by true novices. The small sample size precluded an in-depth analysis of multiple letters by the same resident. Thus, the effect of the multiple sources of variance in the scores could not be calculated. This would have allowed a calculation of number of letters that need to be marked in order to obtain a reliable assessment of a candidate's ability to dictate a letter.

The omission of novices in this study resulted in a decrease range in scores, which can have an adverse effect on the correlation coefficients. Inclusion of letters from novices would have most likely increased the reliability of the CLAT by increasing the variance in scores.

Minor weaknesses in the study include the use of multiple examiners in the oral examination and the use of non-standardised patients. Despite these sources of variation the reliability of the oral examination was excellent.

6. Perceived Role of the CLAT in Medical Education

The CLAT has a role to play in the evaluation of Internal Medicine residents and consultants. Its major advantage over the other two generic evaluation tools is the ability to assess the management of individual medical problems in patients with multi-system disease. This is an important attribute in view of the case specificity of medical knowledge. It allows more detailed feedback, which is a key component in the cybernetic cycle.

The evaluation of consultation letters should be a key component of the assessment of any ambulatory care rotation. Ambulatory rotations offer unique challenges in medical education. They are often characterised by significant time pressure with little time to discuss patients in a busy clinic[9]. The routine evaluation of the letters dictated by the resident affords an excellent opportunity to provide feedback. In addition, the systematic review of the consultation letter may reveal knowledge gaps that can then be probed to see if this is indeed the case. Structured feedback on a resident's consultation letter allows the faculty to teach and role model some of the professional attributes of a consultant including: the importance of ongoing education, the value of clear communication and the responsibilities of consultant physicians with respect to the referring physicians in the management of patients. This is becoming increasingly important as more and more medical problems are being treated in the outpatient setting.

With its excellent reliability the CLAT can also be used in formal examination settings. In this setting, in order to minimise sources of variance, standardised information (in the form of the data provided or the use of standardised patients) should be used to maximise reliability and ensure that all the residents are presented with the same clinical challenge.

In summary the CLAT is a valid and reliable tool for the assessment of an internist's ability to write a consultation letter. However, its use in a formal evaluation setting should be combined with other assessment modalities to ensure that the construct is adequately sampled.

VI. CONCLUSION

The modified CLAT has a high degree of reliability and is a valid tool for assessing consultation letters in Internal Medicine. It can be used in both informal settings such as ambulatory care blocks or in formal examination settings. Further research is required in the evaluation of the modified CLAT. This would include its use on multiple letters dictated by multiple faculty and residents in order to understand further the sources of variance in the scores. The role of knowledge in the generation of the consultation letter needs to be explored. The contribution of the consultation letter as part of the construct of an internist needs to be further defined. This will allow its appropriate use in the assessment of training. No work has been done in the area of the assessment of ongoing competency and performance in practising physicians using their consultation letters. The large number of letters dictated by consulting internists makes this an attractive area to study. Finally, the impact of an ideal consultation letter on the management of patients should be studied to see if an excellent letter could indeed effect outcome.

Reference List

1. The Royal College of Physicians and Surgeons of Canada. Skills for the New Millennium: Report of the Societal Needs Working Group. Can Meds 2000 project. 1996. Ottawa, RCPSC.

Ref Type: Report

2. Pullen IM, Yellowlees AJ. Is Communication Improving between General Practitioners and Psychiatrists? BMJ 1985;290:31-33.

3. Keely EJ, Dojeiji S, Myers KA. Writing Effective Consultation Letters: 12 Tips for Teachers. Medical Teacher 2002;24:585-89.

4. Babington S, Wynne C, Atkinson CH, Hickey BE, Abdelaal AS. Oncology Service Correspondence: Do We Communicate? Australasian Radiology 2003;47:50-54.

5. McPhee SJ, Lo B, Saika GY, Meltzer R. How Good Is Communication between Primary Care Physicians and Subspecialty Consultants. Archives Internal Medicine 1984;144:1265-68.

6. Crossley JG, Howe A, Newble D, Jolly B, Davies HA. Sheffield Assessment Instrument for Letters (SAIL): Performance Assessment Using Outpatient Letters. Medical Education 2001;35:1115-24.

7. The Royal College of Physicians and Surgeons of Canada and the College of Family Physicians of Canada Task Force on the Relationship Between Family Physicians and Specialists/Consultants. page14. 1993. Ottawa, RCPSC/CFCP.

Ref Type: Report

8. McCain GA, Molineux JE, Pederson L, Stuart RK. Consultation Letters as a Method for Assessing In-Training Performance in a Department Of Medicine. Evaluation and the Health Professions 1988;11:21-42.

9. Laidely TL, Braddock III CH. Role of Adult Learning Theory in Evaluating and Designing Strategies for Teaching Residents in Ambulatory Settings. Advances in Health Sciences Education 2000;5:43-54.

10. Myers KA, Keely EJ, Dojeiji S, Norman GR. Development of a Rating Scale to Evaluate Written Communication Skills of Residents. Acad Med 1999;74(10): S111-113 1999;74:S111-S113.

11. Sandler G. Costs of Unnecessary Tests. BMJ 1979;2:21-24.

12. Spooner MK. News: UK Physicians May Be Asked to Share Consultation Letters With Patients. Canadian Medical Association Journal 2003;168:1030.

13. Metcalfe D, Sischy D. Patterns of Referral from Family Practice. New York State Journal of Medicine 1973;73:1690-94.

14. Williams TF, White KL, Fleming WL, et al. The Referral Process in Medical Care and the University Clinic's Role. Journal Of Medical Education 1961;36:899-907.

15. Westerman RF. A Study of Communication between General Practitioners and Specialists. Br J Gen Prac 1990;40:445-49.

16. Kentish R, Jenkins P, Lask B. Study of Written Communication between General Practioners and Departments of Child Psychiatry. Journal of the Royal College of General Practioners 1987;37:162-63.

17. Tattersall MH, Griffin A, Dunn SM, Monaghan H, Scatchard K, Butow PN. Writing to Referring Doctors after a New Patient Consultation. What is Wanted and What Was Contained in Letters From One Medical Oncologist. Aust NZ J Med 1995;25:479-82.

18. McConnell D, Butow PN, Tattersall MH. Improving the Letter We Write: An Exploration of Doctor-doctor Communication in Cancer Care. British Journal of Cancer 1999;80:427-37.

19. Schmidt HG, Norman GR, Boshuizen HP. A Cognitive Perspective on Medical Expertise: Theory and Implications. Acad Med 1990;65:611-12.

20. Bordage G. Elaborated Knowledge: A Key to Successful Diagnostic Thinking. Acad Med 1994;69:883-85.

21. Reitman Olson J, Rueter HH. Extracting Expertise from Experts: Methods For Knowledge Acquisition. Expert Systems 1987;4:152-68.

22. Eva KV, Brooks LR, Norman GR. Forward Reasoning As A Hallmark Of Expertise In Medicine: Logical, Psychological, Phenomenological Inconsistencies. In: Shohov SP, editor. Advances in Psychology Research, vol. 8, first ed. Nova Science Publishers; 2002.

23. Ridderikoff J. Medical Problem-Solving: An Exploration of Strategies. Medical Education 1991;25:196-207.

24. Norman GR, Tugwell P, Feightner JW, Muzzin LJ, Jacoby LL. Knowledge and Clinical Problem-Solving. Medical Education 1985;19:344-56.

25. Kuskniruk AW, Patel VL, Marley AA. Small Worlds and Medical Expertise: Implications for Medical Cognition and Knowledge Engineering. International J Med Informatics 1998;49:255-71.

26. Boudreau D, Tamblyn R, Dufresne L. Evaluation of Consultative Skills in Respiratory Medicine Using a Structured Medical Consultation. Am J Respir Crit Care Med 1994;150:1298-304.

27. Newton J, Eccles M, Hutchinson A. Communication between General Practitioners and Consultants: What Should Their Letters Contain? B. BMJ 1992;304:821-24.

28. Keely , E. J, Dojeiji , S, and Myers, K. A. Faculty Guide for Learning to Write Effective Consultation Letters. 1999. Toronto.

Ref Type: Serial (Book,Monograph)

29. Dojeiji S, Marks M, Keely EJ. Referral and Consultation Letters: Enhancing Communication between Family Physicians And Specialists. Clin Invest Med 1997;20:S49.

30. van Walraven C, Duke SM, Weinberg AL, Wells PS. Standardized or Narrative Discharge Summaries. Which Do Family Physicians Prefer? Canadian Family Physician 1998;44:62-69.

31. Lloyd BW, Barnett P. Use of Problem Lists in Letters between Hospitals Doctors and General Practioners. BMJ 1993;306:247.

32. Rawal J, Barnett P, Lloyd BW. Use of Structured Letters to Improve Communication between Hospital Doctors and General Practioners. BMJ 1993;307:1044.

33. Stillman PL, Regan MB, Haley HL, Friedman M, Setnick AI. The Use of a Patient Note to Evaluate Clinical Skills of First-Year Residents Who Are Graduates of Foreign Medical Schools. Acad Med 1992;67:S57-S59.

34. Keely EJ, Myers KA, Dojeiji S. Can Written Communication Skills Be Tested in an Objective Structured Clinical Examination Format? Acad Med 2002;77:82-86.

35. Streiner DL, Norman GR. Health Measurement Scales. A Practical Guide to Their Development and Use, second ed. 1995. p. 69-83.

APPENDIX A: SAIL

Sheffield Assessment Instrument for Letters (SAIL) October 1999

CIRCLE THE APPROPRIATE OPTION: PLEASE READ ACCOMPANYING NOTES

SpR: Code no..... **Source of letter:** Self-selected/ Random

Type of letter: New-patient/ Follow up/ Referral

Complexity of case: Low/ Average/ High *In your own view*

Assessor: Initials..... **Status:** Consultant/ GP/ Peer/ Self

PROBLEM LIST

1. Is there a problem list? Yes/ No
2. Any obvious and significant problems omitted? Yes/ No
3. Any obviously irrelevant problems listed? Yes/ No

If no problem list answer yes to 2 and 3

HISTORY

4. Current problems/ well-being clarified? Yes/ No
5. Documented history appropriate to the case? Yes/ No

EXAMINATION

6. Documented examin^{n} appropriate to the case? Yes/ No

6 - no examination appropriate and none documented equals yes

7 - no new conclusions appropriate and none documented equals yes

OVERALL ASSESSMENT

7. New conclusions recorded if appropriate? Yes/ No

MANAGEMENT

8. Clear plan of investig^{n}/ non-investig^{n}? Yes/ No/ NK
9. Are the reasons for the above made clear? Yes/ No
10. All (known) treatments clearly listed? Yes/ No/ NK
11. All doses clearly stated (in formal units)? Yes/ No/ NA
12. Explanation given for any changes to treatment? Yes/ No/ NA
13. Documentation of information shared with family? Yes/ No/ NA

If no drugs or changes answer NA to 11 and 12

13 - no information needs sharing and none documented equals yes

FOLLOW UP

14. Is it clear whether or not hospital follow-up planned? Yes/ No
15. Is the purpose of follow-up (if planned) clear? Yes/ No/ NA

CLARITY

16. Is there much unnecessary information? Yes/ No
17. Did the structure of the letter flow logically? Yes/ No
18. Was there any English you did not understand? Yes/ No

PLEASE MARK ON THE SCALE BELOW HOW COMPELTELY YOU AGREE WITH THE FOLLOWING STATEMENT:

"This letter clearly conveys the information I would like to have about the patient if I were the next doctor to see him/her"

1 2 3 4 5 6 7 8 9 10

NOT AT ALL COMPLETELY

APPENDIX B: Consultation Letter Evaluation Tool by Myers et al.

18271

Consultation Letter Evaluation

Resident ID ☐☐☐ Specialty ☐☐ PGY ① ② ③ ④ ⑤ *for office use only:* 023

Site ① ②

Rater ① ② ③

Section A

History of Presenting Illness	1. Identifies chief problem/reason for referral	no ○ yes ○
	2. Describes chief complaint	no ○ yes ○
	3. Identifies associated conditions relevant to the chief complaint	no ○ yes ○
Other Details of History	4. Identifies relevant past history	no ○ yes ○
	5. Lists current medications	no ○ yes ○
	6. Provides other history appropriate to the presenting problem (i.e. family history, psychosocial history, review of systems etc.)	no ○ yes ○
	7. Completeness of Section A	Incomplete ① ② ③ ④ ⑤ Complete
	8. Clarity/organization of Section A	Poor ① ② ③ ④ ⑤ Clear/organized
	9. Brevity of Section A	Wordy ① ② ③ ④ ⑤ Concise
	10. Overall rating of Section A	Poor ① Barely Acceptable ② Acceptable ③ Good ④ Optimal ⑤

Section B

Physical Examination	11. Describes physical findings relevant to the presenting problem	no ○ yes ○
	12. Completeness of Section B	Incomplete ① ② ③ ④ ⑤ Complete
	13. Clarity/organization of Section B	Poor ① ② ③ ④ ⑤ Clear/organized
	14. Brevity of Section B	Wordy ① ② ③ ④ ⑤ Concise
	15. Overall rating of Section B	Poor ① Barely Acceptable ② Acceptable ③ Good ④ Optimal ⑤

18271

Section C

for office use only: 0 2 3

Impression and Management

		no	yes
16.	Provides differential diagnosis	○	○
17.	Provides a management plan	○	○
18.	Provides a rationale for management plan	○	○
19.	States whether management plan discussed with patient	○	○
20.	States who is responsible for follow-up	○	○

21.	Completeness of Section C	Incomplete ①	②	③	④	Complete ⑤
22.	Clarity/organization Of Section C	Poor ①	②	③	④	Clear/organized ⑤
23.	Brevity of Section C	Wordy ①	②	③	④	Concise ⑤
24.	Overall rating of Section C	Poor ①	Barely Acceptable ②	Acceptable ③	Good ④	Optimal ⑤

Section D

25.	Overall rating of consultation letter	Poor ①	Barely Acceptable ②	Acceptable ③	Good ④	Optimal ⑤

Comments

Analysis of Writing Style of Consultation Letter

		no	yes
Words	Uses active voice	○	○
	Avoids jargon	○	○
	Avoids repetition	○	○
Sentences	One idea per sentence	○	○
	Less than three lines in length	○	○
Paragraphs	Contain one topic	○	○
	Less than 4-5 sentences	○	○
Overall Style	Uses headings appropriately	○	○
	Layout visually appealing	○	○

Comments

APPENDIX C

CONSULTATION LETTER ASSESSMENT TOOL (CLAT)

FORMAT

1. The assessment is clearly identified in the text*.
   SD D U A SA

2. The plan is clearly identified in the text
   SD D U A SA

3. I can quickly scan this letter to obtain key information
   SD D U A SA

4. The letter is succinct while maintaining an appropriate level of detail
   SD D U A SA

5. The tone of the letter is professional
   SD D U A SA

6. The paragraphs dealt with one idea
   SD D U A SA

EDUCATIONAL VALUE OF THE LETTER

7. This letter would contribute the referring physician's continuing education through its discussion of the patient's problem
   SD D U A SA

8. This letter provides an explanation for the recommendations
   SD D U A SA

9. This letter highlights areas of controversy or areas of new developments
   SD D U A SA

DATA SYNTHESIS

10. The chief complaint/reason for referral is clearly identified in the first paragraph of the letter
    SD D U A SA

* SD: strongly disagree, D: disagree, U: undecided, A: agree, SA: strongly agree

11. Patient's history

1	2	3	4	5
Rambling, foggy difficult to follow		easy to follow but not concise		concise, relevant

12. History of presenting complaint

1	2	3	4	5
Missing the majority of key positive or negative items		missing some pertinent key negative or positive items		all key items

13. Pertinent ancillary history

1	2	3	4	5
Missing key items such as PMHx, SHx, smoking status		all key items present		all key items present

14. Medications and allergies

1	2	3	4	5
Not listed or incomplete		listed but no doses		complete with dosage

15. Physical examination

1	2	3	4	5
Missing key items		most key items listed		concise/relevant to problem all key items listed

16. Ancillary tests (if applicable)

1	2	3	4	5
no information provided		information provided no interpretation		information provided with interpretation

ASSESSMENT AND MANAGEMENT

problem	assessment	investigations	Management-pharmacologic	Management-nonpharmacologic
1.				
2.				
3.				
4.				

Key

17. Assessment

1	2	3	4	5
Wrong diagnosis or no differential diagnosis	correct diagnosis no discussion re: diagnostic certainty or clinical reasoning			correct diagnosis discussion re diagnostic certainty or clinical reasoning

18. Investigations

1	2	3	4	5
Inappropriate, excessive incomplete investigations	most of tests appropriate			appropriate for problem rationale explained

19. Management-pharmacological

1	2	3	4	5
Incorrect or no management plan Unrecognized drug interactions		plan given, no specific guidelines re: dosing, adverse side effects		plan given guidelines re: dosing, adverse side effects. Goals of therapy discussed

20. Management-nonpharmacological

1	2	3	4	5
Incorrect or no information provided		alluded to but no specific guidelines		addresses support groups physical aids, rehab, vaccines, preventative measures where appropriate

COMMUNICATIONS/FOLLOW UP

21. Follow up plans

1	2	3	4	5
No clear follow-up mentioned Not clear who is responsible for ordering tests, dealing with results		incomplete follow up plans		clearly identified who does what and when

22. Communication

1	2	3	4	5
No mention of discussion with patient		some of discussion mentioned		discussion with patient including patient's understanding reaction to discussion

23. This letter answered all of the family doctor's questions posed in the referral letter

SD D U A SA

GLOBAL RATING

23. Global rating scale

Based on this letter, I believe that the author can manage this patient at the level of a:

1	2	3	4	5
clinical clerk	junior resident	senior resident	internist	internist with special expertise in the area

COMMENTS:

APPENDIX D

MODIFIED CLAT-REVISED APRIL 2, 2001

FORMAT

1.The assessment is clearly identified in the text.
SD D U A SA

2. The plan is clearly identified in the text
SD D U A SA

3. I can quickly scan this letter to obtain key information
SD D U A SA

4.The letter is succinct
SD D U A SA

5. The tone of the letter is polite
SD D U A SA

6.The paragraphs dealt with one idea (i.e. 1 aspect of a medical problem)
SD D U A SA

EDUCATIONAL VALUE OF THE LETTER

7. This letter would contribute the referring physician's continuing education through its discussion of the patient's problem
SD D U A SA

8.This letter provides an explanation for the recommendations
SD D U A SA

DATA SYNTHESIS

9. The chief complaint/reason for referral is clearly identified in the first paragraph of the letter

SD D U A SA

10. Description of the patient's history

1	2	3	4	5
Rambling, foggy Difficult to follow		Easy to follow but not concise		Concise, relevant

11. Completeness of the history of presenting complaint: inclusion of key negative and positive key items in the history

1	2	3	4	5
Missing the majority of key positive or negative items		Missing 3 or 4 key negative or positive items		All key items present

12. Pertinent ancillary history (PMHx, pertinent FHX, occupation, smoking history, alcohol use)

1	2	3	4	5
Missing 3 of the 5	Missing 2 of the 5	Missing 1	All items present	All items present PMHx in point form

13. Medications (score as 5 if the letter states the patient is on no medications)

1	2	3	4	5
Not listed	incomplete	Drug names listed but not the dosages		Complete with dosages

14. Allergies(score as 5 if the letter states the patient has no allergies)

1	2	3	4	5
Not listed		Listed but no mention of type of reaction		Listed with description of the type of reaction

15. Physical examination

1	2	3	4	5
Missing key items*		1 or 2 key items missing		Concise/relevant To the medical problems

*key items refer to pertinent negative and positive findings for the symptoms/diseases discussed in the history as well as a vitals (bp and heart rate)

16. Ancillary tests (if applicable)

1	2	3	4	5
No information provided		Results recorded but no interpretation		Results provided with an interpretation

Assessment and management

problem	assessment	investigations	Management-pharmacologic(19)	Management-pharmacology(20)	Management-nonpharmacological
1					
2					
3					
4					

Key

17.Assessment

1	2	3	4	5
not addressed	wrong diagnosis	correct diagnosis no discussion re diagnostic certainty or clinical reasoning		correct diagnosis discussion re diagnostic certainty or clinical reasoning

18. Investigations

1	2	3	4	5
not addressed	inappropriate tests or missing key tests	most of tests appropriate	Appropriate for problem, complete	appropriate for problem rationale explained

19.Management-pharmacological: dosages and duration of therapy

1	2	3	4	5
Not addressed	Incorrect drugs recommended	Incomplete Correct drugs but no dosing schedule, no mention of maximum dose		Correct dosing Complete with schedule and target dose

20. Management-pharmacological: guidelines for drug usage

1	2	3	4	5
No mention of possible side effects, goals of therapy or drug drug interactions		Incomplete guidelines		Complete guidelines with advice re: adverse drug effects, total amount of drug to be given, goals of therapy

21. Management-nonpharmacological

1	2	3	4	5

not addressed	incorrect information	alluded to but no specific guidelines		addresses support groups physical aids, rehab vaccines, preventative measures where appropriate

22. Follow up plans: ordering of tests

1	2	3	4	5
Not clear what test are required		Clear who is to order test but not clear who will follow up with the results		Clear who will be ordering tests and follow up of results

23. Follow up plans: patient follow up

1	2	3	4	5
No clear statement about patient follow up				Clearly stated if or when the consultant will follow up patient Clearly stated when referring MD should see patient

24. Communication with patient

1	2	3	4	5
No mention of what patient was told		some mention of the discussion with the patient		Documentation of discussion with patient including patient's understanding and reaction

25. This letter answered all of the family doctor's questions posed in the referral letter
SD D U A SA

GLOBAL RATING

26. Global rating scale
The quality of the letter is:

1	2	3	4	5
poor	borderline	satisfactory	Very good	Excellent

Comments: