MANUAL FOR
IMETS
TEST DEVELOPERS AND ITEM WRITERS

International Medical English Testing System

2nd edition 2021

VILMOS WARTA, TIMEA NÉMETH, ALEXANDRA CSONGOR, ANNAMÁRIA GYŐRFFY, IWONA MISZTAL, ANIKÓ HAMBUCH, ANITA HEGEDŰS, GABRIELLA HILD, RITA KRÁNICZ, ERIKA MEISZTER, GABOR-REBEK-NAGY

Table of contents

Preface

International Medical English Testing System (IMETS) is planned to be a set of computer based specialist English for Medical Purposes (EMP) tests to be constructed by a European Consortium of testing and language communication experts from five countries of the European Union (Hungary, Italy, Malta, Poland, Spain). The ultimate aim of the Consortium, operating under the auspices of the 'ERASMUS +' Project financed by the EU, is to standardise the requirements concerning EMP skills all over Europe at levels B1, B2 and C1 in harmony with the recommendations of the Common European Framework of Reference for Languages (CEFR). In addition, as a response to the preferences of our 'computer dependant society', the Consortium targets to develop an electronically available testing system.

IMETS intends to test the four basic EMP skills i.e. reading, writing, speaking and listening at levels B1, B2 and C1. Candidates may choose to take IMETS Written Test or IMETS Oral Test or both at each level.

This Manual for Test Developers and item writers is also meant to serve as a standardization document. It is intended to serve as a common reference framework in the process of test development as well as a set of written rules that will increase both the validity and reliability of this computer based EMP test.

This Manual is based on the basic concepts the authors presented in the Manual for sTANDEM test developers and item writers in 2012.

1 LANGUAGE TESTING CONCEPTS AND THEIR RELEVANCE TO IMETS

There are four basic language testing concepts which will be addressed in this chapter i.e. reliability, validity, backwash and practicality. All four are thought to be crucially important criteria that must be taken into account when designing a system of language tests. All four are interrelated and influence the process of language learning and teaching.

1.1 TEST RELIABILITY

Reliability is a major requirement of language tests which ensures that the same test takers taking the same test on different testing occasions achieve nearly the same score on each of these occasions. This definition may sound somewhat artificial, as it is highly unlikely that one and the same test taker takes the same test more than once. Moreover, if such a thing did happen, the learning factor ought to be taken into account, i.e. on the second and additional testing occasions, the same test takers the learning of the first test, so distorting the results of a second or additional test. Because of these practical interfering factors, reliability is defined in more general terms i.e. the consistency of the results on repeated administrations. The types of reliability are not going to be discussed further in this chapter, since this falls into the realms of the assessment phase of language testing, and will be included in a separate IMETS tests manual. However, the relationship between reliability and the other three basic testing concepts will be touched upon.

1.2 TEST VALIDITY

If we wish to give a very simple and practical definition of test validity, it could be as follows: "test validity is that requirement of language tests which ensures that the test measures that which is test developers intend to be measured. A first glance, this definition seems self-evident, however in practical terms it is not so simple. The reason for this is that there can be so many aspects of the testing focus which simply makes it impossible to give a guarantee as to exactly what the test measures. Another difficulty lies in the vagueness of the testers' intention. They would claim that they know what a particular measurement is supposed to focus on, but can this really be proved to be so? Possible targets of language tests, like speaking or listening skills, ability to take a case history in a particular language, or skills of writing an official letter, are concepts which cannot be defined with a hundred percent accuracy.

Analytical approaches trying to define language skills, tend to provide taxonomies concerning the constituent sub-skills, but are these lists complete? Integrative views claim that it is impossible to further divide these skills and. Is this correct? Is there any evidence that either view is true or false? Obviously not? What remains to be done is to try and substantially narrow down the focus thus making possible a more careful specification of the test target. One way of doing so is identifying certain aspects and types of test validity. In this chapter we are going to look into 7 types of test validity. Five of them, face, concurrent, content, construct, and predictive validity, will be discussed relying on the works of Davies (1968:9), Henning (1987:94) and Bachman (1990:248).

1.2.1   Face validity

Face validity means how the test takers view and value the test. The test-takers' ideas about the test seemingly plays no part in successful testing, it has no direct influence on the content or form of the test papers. It has been shown that there is, indeed, no direct influence. However, what the test takers think about the content and form of tests remains an important issue for test developers. Test takers' opinion on test content is crucial, since, to a certain extent, all kinds of testing requires the cooperation of the test taker. Trust and respect are pre-requisites for such cooperation. Test takers need to trust that test requirements are relevant and they also need to respect the professionalism of the test developers, in other words, they need to accept them as sufficiently knowledgeable, trained, educated and prepared for testing other peoples' professional skills. This is especially important in the case of LSP testing, where, more often than not, the testers are not specialists in the test takers' field of knowledge.

What should be focused on is test takers' expectations of what constitutes an ideal with respect both to their knowledge and skills. This ideal includes content, which relates face validity to content validity and this will be described in detail later in this chapter. It also concerns a construct, which in practical terms relates to test takers' concepts of structure and the practical application of language skills in general. Additionally, it relates to the applicability of their language skills, and the association of face validity with predictive validity. It is obvious that, face validity is a complex phenomenon and, as such, should by no means be excluded from the issues to be taken on board at all stages of test design, development and administration.

In the case of IMETS tests, face validity represents especially doctors', pharmacists' and allied health care providers' views on the use of a specialist language. These professionals are highly qualified in their own fields and most of them also acknowledge the importance of highly developed LMP skills. They tend to accept that the testing of their LMP skills is performed at a professional level. However, knowing that the testers are not specialized in their particular professional fields, they can be very critical of the content and form of LMP tests.

Content-wise they often require that the texts used for testing are professionally up-to-date and contain relevant information. This is a requirement that is easy to comply with, though testing experts argue that even 20-year-old texts could be used effectively use in LMP tests. It should always be emphasised by IMETS, that it is not professional knowledge but LMP skills that comprise the test target. It is also important to make clear that some degree of specialist knowledge and skills are necessary for relevant LMP performance and the required degree of specialist knowledge increases with the level of language skills assessment.

Many test takers believe that language testing should include a separate paper on grammar and lexis, including terminology. This view reflects an old view that individuals learn their second and other languages, by learning rules of grammar and a sufficient number of words and phrases followed by but the mechanical application of these rules to acquired words and phrases. It is exceedingly difficult for test-developers to convince test-takers that they accept the concept of modern testing methodology, which by definition excludes the use of old testing methods.

As can be seen, face validity is related to most or even all other types of test validity, which makes it an important factor at the stage of designing and developing tests. In progressing this, the philosophy is to arrive at a balance between the possibilities of using professionally up-to-date texts with the restrictions of having no separate grammar or vocabulary test papers.

A natural way of increasing face validity is to make test task as near to real-life situations as possible. IMETS test tasks such as history taking, conference presentations, official letter writing, and analysing graphs and other non-linguistic materials are almost 100% in accord with the communicative tasks undertaken in test takers' professional life. Reading and listening comprehension tasks, though they represent a somewhat lesser degree of criterion referenced testing, can be organized so that the items concern professionally relevant features of the text.

Another important area concerning face validity is the specificity of the LSP text. IMETS is a set of LSP tests for doctors, allied health care providers and pharmacist, which has been created to exclude problems of specificity. However today, the degree of specialization in such areas of medicine as surgery, which increasingly is practiced in specialised areas such as, hand surgery, neck surgery, chest surgery, brain surgery, traumatological surgery, orthopaedic surgery, etc., makes it difficult to decide which surgical sub-field should be selected by test developers as test texts. If test developers select a particular sub-fields for test use, professionals working in the other fields may well feel disadvantaged, as well as feeling that colleagues whose field the sample text comes, from are unjustly advantaged by the test. The same holds true for nursing and pharmacy.

It is acknowledged that this is a difficult issue to deal with. However, there are good practices, which can be implemented to exclude or reduce this kind of bias. A possible method is to include more than one text to measure each skill. One of them should definitely be on general topics within the subfield, and relate to the general background knowledge of any professional working in the field. Another text may also be on general topics, part of the general knowledge background of the profession, i.e. knowledge they are supposed to have before they specialize in a sub-field.

Taking face validity requirements into account when designing and developing test papers and giving thorough orientation and, if necessary, explanation to questions concerning test content and form will contribute greatly to successful LSP testing.

1.2.2 Content validity

Some testing experts consider content validity and face validity to be the same (Henning 1978:94). However, it seems useful to make a distinction between the two types of validity, as it allows for discussing different aspects of the same ideas. While face validity is concerned with whether the test looks as though it tests what it intends to test in an appropriate way, content validity is concerned with sampling.

Sampling is a crucial procedure in constructing a test. The problem lies in the fact that it is impossible to create exhaustive tests, i.e. such texts that contain all the features necessary to be tested. Therefore, some kind of compromise has to be reached. When part of the testing philosophy is to use authentic texts related to the professional development of test takers and

not created solely for the purposes of language testing. In order to obtain representative samples, which include most of the characteristic features of specialist texts, it is necessary to seek the advice of distinguished specialists', who use such texts as part of their routine activity. Another possible way of effective sampling is collecting a very large number of appropriate texts and taking a random selection on each testing occasion. The question of sampling in IMETS tests will be discussed in more detail later on in this Manual.

To some extent, the foci of the test items should also be related to content validity. In practical terms it means that no professionally evident aspects should be used for test items just because they easily measured language-wise. On the other hand, no linguistically evident phenomenon should serve as a required solution just because it has a particular role in a profession. The reason for these restrictions are not so much that the items may become too easy for test takers, but, by becoming too easy they may tend to measure general knowledge rather than measuring important aspects of LMP skills.

1.2.3   Concurrent validity

Concurrent validity is the type of validity obtained on correlating scores on one test, e.g. IMETS with those of another similar but established test such as the PROFEX test system, in Hungary developed in harmony with CEFR according to existing recommendations. In the case of languages for which no established LMP tests are available, it seems that the only possibility to ascertain concurrent validity would be to resort to well established tests assessing general language competence. A test can be said to have concurrent validity if it correlates highly with the criterion measure that is, if students return a similar score with IMETS and PROFEX. This holds true even when the second test is not exactly congruent with the test to be validated. In the case of comparing IMETS with PROFEX we need to realize that the former is monolingual while the latter is bilingual, the former does not have a mediation module while the latter does. In reality, it is highly unlikely that test takers who take the IMETS tests will also take PROFEX within a relatively short period of time, as this would only result in obtaining two similar test certificates without any achieved advantage. Concurrent validity is usually established as part of the validation process. Test validation in general is meant to prove that testers, test what they claim to test. Concurrent validity is usually established as a final stage of the validation process, and is termed external validation. External validation links the test to be validated to previously established measuring tools used for measuring similar proficiencies. Concurrent

validity relies on statistical methods which are claimed to increase its power, though however, statistics cannot solve all the associated problems. The major difficulty in establishing concurrent validity lies in the restricted similarity between the test to be validated and the criterion test. Differences in the two scores may originate from the differences in testing methods. For instance, reading comprehension in IMETS is tested by using more multiple choice items which gives several times more chances of giving the right answer than filling in a summary cloze in PROFEX.

1.2.4   Predictive validity

Like concurrent validity, predictive validity is based on correlating two tests. For example, the results of IMETS tests are correlated to the results obtained by high fliers in a group of LMP students. In both cases, some kind of criterion-referenced testing is taking place, and teachers of LMP courses may have experience with certain scores achieved by course participants correlated to real-life criteria, like, for instance, being admitted to a job where a certain level of LMP skills is required. Such comparison can be very useful in predicting performances, although, similarly to concurrent validity, some differences may originate from different testing methods.

In terms of test development, it is crucial to take predictive validity into account, since IMETS, unlike general EFL tests, is meant to certify language skills closely related to the test takers' profession. This is an essential goal which each item of IMETS should reflect and serve. No language phenomena should be measured solely for their own sake: the phenomenon selected for testing should be important functionally rather than structurally.

1.2.5   Criterion validity

This term denotes a set of specific information that indicates relationship between test score, and a criterion, which is thought to present the ability or skills being tested. The criteria may include "level of ability as defined by group membership, individuals' performance on another test of the ability in question or their relative success in performing some task that involves this ability" (Bachman 1990:248). It is easy to conclude that some of these criteria are not related to institutional testing but rather to practical efficiency. This boils down to the English proverb 'The proof of the pudding is in the eating', however, the abilities can manifest themselves at different levels. If efficacy in performing a communicative task in LMP is defined as simply

solving the task without regard to the quality then we agree that simply sending fluent verbal messages to clients is sufficient. However, we know that accuracy is also very important in LMP interactions, so the criterion should not be defined as simply sending verbal messages to clients but sending meaningful messages in accordance with the conventions of the profession and finding out about the effect achieved by those messages.

In certain task-types of IMETS (e.g. history taking) pragmatics is just as important as fluency and accuracy. Remaining with the example of simulated history taking, the way how the message is sent to the client (the examiner) and how his / her responses are taken into account by the professional (the test taker) should also be taken into account when designing such items. Another example is reading comprehension, where "reading between the lines", i.e. making inferences from implicated information is not just a matter of amusement but a professionally important activity. The degree of the author's commitment to the claims s/he makes is very rarely explicit. A clear understanding of it is absolutely necessary for making important professional decisions. Consequently, in a simulated reading comprehension task items focusing on this aspect of the text is inevitable. This also applies to listening comprehension items, where the presenter of a simulated lecture may give different relative weight to different facts.

An important and very much hoped 'side effect' of this Manual should be a relatively detailed description of the criterion, i.e. a summary of the knowledge, skills and abilities related to this specific purpose language use required by a professional when performing communicative tasks in English. If this summary is well-designed and contains the most important and typical foci of EMP language use, the item writers should use it as a springboard to which they can always return for 'ammunition'. In this way the items can be consciously designed to reduce the degree of 'arbitrariness' some language tests, especially EFL tests for general purposes suffer from.

1.2.6 Response validity

Response validity, which statistically is measured by the so-called discrimination index, is related to the fact that learners respond to a test in a predictable way. It means that high fliers in a learners' group are expected to give right answers to the test items in the majority of the cases while those placed towards the bottom of the class will score less. If this is not the case, especially when those in the high fliers' group give a lot of wrong responses, response validity

is undermined. This can be caused by several factors including the following. Test takers perceive a test to have low face validity and, as a consequence, they develop negative attitudes towards the test and these negative attitudes affect their responses. The instructions in the test may be ambiguous or unclear and the test takers give up trying to understand what they are required to do in the test. Test takers health or the environment in which the test is administered may be unsatisfactory, which may also affect the given responses. Henning (1987:96) suggests that it is possible to check this type of validity by interviewing learners who provide unexpected responses.

Ensuring this type of validity is mainly, but not entirely, the responsibility of the institution administering the test. As far as face validity is concerned, it is necessary to give substantial information and direction to future test takers. Sample tests should be made available for them to give them opportunities for making themselves familiar with the test format. The timing of each paper should be appropriate, allotting sufficient time to perform the tasks without haste. However, allocating too much time may have a negative effect on test takers.

The test papers should be attractive in their physical appearance and they should be easy to follow with clear and brief rather than verbose instructions. The scores achievable should also be added to the instructions, indicating what score is available for each subtask. If there is a minimum requirement for each paper, as in IMETS tests [where it is necessary to achieve 40% in each constituting paper, 60% being the pass grade for the whole test in both the oral and written sections], this fact should be included in the general instructions accompanying the papers. Sufficient time should be provided for reading the instructions and any possible misunderstanding related to them should be corrected before the test is administered. The time should be carefully and fairly administered, no advantages or disadvantages should be related to inappropriate time management on the part of the test administrators.

The physical environment of test administration should be faultless, stimulating and should exclude any disturbing factors coming from the test takers or the outside. The room where the test takes place must be well lit and well-aired and be large enough to accommodate the number of test takers on any particular occasion. The predetermined number of test takers should never be exceeded for any reason. In the case of written test papers a prescribed distance of one meter in each direction should be maintained between test takers. Test administrators should not permit any kind of communication between test takers or between test takers and outsiders. No

electronic devices other than provided by IMETS test invigilators should be used during test administration. The use of any types of aids, not listed in the test specifications, should be excluded. The invigilators should constantly be on guard and warnings should be given to anyone appearing to violate the rules. Test takers caught cheating in any way must be suspended. The procedure on how this should be done should be carefully detailed, and should be incorporated into the test rules and regulations. Test takers should be informed about the rules and possible consequences of violating them. When an apparent breach of regulations arises, appropriate action should be applied without hesitation.

1.2.7 Construct validity

Construct validity is present when a test or part of it or a testing method or technique can be proven to measure "just the ability which it is supposed to measure" (Hughes 1989: 26). This is especially true for LSP tests like IMETS, where the whole test can be demonstrated to measure the ability it is intended to measure. For instance, in the history-taking task there is simulated interaction between doctor and patient. The simulation is based on prompts given to both the test taker and the interlocutor, the former playing the doctor's part. Test takers are aware that they must perform this task in accordance with the rules of the profession, relying on the given prompts. The prompts are based on authentic real-life situations and are revised by specialists in the field in question, before test administration. Test takers are only required to produce verbal behaviour in accordance with the information at their disposal. The construct is history taking in English, which is almost exactly reflected by the task and the requirements. This task seems to be very close to direct testing, whereby the exam would take place in a doctor's consulting room, where the doctor's part would be played by the testee. Differences between the two situations include the location and the lack of responsibility for the case. Although these are important factors, this kind of simulation still has a very strong similarity with the corresponding real-life situation. In this case, and many others, the question arises, why not use a one hundred percent direct testing method. The reason is that the obstacles involved are virtually insurmountable, including patients' human rights and claim for privacy, difficulties of logistics, etc.

Another example in the IMETS tests is part of the oral test whereby testees are required to analyse some non-linguistic information such as a graph depicting mortality and morbidity rates among men and women suffering from cardio-vascular diseases. In this case, the simulation is

almost perfect: the only parameter missing is the real time scientific conference environment, but the task is the same, as if it was a part of a presentation at an actual conference.

A third example from IMETS tests is the letter writing task in the written exam. Official letter writing takes place as a simulated task. The topic of the letter is given in the instructions and guidance for the test takers is given in the form of prompts. This task can also be regarded as nearly perfect simulation, the only difference between the testing and real-life situation being that the former one is usually not sent to the addressee.

Yet another IMETS task which has very high construct validity is the interaction between two professionals, e.g. doctors discussing a case on handing over ward duties. This is an entirely relevant situation that takes place daily in hospitals. The difference between the real-life situation and the testing situation lies in the location and the lack / presence of professional responsibility. The prompts used for eliciting language use are reviewed by specialists working in the field.

In conclusion we can say that construct validity is an important parameter in the case of LSP tests and any effort to increase construct validity will increase the efficacy of such language exams by making it as near as possible to real-life situations using linguistic behaviour normally relevant in such situations.

A thorough description of the construct should be at IMETS test developers' disposal. This kind of description can be register-based, discourse-based or genre-based. In the case of IMETS genre-based descriptions are preferred, for this kind of description seems to embrace the other two and may provide grounds for producing tasks (items) having strong construct validity.

1.3   GENRES AND BACKWASH EFFECT

In simple terms backwash is the effect of the test on language learning and is very closely related to construct validity. As has been mentioned, the construct in the case of each task in IMETS will be genre-based.

Genre [Swales; 1981,1985,1990] can be defined as a highly structured and conventionalised communicative event characterised by communicative purposes and used by a professional

community. A more detailed description of the individual features of genre includes the following points:

1) A genre is a class of communicative events. This means that each paper included in IMETS can be subdivided on the basis of a genre.

2) A mass of communicative events is turned into a genre by a set of shared communicative purposes. The consortium, having undertaken to develop the IMETS testing system, agreed upon focusing on one particular type of the possible communicative purposes in EMP, namely, that of the professional rather than that of clients. In principle it would be possible to use popular scientific articles instead of professional research articles, or retailers' presentations used for sale purposes, instead of biomedical conference presentations, or a talk between two clients about the advantages and drawbacks of the National Health Service. However, these tasks would be irrelevant to professionals, since their exchange in all the three situations has the same communicative purpose i.e. to communicate and / or elicit information for professional medical purposes. In other words, a shared communicative purpose necessitates the use of specific genres rather than the use of some kind of discourse.

3) Instances of genres are prototypical to various degrees. It is clear that one presentation cannot be exactly like another one. Biomedical research articles can be remarkably different in length, form and content, among other things. However, for each genre there is a prototype which comprises many more elements than any particular sample of that genre could indicate. Therefore, the prototypical character of genres is just another argument for using the genre-based approach to EMP testing. IMETS is also meant to measure test takers' flexibility in producing a concrete sample of the biomedical genres included in it. A pre-requisite for this flexibility is firm knowledge of these genres, which seems to offer an extremely positive backwash effect on learning EMP.

4) Genre can be interpreted as a set of constraints concerning content, arrangement and form. If IMETS test takers are aware of these constraints, as well as the unwritten rules of arrangement and form, they should be considered able to cope with communicative tasks which they will face in professional settings. Consequently, some of the test items may focus on these aspects of the genre in question. However, this does not mean asking direct questions about the rules.

5) The way in which the discourse community regard their own genre(s) can be very revealing. In terms of backwash, this means that all test items should be pre-tested

on professionals and their opinions should be asked regarding the practicality and relevance of the items from a professional perspective. Doing so, will contribute to test developers' effort to increase test validity in general, and will ensure that, in addition to the content of the texts chosen, the test will convey the correct ways for test takers and their tutors to prepare for IMETS tests.

As can be seen, the genre-based approach is an essential feature of IMETS tests, which, in addition to providing favourable backwash, is also associated with the other three basic aspects of EMP testing: reliability, validity and practicality.

1.4 PRACTICALITY

A good test should also be practical in terms of financial limitations, time constraints, ease of administration, scoring and interpretation. *(http://www.tesol-taiwan.org/index.php?title =Languagetesting)*. There is constant discrepancy between practicality and the other three basic testing concepts. Practicality somehow seems to undermine the other three by imposing restrictions and limitations on them. It might be thought that restricting the time and financial means invested into designing, developing and administering the test, lessens thorough, valid and reliable measurement of language skills and, to some extent, hinders the desired backwash effect. These fears may be partially true justified, and test designers, developers and administrators may have to be pragmatic in order to economize on time, finances and human resources. However, these efforts serve both parties' interests.

The test takers' efforts to give account of their linguistic ability should be proportionate to the importance of having their language skills certified and they should also be cost effective in terms of the test fee. Ideally, testing language skills, especially LSP skills such as is intended by IMETS, would last for several days. If it were possible, all the major aspects of each skill could be thoroughly tested and even some of the simulation tasks would be more detailed and effective. However, for practical reasons this is ruled out as impossible. Very few, if any, test takers would be willing to devote so much time to having their language skills tested. Very few exams last for several days and if there are such exams, much more is at stake, e.g. entering a competitive and highly rewarding job, which is not quite true for most language tests.

The test developers and administrators do high quality professional work for which they should expect remuneration. The work performed by them include a wide range of activities, among which are such things as familiarizing themselves with the test specifications, acquiring knowledge on skills levels, writing the preliminary drafts of the test, trialling the test, i.e. pretesting it on a target population similar to the real test population, assessing scores of the pre-test, discarding the irrelevant items and replacing them by appropriate ones, finalizing the test paper and making sure that all the instructions are clear, as well as excluding spelling and grammatical mistakes.

1.5   MORAL AND ETHICAL CONSIDERATIONS IN IMETS

1.5.1   Moral considerations

IMETS is intended to be an international EMP exam developed by a multinational consortium of LSP teaching and testing experts to be used in as many countries of Europe and Asia as possible. This fact in itself should provide guarantee against any offence concerning the test takers' ethnic origin, religious beliefs, national identity, political views, sexual orientation or any other sensitivity, e.g. obscenity, pornography or inhumane practices like torturing or humiliating fellow humans or using biomedical knowledge for doing harm rather than good to human beings.

It is accepted as obligatory, that all IMETS texts, tasks, test items, instructions and activities related to test development, administration and assessment must be free of any bias that could be introduced by the particular views or circumstances of the test takers and test administrators.

Test developers, item writers and those in charge of editing IMETS tests, are responsible for double checking all products related to IMETS, including texts, test items, tasks, instructions and assessment criteria, and ensuring that they are free of any such bias in relation to the issues listed above.

1.5.2   Ethical considerations

When considering ethical considerations in relation to IMETS, the concept of fairness should be considered of paramount importance. This concept is both very easy and very difficult to

define. It is easy because, in general, it refers to a kind of behaviour "free from bias or injustice, even handedness*" (http://dictionary.reference.com).* And it is difficult, because its application needs to be very carefully fine-tuned to all aspects of life including all aspects of testing.

In terms of language testing, it means keeping strictly to testing principles, especially reliability, and so offer a kind of guarantee to test takers that the tests will present the same level of difficulty on each testing occasion.

In LSP testing, in addition to preserving reliability, fairness also includes the avoidance of favouring any groups of test takers, for instance, representatives of a particular subfield should not be advantaged by a particular test focusing on their specific area of work. Sampling in IMETS is a delicate issue. Test developers and item writers are required to sample from professional topics that are common to the different professional groups for which the examination is intended. It is considered totally unfair to select texts with a narrow specialist focus. An example of this could be the description of a recently published procedure in brain surgery, which could be relevant to the few brain surgeons taking the test, but would be especially difficult for a dermatologist, even if he/she is very experienced in surgical dermatology.

Another ethical issue in IMETS is related to the specific background knowledge required for completing any of the IMETS tests. First of all, it should be made clear that background knowledge is not a focus of assessed in IMETS tests, since they are meant to measure the level of development of LSP skills, not professional skills or knowledge. At the same time it should be accepted that professional background knowledge is a pre-requisite for successful completion of mostly simulated professional tasks. This may seem to contradict the principle of excluding professional knowledge from the test targets, but in actual fact it does not. Fairness in this respect means bearing in mind the following two important rules.

Firstly, no item should directly relate to any specific background knowledge (e.g. biomedical knowledge). All IMETS test developers and item writers, no matter whether they are biomedical professionals or LSP experts, should refrain from writing test items or producing tasks, directly concerning biomedical issues. IMETS test items and tasks should always relate to some aspects of EMP skills, such as comprehending, conveying or eliciting biomedical information. If followed to the extreme, this could result in specialist texts selected for IMETS

tests, containing biomedical information which, due to the development of biomedical sciences, is regarded by some members of the medical profession as controversial or outdated. An adverse effect of selecting such texts is that they may undermine face validity and it is therefore desirable to striving to selecting up-to-date specialist texts. Still, as a consequence of the fast developments in modern science that there can be no guarantee for all texts selected to represent the latest advances.

Secondly, IMETS test developers and item writers must be aware of the fact, that the amount of required biomedical background knowledge, which is an important pre-requisite of successful EMP testing increases with skills levels. In practical terms it means that items and tasks given at level B1 require less background knowledge that those designed for level B2 and the ones used at level C1 require biomedical background knowledge remarkably exceeding that at the two lower levels both in quality and quantity. In practical terms it means that fairness requires test developers and item writers to adjust items and tasks to meet the appropriate professional skills level. At the lower levels, test items requiring too much or too sophisticated background knowledge are thought to be as unfair as the inclusion of trivial background knowledge at level C1. An example of this is that understanding implied information can be unfair at level B1 or even B2, which in most of the cases is only possible if the test taker has firm background knowledge, while setting professionally trivial tasks at level C1 is equally unfair as it can undermine objective judgement of the skill being focused on.

Copyright is a crucial issue all over IMETS skills testing. At all three levels of competence, it is especially important in the case of selecting audio texts serving for stimuli in LC tests, when selecting texts for RC tests and also when graphs are used for writing papers. Only texts with previously arranged copyright can be used in IMETS tests. The copyright holder may donate or sell the copyright to IMETS. Whichever is the case, it should be properly documented by IMETS management. Before starting work on a text, test developers and item writers must check that copyright issues are properly in place.

2 PRACTICAL GUIDE FOR IMETS READING PAPER DEVELOPERS AND ITEM WRITERS

2.1 Background

The purpose of the reading comprehension (RC) paper is to find out how test takers, most probably practitioners or students of medicine, nursing and pharmacy, can use their skills in real-life professional encounters such as reading professional literature, medical records and drug dosage instructions.

It is common knowledge in language testing that RC can be tested at three different levels: literal comprehension, interpretive or referential comprehension and critical comprehension. The following three paragraphs represent a summary of RC at these three levels as it is given in Akmar (1999).

Comprehension at the literal level involves surface meanings. What is tested here is the comprehension of information and ideas that are explicitly stated in the text. In addition, it is also appropriate to test vocabulary, which is "influenced by one's mastery of word meanings in context" (Akmar, 1999).

At the interpretive or referential level, readers go beyond concrete ideas. "They need to be able to see relationships among ideas, for example how ideas go together and also see the implied meanings of these ideas. It is also obvious that before they can do this, they have to first understand the ideas that are stated (literal comprehension). Interpretive or referential comprehension includes thinking processes such as drawing conclusions, making generalizations and predicting outcomes. At this level, more challenging questions tasks can be given such as:

- Re-arrange the ideas or topics discussed in the text.
- Explain the author's purpose of writing the text.
- Summarize the main idea when this is not explicitly stated in the text.
- Select conclusions which can be deduced from the text they have read.

At the third level of comprehension is critical reading whereby ideas and information are evaluated. Critical evaluation occurs only after readers have understood the ideas and

information that the writer has presented. At this level, students can be tested on the following skills:

- The ability to differentiate between facts and opinions.
- The ability to recognize persuasive statements.
- The ability to judge the accuracy of the information given in the text.

2.2 RC at level B1

Task and time available

Test takers at level B1 should be able to understand global and factual information in two narrative medical texts on general topics within about 60 minutes.

Genres and source of texts

Since authenticity is a most important feature of IMETS tests, not only the communicative situations but the genres to which test takers are exposed should be authentic even at level B1. What makes them authentic is that they appear in the test takers' professional life in a natural way, as part of their professional daily routine activity. IMETS test specifications include the following target genres for testing RC skills at level B1:

1. textbook extracts
2. product descriptions
3. instructions (e.g. for treatments, instrument use, etc.)
4. abstracts
5. information leaflets
6. case reports
7. referral letters
8. internal communication

The test developers shall guarantee that the texts selected, considering, especially, authorship rights, are allowed to be used for the IMETS Reading Paper. The source of the texts shall be given.

Length

In line with the concept of representability, two texts are provided altogether in 590-610 words (290-310 words each). A note concerning the length of the texts should be made here. It is

almost impossible to find texts exactly fitting into the range of 290-310 words. In order to preserve the main feature accepted in IMETS specifications, there are some ways of adjusting texts to the limits in length.

1. It is possible to cut the end of the text off, provided it does not contain crucial information without which the genre would be changed.
2. Although figures in parentheses may provide important information to readers about what the authors meant by the concepts they use in the text, they do not aid understanding, consequently, their lack does not disturb understanding.

Types of tasks

First of all, it is important to clarify that RC tests are about what the term suggests, i.e. reading comprehension. By no means should it aim at language accuracy, i.e. grammar points or professional knowledge, i.e. biomedical knowledge. These factors are necessary conditions and not aims of RC testing. This rule makes it clear that grammatically false forms should NEVER appear in the test, not even in the form of distractors. Scientifically false information may appear in the text, but ONLY as distractors.

In harmony with the concept of face validity IMETS applies Reading Paper task sheet templates, which are to be applied at each level. They not only present general requirements but also the format (character type, character size, justification etc.) to be followed.

The first task aims to evaluate skimming and scanning reading skills by a *matching headings* exercise, with four correct answers and two distractors provided (4 items = 4 points).

The second task, following the text-flow, is intended to measure analytical reading skills applying *a multiple choice (MC)* exercise with one correct answer and two distractors (text 1: 8 items = 8 points; text 2: 9 items = 9 points) provided. When writing MC items, the following checklist can be of great help.

Checklist for multiple choice items:

1. Items follow the text-flow, i.e. the order of items are in harmony with the structure of the text, which means the test takers' do not need to jump forwards and backwards in the text.

2. Items should not be based on common knowledge of the world or professional knowledge. It can be checked by trying to solve the question without reading the text. If it is successful, the item should be discarded or modified.

3. There must always be 1 correct solution and 2 distractors. The correct solution must be clearly correct and the distractors should completely be false. In other words, the distractors should really distract the test taker's attention from the right solution, which means that they must be worth considering as the correct solution. Distractors that can be judged as false at first glance should be discarded or amended.

4. A common mistake of item writers is that they make the right solution conspicuous in some way. It should be neither too long nor too short as compared to the length of the distractors. If the item writer uses logical manipulations such as affirmative and negative sentences, the right solution should not be at one extreme point (say, negative) against all the distractors at the other extreme point (affirmative).

5. Items requiring too complicated logical manipulation or interpretation, say double negation, are forbidden. So is any kind of manipulation that may mislead or confuse test takers. The items may require making inferences even at level B1, but they may never require following a complicated, shrewd-minded way of thinking.

6. It is very important that all the items can be related to the text only, no knowledge of the world or professional knowledge based on recent research findings may be used for giving the right solution, however up-to-date it is. In extreme cases (which should be avoided if possible) the solution required may contradict the latest research findings. Even in such situations only reading comprehension can be what matters. Complaints concerning outdated professional information should be taken with thanks, however, such complaints cannot serve as the grounds for challenging test results. Even for professionals or researchers working in the field it is not easy to keep pace with the developments of all branches of science, especially of biomedical science, which may change overnight.

7. It is also important that the language level of the test item should not exceed that of the text or the test taker's required language knowledge. This rule applies to vocabulary as well as structure.

Language of instructions

Provided in an easy-to-understand target language, they are to be used in the way they are presented in the task sheet templets.

Reading Paper layout

This booklet is also meant to serve as a set of templates for the test papers, and, in order to provide for uniform layout their digital variety will be available for item writers.

Apart from the elements mentioned above it seems necessary and possible to define the more important features and guidelines with regard to test layout.

Fonts: Times New Roman, size: 12

Headers of the text should contain:

1) the logo including IMETS
2) the name of the skill (Reading Paper)
3) the skills level (B1, B2, C1)
4) the task number (Task 1, Task 2)
5) the scores available (…points)
6) the text (four parts of the text is indicated by numbers from 1-4 in brackets)
7) the length of the text (in words)
8) source of the text (in brackets, font size: 8, in italics)

Task Sheet:

1) the task sheet number (Task Sheet 1, Task Sheet 2)
2) instructions

Instructions should contain:

1) number of the exercise: *1)* or *2)*
2) a detailed description of the task, i.e. what the test taker is supposed to do
3) the text is in bold and italics
4) the scores available (…points)

Key should contain:

1) the logo including IMETS
2) the name of the skill (Reading Paper)
3) the skills level (B1, B2, C1)
4) the task number (**KEY - Task 1, KEY - Task 2**)
5) the total scores available (…points)
6) answer key to task 1
7) the scores available (…points)

8) answer key to task 2

9) the scores available (…points)

2.3 RC at level B2

Task and time available

Test takers at level B2 should be able to understand global, factual and implied information in two narrative or descriptive medical texts on general and not deeply specialised topics within about 60 minutes.

Testing understanding of implied information imposes some challenges for the test developers. The complication lies in the fact that this kind of information may be communicated in a number of ways and it is demanding or impossible to give a full classification of these ways. Another difficulty of recognizing and testing implied information is that in most of the cases we need to make some logic-based inferences to obtain it. The question always arises: how far this procedure should go, especially if the target is to create a test item checking understanding. The answer is very simple: far-reaching implications should never be a target in testing reading comprehension in a foreign or second language. Test developers and item writers must remember: what they should check is the skill of language comprehension rather than the ability of logical information processing. It is clear that one cannot exist without the other, however, in IMETS tests, priority must be given to the language comprehension aspect.

Another question is the ratio of items on factual information and applied information. Due to the implicit character of applied information it requires a special effort to decipher, while factual information is usually explicit. To be fair, in a test at level B2 the vast majority of test items should concern factual information and only a few should focus on implied information. The ratio of the latter type of information should not exceed 20 percent in IMETS RC tests.

Genres and source of texts

Since authenticity is a most important feature of IMETS tests, not only the communicative situations but the genres to which test takers are exposed should be authentic even at level B2. What makes them authentic is that they appear in the test takers' professional life in a natural way, as part of their professional daily routine activity. IMETS test specifications include the following target genres for testing RC skills at level B2:

24

1. textbook extracts
2. official letters on professional topic
3. research articles
4. case reports
5. referral letters
6. ethical approvals
7. motivation letters
8. abstracts
9. informed consent statements
10. letters of application
11. grant applications
12. internal communication

The test developers shall guarantee that the texts selected, considering, especially, authorship rights, are allowed to be used for the IMETS Reading Paper. The source of the texts shall be given.

Length

In line with the concept of representability, two texts are provided altogether in 890-910 words (440-460 words each). A note concerning the length of the texts should be made here. It is almost impossible to find texts exactly fitting into the range of 440-460 words. In order to preserve the main feature accepted in IMETS specifications, there are some ways of adjusting texts to the limits in length.

1. It is possible to cut the end of the text off, provided it does not contain crucial information without which the genre would be changed.
2. Although figures in parentheses may provide important information to readers about what the authors meant by the concepts they use in the text, they do not aid understanding, consequently, their lack does not disturb understanding.

Types of tasks

First of all, it is important to clarify that RC tests are about what the term suggests, i.e. reading comprehension. By no means should it aim at language accuracy, i.e. grammar points or professional knowledge, i.e. biomedical knowledge. These factors are necessary conditions and not aims of RC testing. This rule makes it clear that grammatically false forms should NEVER appear in the test, not even in the form of distractors. Scientifically false information may appear in the text, but ONLY as distractors.

In harmony with the concept of face validity IMETS applies Reading Paper task sheet templates, which are to be applied at each level. They not only present general requirements but also the format (character type, character size, justification etc.) to be followed.

The first task aims to evaluate skimming and scanning reading skills by a *matching headings* exercise, with four correct answers and two distractors provided (4 items = 4 points).

The second task, following the text-flow, is intended to measure analytical reading skills applying *a multiple choice (MC)* exercise with one correct answer and three distractors (text 1: 8 items = 8 points; text 2: 9 items = 9 points) provided. When writing MC items, the following checklist can be of great help.

Checklist for multiple choice items:
1. Items follow the text-flow, i.e. the order of items are in harmony with the structure of the text, which means the test takers' do not need to jump forwards and backwards in the text.
2. Items should not be based on common knowledge of the world or professional knowledge. It can be checked by trying to solve the question without reading the text. If it is successful, the item should be discarded or modified.
3. There must always be 1 correct solution and 2 distractors. The correct solution must be clearly correct and the distractors should completely be false. In other words, the distractors should really distract the test taker's attention from the right solution, which means that they must be worth considering as the correct solution. Distractors that can be judged as false at first glance should be discarded or amended.
4. A common mistake of item writers is that they make the right solution conspicuous in some way. It should be neither too long nor too short as compared to the length of the distractors. If the item writer uses logical manipulations such as affirmative and negative sentences, the right solution should not be at one extreme point (say, negative) against all the distractors at the other extreme point (affirmative).
5. Items requiring too complicated logical manipulation or interpretation, say double negation, are forbidden. So is any kind of manipulation that may mislead or confuse test takers. The items may require making inferences even at level B2, but they may never require following a complicated, shrewd-minded way of thinking.

6. It is very important that all the items can be related to the text only, no knowledge of the world or professional knowledge based on recent research findings may be used for giving the right solution, however up-to-date it is. In extreme cases (which should be avoided if possible) the solution required may contradict the latest research findings. Even in such situations only reading comprehension can be what matters. Complaints concerning outdated professional information should be taken with thanks, however, such complaints cannot serve as the grounds for challenging test results. Even for professionals or researchers working in the field it is not easy to keep pace with the developments of all branches of science, especially of biomedical science, which may change overnight.

7. It is also important that the language level of the test item should not exceed that of the text or the test taker's required language knowledge. This rule applies to vocabulary as well as structure.

Language of instructions

Provided in an easy-to-understand target language, they are to be used in the way they are presented in the task sheet templets.

Reading Paper layout

This booklet is also meant to serve as a set of templates for the test papers, and, in order to provide for uniform layout their digital variety will be available for item writers.

Apart from the elements mentioned above it seems necessary and possible to define the more important features and guidelines with regard to test layout.

Fonts: Times New Roman, size: 12

Headers of the text should contain:

1) the logo including IMETS
2) the name of the skill (Reading Paper)
3) the skills level (B1, B2, C1)
4) the task number (Task 1, Task 2)
5) the scores available (…points)
6) the text (four parts of the text is indicated by numbers from 1-4 in brackets)
7) the length of the text (in words)

8) source of the text (in brackets, font size: 8, in italics)

Task Sheet:

1) the task sheet number (Task Sheet 1, Task Sheet 2)

2) instructions

Instructions should contain:

1) number of the exercise: *1)* or *2)*

2) a detailed description of the task, i.e. what the test taker is supposed to do

3) the text is in bold and italics

4) the scores available (…points)

Key should contain:

1) the logo including IMETS

2) the name of the skill (Reading Paper)

3) the skills level (B1, B2, C1)

4) the task number (**KEY - Task 1, KEY - Task 2**)

5) the total scores available (…points)

6) answer key to task 1

7) the scores available (…points)

8) answer key to task 2

9) the scores available (…points)

2.4   RC at level C1

Task and time available

Test takers at level C1 should be able to understand global, factual and implied information in two argumentative medical texts on general and not deeply specialised topics within about 60 minutes.

The main focus of RC at level C1 in the IMETS system is argumentative style. The ideal candidate is proficient in extracting all types of information, including global, factual and implied information from EMP texts on not deeply specialized biomedical topics. As compared to levels B1 and B2, at this level the requirement is a very high independence of the reader, a skill that is also associated with the ability of 'reading between the lines', i.e. easily understanding implied information that in most of the cases would be impossible to locate in the text.

Testing understanding of implied information imposes some challenges for the test developers. The complication lies in the fact that this kind of information may be communicated in a number of ways and it is demanding or impossible to give a full classification of these ways. Another difficulty of recognizing and testing implied information is that in most of the cases we need to make some logic-based inferences to obtain it. The question always arises: how far this procedure should go, especially if the target is to create a test item checking understanding. The answer is very simple: far-reaching implications should never be a target in testing reading comprehension in a foreign or second language. Test developers and item writers must remember: what they should check is the skill of language comprehension rather than the ability of logical information processing. It is clear that one cannot exist without the other, however, in IMETS tests, priority must be given to the language comprehension aspect.

Another question is the ratio of items on factual information and applied information. Due to the implicit character of applied information it requires a special effort to decipher, while factual information is usually explicit. To be fair, in a test at level C1 the vast majority of test items should concern factual information and only a few should focus on implied information. The ratio of the latter type of information should not exceed 20 percent in IMETS RC tests.

Genres and source of texts

Since authenticity is a most important feature of IMETS tests, not only the communicative situations but the genres to which test takers are exposed should be authentic even at level C1. What makes them authentic is that they appear in the test takers' professional life in a natural way, as part of their professional daily routine activity. IMETS test specifications include the following target genres for testing RC skills at level C1:

1. textbook extracts
2. official letters on professional topic
3. research articles
4. case reports
5. monographs
6. referral letters
7. letters to the editor
8. ethical approvals
9. editorials
10. reviewers comments

11. motivation letters

12. abstracts

13. informed consent statements

14. letters of application

15. grant applications

The test developers shall guarantee that the texts selected, considering, especially, authorship rights, are allowed to be used for the IMETS Reading Paper. The source of the texts shall be given.

Length

In line with the concept of representability, two texts are provided altogether in 1290-1310 words (640-660 words each). A note concerning the length of the texts should be made here. It is almost impossible to find texts exactly fitting into the range of 440-460 words. In order to preserve the main feature accepted in IMETS specifications, there are some ways of adjusting texts to the limits in length.

3. It is possible to cut the end of the text off, provided it does not contain crucial information without which the genre would be changed.

4. Although figures in parentheses may provide important information to readers about what the authors meant by the concepts they use in the text, they do not aid understanding, consequently, their lack does not disturb understanding.

Types of tasks

First of all, it is important to clarify that RC tests are about what the term suggests, i.e. reading comprehension. By no means should it aim at language accuracy, i.e. grammar points or professional knowledge, i.e. biomedical knowledge. These factors are necessary conditions and not aims of RC testing. This rule makes it clear that grammatically false forms should NEVER appear in the test, not even in the form of distractors. Scientifically false information may appear in the text, but ONLY as distractors.

In harmony with the concept of face validity IMETS applies Reading Paper task sheet templates, which are to be applied at each level. They not only present general requirements but also the format (character type, character size, justification etc.) to be followed.

The first task aims to evaluate skimming and scanning reading skills by a *matching headings* exercise, with four correct answers and two distractors provided (4 items = 4 points).

The second task, following the text-flow, is intended to measure analytical reading skills applying *a multiple choice (MC)* exercise with one correct answer and three distractors (text 1: 8 items = 8 points; text 2: 9 items = 9 points) provided. When writing MC items, the following checklist can be of great help.

Checklist for multiple choice items:

1. Items follow the text-flow, i.e. the order of items are in harmony with the structure of the text, which means the test takers' do not need to jump forwards and backwards in the text.

2. Items should not be based on common knowledge of the world or professional knowledge. It can be checked by trying to solve the question without reading the text. If it is successful, the item should be discarded or modified.

3. There must always be 1 correct solution and 2 distractors. The correct solution must be clearly correct and the distractors should completely be false. In other words, the distractors should really distract the test taker's attention from the right solution, which means that they must be worth considering as the correct solution. Distractors that can be judged as false at first glance should be discarded or amended.

4. A common mistake of item writers is that they make the right solution conspicuous in some way. It should be neither too long nor too short as compared to the length of the distractors. If the item writer uses logical manipulations such as affirmative and negative sentences, the right solution should not be at one extreme point (say, negative) against all the distractors at the other extreme point (affirmative).

5. Items requiring too complicated logical manipulation or interpretation, say double negation, are forbidden. So is any kind of manipulation that may mislead or confuse test takers. The items may require making inferences, but they may never require following a complicated, shrewd-minded way of thinking.

6. It is very important that all the items can be related to the text only, no knowledge of the world or professional knowledge based on recent research findings may be used for giving the right solution, however up-to-date it is. In extreme cases (which should be avoided if possible) the solution required may contradict the latest research findings. Even in such situations only reading comprehension can be what matters. Complaints concerning outdated professional information should be taken with thanks, however, such complaints cannot serve as the grounds for challenging test results. Even for

31

professionals or researchers working in the field it is not easy to keep pace with the developments of all branches of science, especially of biomedical science, which may change overnight.

7. It is also important that the language level of the test item should not exceed that of the text or the test taker's required language knowledge. This rule applies to vocabulary as well as structure.

Language of instructions

Provided in an easy-to-understand target language, they are to be used in the way they are presented in the task sheet templets.

Reading Paper layout

This booklet is also meant to serve as a set of templates for the test papers, and, in order to provide for uniform layout their digital variety will be available for item writers.

Apart from the elements mentioned above it seems necessary and possible to define the more important features and guidelines with regard to test layout.

Fonts: Times New Roman, size: 12

Headers of the text should contain:

1) the logo including IMETS
2) the name of the skill (Reading Paper)
3) the skills level (B1, B2, C1)
4) the task number (Task 1, Task 2)
5) the scores available (…points)
6) the text (four parts of the text is indicated by numbers from 1-4 in brackets)
7) the length of the text (in words)
8) source of the text (in brackets, font size: 8, in italics)

Task Sheet:

1) the task sheet number (Task Sheet 1, Task Sheet 2)
2) instructions

Instructions should contain:

1) number of the exercise: *1)* or *2)*
2) a detailed description of the task, i.e. what the test taker is supposed to do

3) the text is in bold and italics

4) the scores available (…points)

<u>Key</u> should contain:

1) the logo including IMETS

2) the name of the skill (Reading Paper)

3) the skills level (B1, B2, C1)

4) the task number (**KEY - Task 1, KEY - Task 2**)

5) the total scores available (…points)

6) answer key to task 1

7) the scores available (…points)

8) answer key to task 2

9) the scores available (…points)

3. PRACTICAL GUIDE FOR IMETS LISTENING PAPER DEVELOPERS AND ITEM WRITERS

3.1 Background

The purpose of the listening comprehension (LC) paper is to find out how test takers, most probably practitioners or students of medicine, nursing or pharmacy, can use the skill in real-life professional encounters.

3.2 The construct of listening comprehension skills

Before getting down to the individual skills levels, it seems inevitable to clarify the construct of LC.

> *The ability to process extended samples of realistic spoken language, automatically and in real time, to understand the linguistic information that is unequivocally included in the text, and to make whatever inferences are unambiguously implicated by the content of the passage.*
>
> *http://www.lidgetgreen.org/index.php/raising-the-validity-bar/testing-listening-comprehension/the-listening-comprehension-construct/*

The above definition is far too general, however, it seems to be sufficient for explaining IMETS test developers' most important foci of interest.

3.3 Requirements for the texts to be used as stimuli in testing IMETS LC skills

The phrase '*extended samples of realistic spoken language*' basically means that authentic texts should be selected to serve as stimuli. The phrases '*realistic, spoken and extended samples*' all refer to this important feature. The phrase '*extended samples*' means that no isolated chunks of language, such as words, phrases or sentences, should be used as textual stimuli in IMETS LC tests, which is in accordance with the requirement of authenticity. The phrase '*processing* [this kind of texts] *automatically and in real time*' suggests the kind of texts to be used. As far as possible they should be natural and accurately reflect the meaning of the source text from which they have been taken. In IMETS LC tests, two occasions of listening are specified. This may be considered as a deviation from the normal process of listening, where it is never possible to have two occasions of listening. However, for testing purposes it is a practical and reasonable solution, since many of the conditions, which otherwise would accompany a listening task in real life, are missing. Such conditions, among others, include the speaker's associated gestures, impersonations and other non-verbal clues, all of which facilitate the listener's comprehension of the spoken words. Test developers should develop and use certain measures which can substitute for these non-verbal clues and by so doing make the tests more user-friendly and reduce the feel the tests being artificially constructed. Some such measures include inserting some pictures relevant to the content of the recording in the task sheet, briefly (in one sentence) summarizing the content of the recording both in the written and spoken instructions, explaining to the test taker the setting and the context of recording and what they are expected to do in this task.

As far as the topics of IMETS listening texts are concerned, they should bear features of the so-called' general specificity', i.e. all three groups of the potential target audience (physicians, nurses or pharmacists) should possess appropriate background knowledge necessary for understanding. As there are several overlaps in the professional background knowledge of these three professions, it is possible, although difficult to select such texts. The requirement of 'general specificity' is necessary for two reasons. First, it would undermine the principle of equal opportunity for all candidates. Extremely specific texts may be 'too specific' even within one of the three groups, e.g., the presentation of a complicated procedure of brain surgery would definitely advantage brain surgeons, though, at the same time, disadvantaging most of the candidates from other branches of medicine, not to mention the nurses or pharmacists.

Secondly, it would undermine comparability. If text were used which were specific and relative to the field of knowledge of each sub-group, on each test occasion special procedures would have to be applied to equalise the degree of difficulty between these texts. No reliable method exists for selecting texts, which are exactly of the same degree of difficulty. Therefore, it is very important to bear in mind that content-wise each LC text sample should suit all the three professional subgroups of the test takers, which is only possible if all the three subgroups are exposed to the same text.

The specificity of LC texts is ensured by selecting authentic text samples, which are health-care oriented and meant for health care professionals in the three subgroups of IMETS test takers. They may come from any of the three subject areas of medicine, pharmacy or nursing. On the one hand, they should be general enough to fall within the range of the expected background knowledge of each subgroup on the one hand and be specific enough to differ from a general scientific text. The prime target audience of IMETS LC texts should be health care providers including the three groups of professionals and students of the three afore mentioned disciplines. As health care professionals are also exposed to all various types of spoken communication arising from or intended for patients, non-professional caretakers and individuals, it is recommended to also include such discourse genres as well.

3.4   General requirements concerning LC in IMETS test specifications

Requirements concerning the text to be selected

The listening texts at all the three levels should be real-life samples of spoken language (authentic, semi-authentic, semi-scripted or unscripted). In other words, they must not be artificially contrived and no such written text should be used as a stimulus for any of the three levels of competence. The speed of speech is part of achieving a realistic assignment. The levels of speed are given in words/min (words per minute, wpm).

General requirements concerning the test tasks

Another aspect in the selection of authentic listening texts is that they should include and accurately reflect the range of tasks that the candidate would normally experience in trying to comprehend conversation in the course of his/her professional duties. This varies with skills levels and will be discussed in more detail there. The same applies to the genres to be used as textual stimuli. An important requirement concerning range of tasks within the listening test is

that they need to facilitate automatic language processing, i.e., they should not require any cognitive activity other than listening comprehension. In practical terms, it means that IMETS listening test items are not supposed to focus on the background knowledge of professional issues or language forms, since, according to the second part of the construct description above, LC skills testing is meant to find out if test takers "… *understand the linguistic information that is unequivocally included in the text, and to make whatever inferences are unambiguously implicated by the content of the passage."* The emphasis here is on "*unequivocally included in the text*", which concerns the nature of linguistic information to be tested and on "*unambiguously implicated"*, which concerns the implications that test takers can be required to understand. In LSP tests, it may be difficult to consistently achieve these two requirements, because many experts consider specific background knowledge as a confounding factor. However, this type of background knowledge should be regarded as a necessary condition of but by no means the target of the test items. As far as lexico-grammatical competence is concerned, it is also necessary but by no means should it be a target skill in LC test tasks.

In other words, if a candidate demonstrates an insufficient knowledge of background knowledge or lexico-grammatical competence, this should be taken as a clear indicator that his / her LC skills are insufficient or underdeveloped in these areas.

3.5   General principles of LC item writing in IMETS

Within the test there exists a relationship between the difficulty of the audio-stimulus and the related task. In practical terms it means that a difficult stimulus can (and should) be compensated by an easier task and vice versa, an easy text can (and should) be made more difficult by a more difficult task. What makes an audio text difficult includes the following:

1. the quality of the recording
2. the amount of background noise
3. text length
4. text speed
5. speaker's accent
6. degree of nominalization – verbalization
7. sentence complexity
8. the degree of structuredness
9. the amount of background knowledge the listener needs about the topic

10. the amount of background knowledge the listener needs about the genre

As can be seen from the above list, 80 percent of the factors influencing the difficulty of audio texts are of a physical character (1-8), which are essentially text related. The remainder relate to "knowledge and skills" (9-10), which are essentially candidate related.

The quality of recording
It should be of the same standard as a normal radio broadcast.
It should have an authentic context, realistic and natural and reflect real life situations at all times.

If the audio text is authentic, it will usually include a certain amount of <u>background noise</u>, which is a characteristic feature of authentic recordings. As you would expect the amount of background noise in these audio texts varies with skills levels. At level B1 there should be a lot less of it than at level C1. The background noise, however, is not supposed to block out the speaker's voice. It should achieve what its name implies, "noise in the background" which does not interrupt or makes understanding impossible.

Text length
At each level it is given in minutes. There are three audio texts at each level, each of them is limited in time. For each skill level the allocated total permitted time is given. Text length increases with the skill level. The role of short-term memory in understanding the texts and solving the tasks is indisputable. However, IMETS test developers and item writers should keep in mind that this is not a memory test but a specific purpose language test. In many cases, it is difficult to state how big a part memory plays in task solution, however, items focusing on figures, numbers, proper names, biodata of individuals or idiosyncratic phrases are considered to be inappropriate. Although such data can be classified as factual information, which is the focus of listening comprehension at each level, such items should be avoided, since apparently, in recalling them, memory plays a greater role than understanding.

Text speed
It is specified for each of the three skills levels is strongly related to text length. Test developers may use some compensatory mechanisms to avoid selecting texts, which are either too difficult or too easy. For instance, if the speed of a text at a certain level is close to the maximum, the

length can be closer to the minimum. This helps to reduce the pressure on test takers who would have the compounded difficulty of having to understand a text delivered at speed, which is, additionally, too long.   A similar compensatory mechanism can also work in reverse, i.e. if a slow text was chosen for a particular skills level within the range determined for that level, the difficulty of the LC task can be increased by increasing the text length.

The speaker's accent

It is another factor that influences understanding. The unwritten rule of fairness requires that test takers know in advance what kind of accents they can expect in the LC test at each level. This information, which is accurately determined in IMETS test specifications, will be made available to them in the brochure for test takers. As is to be expected, the number of accents required to be understood by candidates will increase according to the skill level. Test developers should pay attention not only to the accent but also to any peculiarity in the speaker's individual style of speech.   If an individual's style of speech causes problems with understanding the text then the text should not be used in IMETS testing.

The proportion of nominalization – verbalization

It is also thought to be an important factor that has an influence on text difficulty and information processing. A high degree of nominalization is known to increase text difficulty because of the high degree of compactness and high information density, while verbalized phrases are easier to grasp, since they break down information into more easily understandable portions. In practical terms, this means that nominalised texts should be avoided and wherever possible replaced by highly verbalized ones. An important consideration is the ratio of nominalised to verbalized texts used within any one skill level. It is permissible to have one highly normalised text, but not all three.

Sentence complexity

It is similar to and often is closely associated with nominalization – verbalization. There is general agreement that a text consisting only of simple sentences is easier to understand than one full of complex, compound sentences. The rule of thumb message for test developers is that this problem area should be treated similarly to other textual dichotomies like nominalisation – verbalisation.

By the degree of structuredness

We mean the discourse feature that is often, but not always, characterised by cohesion devices. As it is well known, all texts are coherent and a text can be coherent with or without cohesion markers. Texts without cohesion markers are more difficult to follow and understand than one which includes such markers. A text can also be well-structured with or without cohesion markers. Sometimes accent, intonation, word stress, pauses and other similar elements help to suggest the meaning within the text.

The amount of background knowledge about the genre required from the listener
It is also important, because it gives rise to expectations concerning the form of delivery and arrangement of the information. In point of fact, it can be said that a good knowledge of the genre, together with practice, are the most important pre-requisites for success in this IMETS test paper.

The amount of background knowledge about the topic required from the listener
It is a crucial factor in text difficulty. Not only does it mean that he / she knows the lexis or terminology used in discussing a subject but has a system of hierarchically-built schemata in which he / she stores this knowledge. Since this knowledge is acquired during professional socialization, it is far more complex and profound than the ability to recognize or produce the terms used to communicate the necessary meaning. In other words, the listener knows and understands the issues within the topic, and also knows the ways, in which these issues are related or discussed. This kind of knowledge gives rise to expectations, which are constantly compared to what the listener can hear in a particular situation. In LSP testing these expectations are even more important than in the so-called general language testing. Therefore IMETS test developers need to have it in mind when selecting texts, constructing instructions or writing items. The background knowledge can be divided into two main classes: knowledge of the world and background knowledge specific to the domain of language use in question. It is important to emphasize that neither of these two types of knowledge can serve as targets for IMETS tests at any of the three levels of competence. While it is well known that both of them are indispensable for language testing, neither of them are relevant language test targets. On the contrary, it is absolutely necessary to check each test item in IMETS LC tests to prove that it cannot be solved purely by relying on general knowledge or subject-specific background knowledge. Sometimes it is difficult to make a clear distinction between whether an item was solved by relying on understanding or on knowledge. A very simple way of doing so is trying to solve the item without listening to the stimulus text. If this strategy is successful, the item in

question should be discarded or modified. In extremely practical terms, we can say that issues of the sciences of medicine, nursing or pharmacy should never become the target of IMETS LC test items. The purpose is testing LC skills and not the underlying scientific knowledge.

This raises the issue of whether or not it is acceptable to use texts, which are not up-to date with respect to their time of publication. The answer is that this will depend, in the main, on our priorities. If it is accepted that LC tests are aimed at investigating LC skills and not scientific knowledge, then it is acceptable to conclude that there is no time limit involvement in the choice of texts.

This may leave the way open for candidates to criticise the test as containing outdated texts. However, it should always be remembered that the test is not about the knowledge a candidate has of a subject, but about his/her skills in understanding the spoken word.

It is unreasonable to expect test developers and item writers to keep abreast of all the advances and developments within the three professional sub-disciplines. Therefore, while it seems advisable that, whenever possible, the timing of publication of the original texts should be contemporaneous with the test occasion, knowing the current rapidity of progress in science, no test developer or text composer, can be expected to produce new texts and items in accordance with the latest advances science.

Any complaint by a candidate, concerning the scientific validity of a text should be politely but firmly rejected. The most desirable state of affairs, of course, is avoiding such complaints, which is possible by trying to select up-to-date texts, which will give rise to appropriate and acceptable test items.

3.6 LC at level B1

At level B1, LC skills testing focuses on understanding the main points of factual information in clearly structured and articulated medical audio-texts in standard British English or American English accents.

The texts are monologues and dialogues, each with a speed of 130-140 words/minute.

Genres and source of texts

1. interviews
2. health reports
3. lectures
4. patient's narratives
5. case reports

The test developers shall guarantee that the texts selected, considering, especially, authorship rights, are allowed to be used for the IMETS Listening Paper. The source of the texts shall be given.

Length

The text one and two range from 2.5 to 3.5 minutes in length while the remaining last approx. 1 min.

Types of tasks

In harmony with the concept of face validity IMETS applies Listening Paper task sheet templates, which are to be applied at each level (see later). They not only present general requirements but also the format (character type, character size, justification etc.) to be followed.

Task 1, level B1

Task 1 comprises of a mini-lecture on a medical topic. The task is intended to measure analytical listening skills applying *a multiple choice (MC)* exercise with one correct answer and two distractors provided (5 items = 5 points).

Task 2, level B1

Task 2 is an 8-option multiple-matching task. Test takers hear five short monologues and need to match an option to the correct speaker. Each option may be chosen only once. For example test takers are required to select the five correct health professionals from eight options or select the five correct topics of the monologue from eight options provided (5 items = 5 points). In this part, test takers are being tested on their global listening skills such as identifying main ideas and identifying contexts.

Task 3, level B1

Task 3 is a gap filling exercise based on a dialogue involving two or more speakers. The text typically takes the form of a doctor taking a patient's history or a conversation between two healthcare professionals focused on reporting patient's history. Test takers are presented with a the patient's notes written by a health care professional based on the interview with a patient or a colleague. There are five mistakes in the patient's notes that test takers need to identify and correct (5 items = 5 points).

Task 4, level B1

Task 4 is a gap filling exercise based on a dialogue. Test takers hear a conversation related to the case presented in Task 3 and are required to complete the five gaps in the treatment plan. In this part, test takers are being tested on their note taking skills and need to complete specific factual information (5 items = 5 points).

Task 5, level B1

The task is an 8-option multiple-matching task. Test takers hear five short monologues and need to match a statement to the correct speaker (5 items = 5 points). Each option may be chosen only once. In this part, test takers are being tested on their ability to listen for specific words or phrases focusing on detail.

Generally, gap-filling items in the IMETS LC Paper require exactly a one-word-solution for one item.

Checklist for multiple choice items:

1. Items follow the text-flow, i.e. the order of items are in harmony with the structure of the text, which means the test takers' do not need to jump forwards and backwards in the text.

2. Items should not be based on common knowledge of the world or professional knowledge. It can be checked by trying to solve the question without reading the text. If it is successful, the item should be discarded or modified.

3. There must always be 1 correct solution and 2 distractors. The correct solution must be clearly correct and the distractors should completely be false. In other words, the distractors should really distract the test taker's attention from the right solution, which means that they must be worth considering as the correct solution. Distractors that can be judged as false at first glance should be discarded or amended.

4. A common mistake of item writers is that they make the right solution conspicuous in some way. It should be neither too long nor too short as compared to the length of the distractors. If the item writer uses logical manipulations such as affirmative and negative sentences, the right solution should not be at one extreme point (say, negative) against all the distractors at the other extreme point (affirmative).

5. Items requiring too complicated logical manipulation or interpretation, say double negation, are forbidden. So is any kind of manipulation that may mislead or confuse test takers. The items may require making inferences even at level B1, but they may never require following a complicated, shrewd-minded way of thinking.

6. It is very important that all the items can be related to the text only, no knowledge of the world or professional knowledge based on recent research findings may be used for giving the right solution, however up-to-date it is. In extreme cases (which should be avoided if possible) the solution required may contradict the latest research findings. Even in such situations only reading comprehension can be what matters. Complaints concerning outdated professional information should be taken with thanks, however, such complaints cannot serve as the grounds for challenging test results. Even for professionals or researchers working in the field it is not easy to keep pace with the developments of all branches of science, especially of biomedical science, which may change overnight.

7. It is also important that the language level of the test item should not exceed that of the text or the test taker's required language knowledge. This rule applies to vocabulary as well as structure.

Checklist for gap-filling items

Similar to the multiple matching items, the item writer is supposed to find the text sites which (s)he wishes to use as the bases for writing gap-filling items.

The text extracts to be selected need some alteration or total reformulation to avoid repeating the wording found in the text, which would be an extremely easy task for the test takers to complete.

1. The language level of the re-written text extracts should not exceed the language level of the text or that required from test takers at a given level.

2. It should be avoided to require structure words belonging to the valency of a meaningful word, e.g. *.... insisted ... taking responsibility for ...* If the test item requires the test taker to supply the word 'ON' in the gap between the words 'insisted' and 'taking', we are not assessing the skill of LC but only test takers' grammar knowledge, which should not be the objective of an LC test. Most probably this task could be solved even without listening to the text it is contained in. Therefore, item writers should focus on medical terms exclusively.

3. The opposite of the situation described under point 2 is possible, however, with caution. E.g., *... insisted on taking ... for ...* If the test item requires the test taker to supply the word 'RESPONSIBILITY' in the gap between the words 'taking' and 'for', we may measure LC, since there are more than one word that fits into this gap. It can especially be appropriate in the case of paraphrased extracts.

4. Any kind of leading questions should be avoided. Most of the possible leading questions in gap filling originate from using the same stem. When paraphrasing the selected extracts, it is advisable to use synonyms or synonymous phrases with different stems.

5. The correct solution should be considered acceptable in the sense that it contains what was said on the audio recording and should not merely test the knowledge of grammar or lexis.

6. Test takers should not be able to find the right answer without listening to the audio recording, which means that gaps which can be filled on the basis of general knowledge or professional background knowledge are not acceptable. This can be checked by trying to solve the task sheet without listening to the audio recording.

Language of instructions

Provided in an easy-to-understand target language, they are to be used in the way they are presented in the task sheet templets.

LC test layout at level B1

This booklet is also meant to serve as a set of templates for the test papers, and, in order to provide for uniform layout their digital variety will be available for item writers.

Apart from the elements mentioned above it seems necessary and possible to define the more important features and guidelines with regard to test layout.

Fonts: Arial Narrow, size: 12

Headers should contain:

9) the logo including IMETS
10) the name of the skill (Listening Paper)
11) the skills level (B1, B2, C1)
12) the task sheet number (Task Sheet 1, Task Sheet 2)
13) the task number (Task 1, Task 2)
14) the scores available (…points)

Instructions should contain:

5) a detailed description of the task, i.e. what the test taker is supposed to do
6) the text is in bold and italics

Key should contain:

10) audio script
11) source of recording (in brackets, font size: 8, in italics)
12) length of recording (in words)
13) speed (in words/minute, wpm)
14) answer key

3.7   LC at level B2

At level B2, LC skills testing focuses on understanding global and detailed factual and implied information in   naturally structured and articulated medical audio-texts in British English (including light regional accents), non-native English (C1 or above, incl. light regional accents) or AmE (incl. light regional accents).

The texts are monologues and dialogues, each with a speed of 150-160 words/minute.

Genres and source of texts
1. interviews
2. medical news reports
3. case reports
4. lectures
5. patient's narratives

The test developers shall guarantee that the texts selected, considering, especially, authorship rights, are allowed to be used for the IMETS Listening Paper. The source of the texts shall be given.

Length
The text one and two range from 2.5 to 3.5 minutes in length while the remaining texts range from 1 to 2 minutes in length.

Types of tasks
In harmony with the concept of face validity IMETS applies Listening Paper task sheet templates, which are to be applied at each level (see later). They not only present general requirements but also the format (character type, character size, justification etc.) to be followed.

Task 1, level B2
The first task comprises of a mini-lecture on a medical topic. The task is intended to measure analytical listening skills applying *a multiple choice (MC)* exercise with one correct answer and two distractors provided (5 items = 10 points).

Task 2, level B2

The second task is an 8-option multiple-matching task. Test takers hear five short monologues and need to match an option to the correct speaker. Each option may be chosen only once. For example test takers are required to select the five correct health professionals from eight options or select the five correct topics of the monologue from eight options provided (5 items = 5 points). In this part, test takers are being tested on their global listening skills such as identifying main ideas and identifying contexts.

Task 3, level B2

Task 3 is a multiple choice exercise based on a dialogue involving two or more speakers. The text typically takes the form of a conversation between a healthcare professional and a patient. Test takers are required to select one correct answer from three options provided (3 items = 3 points).

Task 4, level B2

Task 4 is a gap filling exercise based on a dialogue. Test takers hear a conversation related to the case presented in Task 3 and are required to complete the four gaps in the treatment plan. In this part, test takers are being tested on their note taking skills and need to complete specific factual information (4 items = 4 points).

Task 5, level B2

Task 5 is a multiple choice exercise based on a dialogue involving two or more speakers. The text typically takes the form of a conversation between a healthcare professional and a patient. Test takers are required to select one correct answer from three options provided (3 items = 3 points).

Checklist for multiple choice items:

1. Items follow the text-flow, i.e. the order of items are in harmony with the structure of the text, which means the test takers' do not need to jump forwards and backwards in the text.
2. Items should not be based on common knowledge of the world or professional knowledge. It can be checked by trying to solve the question without reading the text. If it is successful, the item should be discarded or modified.

3. There must always be 1 correct solution and 2 distractors. The correct solution must be clearly correct and the distractors should completely be false. In other words, the distractors should really distract the test taker's attention from the right solution, which means that they must be worth considering as the correct solution. Distractors that can be judged as false at first glance should be discarded or amended.

4. A common mistake of item writers is that they make the right solution conspicuous in some way. It should be neither too long nor too short as compared to the length of the distractors. If the item writer uses logical manipulations such as affirmative and negative sentences, the right solution should not be at one extreme point (say, negative) against all the distractors at the other extreme point (affirmative).

5. Items requiring too complicated logical manipulation or interpretation, say double negation, are forbidden. So is any kind of manipulation that may mislead or confuse test takers. The items may require making inferences even at level B1, but they may never require following a complicated, shrewd-minded way of thinking.

6. It is very important that all the items can be related to the text only, no knowledge of the world or professional knowledge based on recent research findings may be used for giving the right solution, however up-to-date it is. In extreme cases (which should be avoided if possible) the solution required may contradict the latest research findings. Even in such situations only reading comprehension can be what matters. Complaints concerning outdated professional information should be taken with thanks, however, such complaints cannot serve as the grounds for challenging test results. Even for professionals or researchers working in the field it is not easy to keep pace with the developments of all branches of science, especially of biomedical science, which may change overnight.

7. It is also important that the language level of the test item should not exceed that of the text or the test taker's required language knowledge. This rule applies to vocabulary as well as structure.

Generally, gap-filling items in the IMETS LC Paper require exactly a one-word-solution for one item.

Checklist for gap-filling items

Similar to the multiple matching items, the item writer is supposed to find the text sites which (s)he wishes to use as the bases for writing gap-filling items.

The text extracts to be selected need some alteration or total reformulation to avoid repeating the wording found in the text, which would be an extremely easy task for the test takers to complete.

1. The language level of the re-written text extracts should not exceed the language level of the text or that required from test takers at a given level.

2. It should be avoided to require structure words belonging to the valency of a meaningful word, e.g. *.... insisted ... taking responsibility for ...* If the test item requires the test taker to supply the word 'ON' in the gap between the words 'insisted' and 'taking', we are not assessing the skill of LC but only test takers' grammar knowledge, which should not be the objective of an LC test. Most probably this task could be solved even without listening to the text it is contained in. Therefore, item writers should focus on medical terms exclusively.

3. The opposite of the situation described under point 2 is possible, however, with caution. E.g., *... insisted on taking ... for ...* If the test item requires the test taker to supply the word 'RESPONSIBILITY' in the gap between the words 'taking' and 'for', we may measure LC, since there are more than one word that fits into this gap. It can especially be appropriate in the case of paraphrased extracts.

4. Any kind of leading questions should be avoided. Most of the possible leading questions in gap filling originate from using the same stem. When paraphrasing the selected extracts, it is advisable to use synonyms or synonymous phrases with different stems.

5. The correct solution should be considered acceptable in the sense that it contains what was said on the audio recording and should not merely test the knowledge of grammar or lexis.

6. Test takers should not be able to find the right answer without listening to the audio recording, which means that gaps which can be filled on the basis of general knowledge or professional background knowledge are not acceptable. This can be checked by trying to solve the task sheet without listening to the audio recording.

Language of instructions

Provided in an easy-to-understand target language, they are to be used in the way they are presented in the task sheet templets.

LC test layout

This booklet is also meant to serve as a set of templates for the test papers, and, in order to provide for uniform layout their digital variety will be available for item writers.

Apart from the elements mentioned above it seems necessary and possible to define the more important features and guidelines with regard to test layout.

Fonts: Times Arial Narrow, size: 12

Headers should contain:

1) the logo including IMETS
2) the name of the skill (Listening Paper)
3) the skills level (B1, B2, C1)
4) the task sheet number (Task Sheet 1, Task Sheet 2)
5) the task number (Task 1, Task 2)
6) the scores available (…points)

Instructions should contain:

7) a detailed description of the task, i.e. what the test taker is supposed to do
8) the text is in bold and italics

Key should contain:

9) audio script
10) source of recording (in brackets, font size: 8, in italics)
11) length of recording (in words)
12) speed (in words/minute, wpm)
13) answer key

3.8 LC at level C1

At level B2, LC skills testing focuses on understanding complex medical audiotexts (expository or argumentative) with complex structures and authentic articulation in BritE (incl. Irish and Scottish accents), AmE (incl. Canadian and regional American accents), standard Australian English, or non-native English (C1 or above, incl. light regional accents)

The first text is always a dialogue and the second one is a monologue text, each with a speed of 170-180 words/minute.

Genres and source of texts

1. interviews
2. lectures
3. research reports
4. conference presentations
5. case reports

The test developers shall guarantee that the texts selected, considering, especially, authorship rights, are allowed to be used for the IMETS Listening Paper. The source of the texts shall be given.

Length

The text one ranges from 3.5 to 4.5 minutes in length while the remaining texts range from 1.5 to 2.5 minutes in length.

Types of tasks

In harmony with the concept of face validity IMETS applies Listening Paper task sheet templates, which are to be applied at each level (see later). They not only present general requirements but also the format (character type, character size, justification etc.) to be followed.

Task 1, level C1

The first task comprises of a mini-lecture on a medical topic. The task is intended to measure analytical listening skills applying *a multiple choice (MC)* exercise with one correct answer and two distractors provided (5 items = 10 points).

Task 2, level C1

The second task is an 8-option multiple-matching task. Test takers hear five short monologues and need to match an option to the correct speaker. Each option may be chosen only once. For example test takers are required to select the five correct health professionals from eight options or select the five correct topics of the monologue from eight options provided (5 items = 5 points). In this part, test takers are being tested on their global listening skills such as identifying main ideas and identifying contexts.

Task 3, level C1

Task 3 is a multiple choice exercise based on a dialogue involving two or more speakers. The text typically takes the form of a conversation between a healthcare professional and a patient. Test takers are required to select one correct answer from three options provided (3 items = 3 points).

Task 4, level C1

Task 4 is a gap filling exercise based on a monologue. Test takers hear a conversation related to the case presented in Task 3 and are required to complete the four gaps in the treatment plan. In this part, test takers are being tested on their note taking skills and need to complete specific factual information (4 items = 4 points).

Task 5, level C1

Task 5 is a multiple choice exercise based on a dialogue involving two or more speakers. The text typically takes the form of a conversation between a healthcare professional and a patient. Test takers are required to select one correct answer from three options provided (3 items = 3 points).

Checklist for multiple choice items:

1. Items follow the text-flow, i.e. the order of items are in harmony with the structure of the text, which means the test takers' do not need to jump forwards and backwards in the text.

2. Items should not be based on common knowledge of the world or professional knowledge. It can be checked by trying to solve the question without reading the text. If it is successful, the item should be discarded or modified.

3. There must always be 1 correct solution and 2 distractors. The correct solution must be clearly correct and the distractors should completely be false. In other words, the distractors should really distract the test taker's attention from the right solution, which means that they must be worth considering as the correct solution. Distractors that can be judged as false at first glance should be discarded or amended.

4. A common mistake of item writers is that they make the right solution conspicuous in some way. It should be neither too long nor too short as compared to the length of the distractors. If the item writer uses logical manipulations such as affirmative and negative sentences, the right solution should not be at one extreme point (say, negative) against all the distractors at the other extreme point (affirmative).

5. Items requiring too complicated logical manipulation or interpretation, say double negation, are forbidden. So is any kind of manipulation that may mislead or confuse test takers. The items may require making inferences even at level B1, but they may never require following a complicated, shrewd-minded way of thinking.

6. It is very important that all the items can be related to the text only, no knowledge of the world or professional knowledge based on recent research findings may be used for giving the right solution, however up-to-date it is. In extreme cases (which should be avoided if possible) the solution required may contradict the latest research findings. Even in such situations only reading comprehension can be what matters. Complaints concerning outdated professional information should be taken with thanks, however, such complaints cannot serve as the grounds for challenging test results. Even for professionals or researchers working in the field it is not easy to keep pace with the developments of all branches of science, especially of biomedical science, which may change overnight.

7. It is also important that the language level of the test item should not exceed that of the text or the test taker's required language knowledge. This rule applies to vocabulary as well as structure.

Generally, gap-filling items in the IMETS LC Paper require exactly a one-word-solution for one item.

Checklist for gap-filling items

Similar to the multiple matching items, the item writer is supposed to find the text sites which (s)he wishes to use as the bases for writing gap-filling items.

The text extracts to be selected need some alteration or total reformulation to avoid repeating the wording found in the text, which would be an extremely easy task for the test takers to complete.

1. The language level of the re-written text extracts should not exceed the language level of the text or that required from test takers at a given level.

2. It should be avoided to require structure words belonging to the valency of a meaningful word, e.g. *.... insisted ... taking responsibility for ...* If the test item requires the test taker to supply the word 'ON' in the gap between the words 'insisted' and 'taking', we are not assessing the skill of LC but only test takers' grammar knowledge, which should not be the objective of an LC test. Most probably this task could be solved even without listening to the text it is contained in. Therefore, item writers should focus on medical terms exclusively.

3. The opposite of the situation described under point 2 is possible, however, with caution. E.g., *... insisted on taking ... for ...* If the test item requires the test taker to supply the word 'RESPONSIBILITY' in the gap between the words 'taking' and 'for', we may measure LC, since there are more than one word that fits into this gap. It can especially be appropriate in the case of paraphrased extracts.

4. Any kind of leading questions should be avoided. Most of the possible leading questions in gap filling originate from using the same stem. When paraphrasing the selected extracts, it is advisable to use synonyms or synonymous phrases with different stems.

5. The correct solution should be considered acceptable in the sense that it contains what was said on the audio recording and should not merely test the knowledge of grammar or lexis.

6. Test takers should not be able to find the right answer without listening to the audio recording, which means that gaps which can be filled on the basis of general knowledge or professional background knowledge are not acceptable. This can be checked by trying to solve the task sheet without listening to the audio recording.

Language of instructions

Provided in an easy-to-understand target language, they are to be used in the way they are presented in the task sheet templets.

LC test layout

This booklet is also meant to serve as a set of templates for the test papers, and, in order to provide for uniform layout their digital variety will be available for item writers.

Apart from the elements mentioned above it seems necessary and possible to define the more important features and guidelines with regard to test layout.

Fonts: Arial Narrow, size: 12

Headers should contain:

1) the logo including IMETS
2) the name of the skill (Listening Paper)
3) the skills level (B1, B2, C1)
4) the task sheet number (Task Sheet 1, Task Sheet 2)
5) the task number (Task 1, Task 2)
6) the scores available (…points)

Instructions should contain:

7) a detailed description of the task, i.e. what the test taker is supposed to do
8) the text is in bold and italics

Key should contain:

9) audio script
10) source of recording (in brackets, font size: 8, in italics)
11) length of recording (in words)
12) speed (in words/minute, wpm)
13) answer key

4. SPEAKING

4.1 Background

Speaking is a complex activity, which can be viewed from a number of different aspects. The three main ways of viewing it are summarized in the following three definitions:

> *"... the action of conveying information or expressing one's thoughts and feelings in spoken languages. "*Oxford Dictionary of Current English (2009: 414)

> *"... speaking is the process of building and sharing meaning through the use of verbal or non-verbal symbols in a variety of contexts"* (Chaney 1998:13)

> *"speaking is the productive oral skill. It consists of producing systematic verbal utterances to convey meaning."* (Nunan 2003:48)

It is obvious from that speaking can be interpreted as an action, a process and a skill. While in language testing the third definition seems to be the most suitable, we should also keep the other two in mind when considering the nature of speaking and the issues of using this skill in real life.

4.2 The construct of speaking skills

Speaking, i.e. speech production is probably the most difficult language skill to test. This is partially due to the fact that speech production requires several sub-skills which are not always directly correlated with the success of this activity. For instance, distorted pronunciation does not always minimise understanding, while a very slight difference in intonation and rhythm may do so. Superficial grammatical mistakes may not lead to communication breakdown while deficiencies in vocabulary may cause serious misunderstanding. If we step a bit higher in the language hierarchy, it can be argued that speech lacking some degree of grammatical correctness may still convey the proper meaning, that violating the rule of appropriacy may cause confusion or even lead to a breakdown in communication.

Another factor, which further complicates speaking, is that, in the majority of the cases, it is related to other skills such as listening or reading. This is especially true of language tests,

where the stimulus for a test item is usually provided by visual, audio or audio-visual aids. Comprehension of the stimulus may heavily influence the performance.

With very few exceptions, speaking is interactive, i.e. there is always an audience to whom information is conveyed. The audience constantly communicates their ideas, giving feedback to the speaker while trying to influence him/her. This may or may not be successful and depends on both parties' ability to communicate. Moreover, this directing and regulating takes place without the audience's direct interference: sheer existence and presence of the audience makes the speaker keep to certain language rules and conventions. This characteristic feature makes it necessary to raise another important aspect of speaking i.e. accuracy and fluency.

Accuracy has two ranges of interpretation. In its original sense it means producing grammatically correct speech., while in a wider sense, it is concerned not only with grammar but also phonology, syntax, semantics and even discourse.

Fluency, on the other hand, is about the necessity of stopping to think (hesitating) while speaking. If the reason for stopping and hesitating is related to planning and formulating subsequent part of the dialogue, it indicates a limited command of language. This is especially true if pauses and hesitations are frequent. Tightly related to fluency is independence, which refers to the degree the speaker can use his / her own words rather than previously learned clichés to express ideas, as well as flexibility, which enables the speaker to give relevant replies to unexpected verbal stimuli. The third aspect of fluency is speed, which is measured in words per minute (wpm). In English, the normal range for native speakers is 180 ± 20 wpm. This speed is required at level C1 in IMETS SP and LC tests. Speech slower than this indicates the speaker's language use is somewhat restricted though it is acceptable at lower skills levels.

Striving for accuracy is another factor that influences fluency. From the psycholinguistic point of view, accuracy and fluency can be placed at two extreme points of a continuum. An individual trying to speak with 100 percent accurately most probably will lose out on fluency and vice versa. One who tries to speak very fluently may have good chances of becoming less accurate. It is always a matter of decision, which of the two contradicting qualities should dominate. In emergencies, for instance, accuracy is far less important than fluency, however, in an expert opinion given in court, accuracy can be crucial to such an extent that it overshadows

fluency. It can be said that there is no such thing as displaying 100 percent fluency or accuracy. Even native speakers violate rules of use and pause to process and formulate speech.

The situation is even more complicated when speaking is to be tested in well-defined, contexts. LSP contexts, in general, are a good example by being rather restricted and specific. However, it is impossible to find general or purely LSP situations. The language used, both in general and LSP situations is the same; in the case of EMP it is English. The difference between general and specific oral language use lies in the circumstances. The most important variables of the circumstances of language use include the personal features of the participants (speaker-listener) their roles and position in the social hierarchy. Speaking as a private person is totally different from speaking as a member of a discourse community (in our case it is the biomedical discourse community), which is associated with some background knowledge and expectations, communicative purposes and communicative tasks the participants are trying to solve. The direction of speaking is also crucial: speaking to a peer within the discourse community (internal discourse) is totally different from speaking to someone outside the discourse community (external discourse). All this applies to the style, register and genre used by the speaker, a fact that has important consequences in identifying the major factors of LSP speaking.

Finally, it should be remembered that culture is an umbrella term which embraces all of these aspects and has a great influence on them. Culture may be interpreted as"*the fabric of meaning in terms of which human beings interpret their experience and guide their action"* (Geertz 1973 quoted in Trompenar 1993:25). Since the topic of this chapter is LMP, 'culture' or 'cultural aspects' refer to a British-American culture, if such a one exists. It is unlikely that there is some kind of neutral, global English unrelated to culture. Such views are especially difficult to accept for LMP and LMP testing, as medicine is always practised as part of a culture and it is widely accepted that verbal communication is a crucial part of culture.

4.3   General principles of IMETS SP items

In testing receptive language skills, the texts serving as stimuli for the test takers are described, while in testing the productive skills the expected outcome has to be defined. The nature of the performance expected at each level can be best described in terms of communicative tasks and genres.

4.3.1 IMETS SP items at level B1

Task 1, level B1

The task is producing a monologue about the profession, education or research activity of the candidate. This task is meant as a warm-up activity, but it also has its value as an occasion where the production of monologues can be tested. It must be made clear that test takers can prepare for this task in advance of the test, and that this in no way reduces its value, especially if the way of delivery is optimised and adjusted to the testing purpose. In general, the most problematic deficiency of prepared speech is the lack of spontaneity, a most important characteristic of live speech. With few exceptions, events seldom exist where the speaker reproduces word for word what he / she has acquired by rote learning. The temptation to do so, at a language test, is enormous, especially at lower levels. Therefore, both test developers and examiners (interlocutors) of IMETS need to do their best to give the assessor a good chance for assessing the test takers' speaking skills and not their memory or ability to learn a text by heart. There are two ways for doing so. On the one hand, a carefully elaborated and annotated list of topics would increase the flexibility of preparation for the exam. At level B1 the phrase 'carefully elaborated' means a list which contains items that can be talked about at a factual and descriptive level. To enhance this feature, annotation, i.e. hints at sub-topics that may belong to the main topic area should be defined. This list will vary with skills levels. It is also a function of the interlocutor to make sure that the answer given by a test taker is spontaneous speech and not the product of rote learning. Their responsibility is to provide for spontaneity and flexibility by making remarks, asking for further information or further detail, all of which will prevent the candidate from reproducing prepared speech.

The list of topic areas for General Medicine Task 1 at level B1 follows below:

1) Choice of profession

2) Education for Health care professionals

3) Daily routine of the professional

4) Future of the profession

5) Research and development of the profession

6) Healthcare professional – patient relationship

7) Healthcare system

8) Prevention

9) Role of foreign languages in the healthcare profession

10) Conventional medicine vs. complementary and alternative medicine

The list of topic areas for Dentistry Task 1 at level B1 follows below:
1) Choice of profession.

2) Education for healthcare professionals.

3) Daily routine of the professional.

4) Future perspectives of the profession.

5) Research and development of the profession.

6) Dentistry professional - client relationship.

7) Healthcare system.

8) Preventive Medicine.

9) Role of foreign languages in healthcare.

10) Conventional vs. alternative medicine

The list of topic areas for Pharmacy Task 1 at level B1 follows below:
1) Choice of profession.

2) Education of pharmaceutical professionals.

3) Daily routine of the professional.

4) Future perspectives of the profession.

5) Pharmaceutical research and development.

6) Pharmacist-client relationship.

7) The healthcare system.

8) Preventive healthcare/ prophylaxis.

9) The role of foreign languages.

10) Conventional vs. alternative medicine.

The list of topic areas for Nursing Task 1 at level B1 follows below:

1) Choice of profession.
2) Education for healthcare professionals.
3) Daily routine of the professional.
4) Future of the nursing profession.
5) Research and development of the nursing profession.
6) Healthcare professional - client relationship.
7) Healthcare system.
8) Preventive Medicine.
9) Role of foreign languages in healthcare.
10) Conventional vs. alternative medicine

Test developers prepare up to five questions on each subject area, with B1 level requirements in mind, focusing on eliciting narrative and descriptive language.

Task 2, B1

According to IMETS specifications, the task sheet should contain five history taking topics which help the test taker elicit information on biodata during a preliminary history taking procedure. Each task sheet provided for the test taker will have its counterpart containing the relevant information for the examiners. The point of this is twofold: supplying the data required for the history taking helps the examiner to give standard answers when in a client's role and thus avoiding having to invent something on the spur of the moment. The latter requirement is very important and aims at increasing reliability. Each candidate, who draws a particular task sheet will be exposed to the task of the same difficulty, on the other hand, it gives good chances for the interlocutor to take the lead in this process and elicit the kind of language use that is appropriate. In this way, the interlocutor can also put forward ideas, to which the test taker can react spontaneously, thus demonstrating flexibility and the knowledge of the genre of medical history taking. A major feature of professional language use is its unrehearsed nature and the above approach allows this to be manifested in simulated situation.

There should be history taking topics on each task sheet prepared for test takers. Taking the part of an interviewer, the test taker asks three questions related to each prompt. The hints of the answers to these questions and some additional communicative tasks will be listed on the interlocutor's card.

When writing the prompts for the candidate, the following principles need to be taken into account:

1) Each prompt should contain upto three instructions on the required communicative task to be performed by the candidate, e.g.

   Enquire…, *Find out about…*, *Make sure …*, **Require repetition …**, **Express satisfaction …**, **Assure the patient about …**, **Signalize going on** / **finishing**, *Ask the patient to relate …*, *Ask the patient to give more details on …* **Relate your question to the data in the client's file**, *Ask the patient to confirm …*, *Encourage the patient* to …, **Greet the patient & introduce yourself., Give the reason for the interview. Give feedback on following the patient., Announce the next topic.,** etc.

   There are two distinct types of communicative tasks in this list. Those *written in italics concern medical facts*, which are the target of the interview, while those **written in bold type are used as meta-discourse**, which is an important feature of medical interviews and especially history taking. Of the communicative tasks listed, or made-up by the item writer, at least 3 different ones must be included within in each prompt listed on the task sheet.

2) The data originating from the (imaginary or real) client's file should be formulated as a list of words or phrases to avoid the use of reading out direct statements.

3) All four topic areas indicated in IMETS specifications should be covered. For General Medicine at level B1 the topic areas include:

   - Biodata
   - Present complaint
   - Past medical history
   - Family history
   - Social history

4) For Dentistry at level B1 the topic areas include:

   - Biodata
   - Presenting complaints
   - Past medical history
   - Past dental history
   - Social and family history

5) For Pharmacy at level B1 the topic areas include:

- Reason for visit
- Presenting complaints
- Medication history
- Recommendations/Advice
- Further steps and closure

6) For Nursing at level B1 the topic areas include:

- Greeting and introduction
- Admission questions
- Present complaints
- Past medical history
- Family, social history and closing

As far as the technique of writing prompts is concerned, it seems reasonable to start with writing the whole dialogue in advance, then extract the information for both test taker and interlocutor, which can then be formulated into appropriate prompts.

Prompts should be written so that, according to the specifications, the time of Task 2 at level B1 takes 7-8 minutes, which includes up to 2 minutes necessary for studying the prompts. This may raise the problem that the test taker is required to engage skills other than SP skills, i.e. skimming and scanning reading comprehension skills, in order to be able to perform the SP task. There may be some truth in this view, however, in real-life history taking, very similar activities are performed. The interviewer, as a rule, has some previous knowledge of the patient before starting the interview. Therefore, this requirement can be considered as an element increasing the reality of the simulation. Below is a sample task sheet for level B1. The task sheet consists of two major parts: a situation box and a prompts box.

Five important features of the situation serving as the basis for a simulated history are found in the situation box. First, the task is described in brief. Second, the location of the interview such as an imaginary hospital unit, clinical department, consulting room, emergency room, etc. is given Third, the time of the interview, if relevant, is given. Fourth, the role of the test taker is clarified. On being introduced, the test taker should use his / her own name and the appropriate

role title. (e.g. dr. XY). Fifth, the description of the patient / client, should contain the following data: name, age in years in brackets, reason for the interview e.g. referral, patient presents with complaint, patient is taken by ambulance, etc., followed by a brief description of the patient's present status, and any planned action. The number of words in the situation card at level B1 must not exceed 100.

The task sheet is a table with two columns. The first column contains the name of the 5 history taking topics as included in IMETS SP test specifications for Task 2 at level B1. Each box of the second column contains 3 instructions, at least one of which is about meta-discourse, while the rest is for eliciting factual information from the client. The item writers should make sure that acting upon the two types of instructions would result in an appropriate medical interview.

Below is an example of a task sheet for task 2 at level B1 for Dentistry candidates:

Situation
Use the information presented in the table below to interact with the client in a professional way.
Location: Dental surgery
You: Dentist
Patient: Mr./Mrs. Campbell, a 51-year-old patient, presents with a swelling on his cheek and severe, persistent, throbbing toothache.

	TASK
Biodata	- Greet the patient and introduce yourself. - Find out about four pieces of biodata. - Signal that you move on to the chief concern.
Presenting complaints	- Clarify the nature of the complaint. - Enquire about the onset and duration of the complaint. - Find out more details about triggering and soothing factors.
Past medical history	Ask the patient about: - respiratory disorders - diabetes - cardiovascular problems and allergies
Past dental history	- Find out about the last dental visit. - Ask about previous oral dental surgeries. - Ask about the loss of teeth.
Social and family history	Ask the patient about: - smoking status and alcohol consumption - complementary therapies used. Tell the patient the possible ways of dealing with this problem (antibiotics and root treatment).

Below is an example of a task sheet for task 2 at level B1 for Pharmacy candidates:

Situation
Use the information presented in the table below to interact with the client in
professional way.
Location: Pharmacy
You: Pharmacist
Patient: Mr./ Mrs. Simpson, a 52-year-old patient

	TASK
Reason for visit	- Greet the client. - Find out about the purpose of the visit (prescription/OTC medication). - Ask about previous experience and drug use.
Presenting complaints	Ask about: - current problem - the onset of the problem, aggravating and relieving factors - prescription/recommended treatment.
Medication history	Find out about: - drugs taken (OTC or prescription drugs) - dosage and frequency, route of administration (oral, topical etc.) - adverse reactions and allergies.
Recommendations/Advice	- Explain to the client what you recommend without prescription. (Valerian/ Melatonin/any other) - Find out what the client knows about the product. - Inform the client about dose, drug interactions; advantages; side effects; precautions. (as regards children, elderly patients, driving, etc.)
Further steps and closure	- Talk about the importance of professional (doctor's) advice. - Summarize the course of therapy (dosage). - Check understanding, encourage questions.

Below is an example of a task sheet for task 2 at level B1 for Nursing candidates:

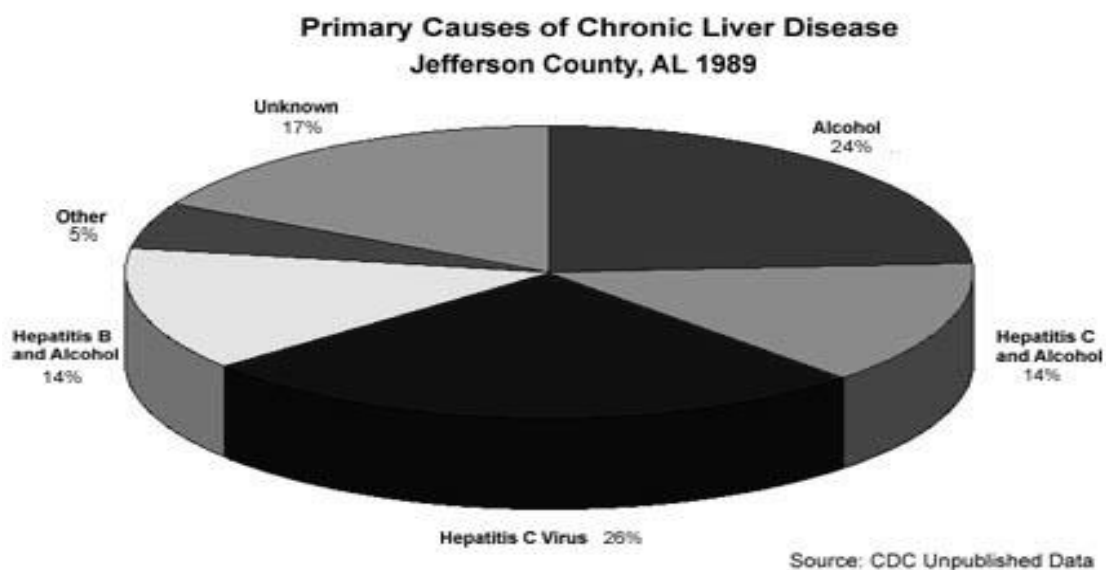| Situation |
|---|
| **Use the information presented in the table below to interact with the patient in a professional way.**
Location: Surgery ward
You: Surgery nurse
Patient: Mr./Mrs. Smith, a 40-year-old patient |

	TASK
Greeting and introduction	- Greet the patient and introduce yourself. - Explain your role and responsibilities. - Signal that you move on to admitting the patient to the ward.
Admission questions	Collect the following data from the patient: - name, age, occupation, address, phone number - marital status - next of kin.
Present complaints	Ask the patient about: - the reason for presenting to the hospital - the onset of the problem - pain related questions (duration, triggering, relieving, aggravating factors, etc.).
Past medical history	Ask the patient about: - previous diseases - treatment received, effectiveness of treatment - chronic diseases.
Family, social history and closing	Ask the patient about: - illnesses that run in the family - smoking, alcohol, coffee - work conditions, stress.

Task 3, B1

The task is graphic stimuli such as graphs, tables and figures containing information in the target language and should be used as the basis for developing test items. The most important criteria for selecting visual materials include the following:

1) The inclusion of visual materials should comply with copyright issues.

2) At level B1 visual aids predominantly contain easy-to-understand facts or descriptions.

3) The visual aids should be within the field of biomedicine

4) When selecting visual materials, test developers should have all three subgroups of the target population in mind, i.e. highly specific materials should be avoided.

5) Visual stimuli should be coloured.

6) In order to ensure the practical applicability of this task, in addition to finding visual materials, test developers should prepare a possible solution to this task, with B1 level test taker in mind. This information will be kept for internal use only.

Some visual material suitable for use as a Task 3 at level B1 is provided below:

Primary Causes of Chronic Liver Disease
Jefferson County, AL 1989

Source: CDC Unpublished Data

4.3.2   IMETS SP items at level B2

Task 1, level B2

Task 1 at level B2 is very similar to that of level B1. The only essential difference lies in the required outcome, which will be reflected in the questions written by the test developers for each topic area. In addition to eliciting narrative and descriptive language the questions at level B2 should provide opportunities for testing speaking about implied and global information. Test developers should use the same set of topics listed in point 4.3.1 above.

Task2, B2

Task 2, similarly to level B1, requires the candidate to take a simulated case history. The first seven history taking topics are similar to those indicated at level B1, i.e. biodata, present complaint, examination and investigation, past medical history, family and social history, diagnosis and recommendations, follow up and closing the examination. These history taking topics vary for the different fields. Please see the sample task sheets below for General Medicine, Dentistry, Pharmacy and Nursing.

Please see below a B2 level task sheet for General Medicine candidates:

Situation
Use the information presented in the table below to interact with the patient in a professional way. Location: Gastroenterology Clinic You: Gastroenterologist Patient: Mr./Mrs. Lewis (66), a recently retired patient was referred to the department by His/her GP. Suspected diagnosis: ulcer. He needs further investigations and advice on appropriate treatment plan.

	TASK
Collecting demographic data	Ask about: - name, age, address, phone number - marital status - health insurance - occupation.
Presenting Complaints	Enquire about: - reason for presenting to the healthcare institution - current health and illness status - the onset of complaint - relieving/aggravating factors.
Examination &Investigation	Inform the patient about the vital signs that you need to take: - respiratory rate, heart rate, blood pressure - temperature - talk about the investigations ordered by the GP - advise psychological counselling.
Past medical history	Ask about: - previous visits at the hospital/GP. - medication history including over-the-counter and herbal remedies - cardiovascular, respiratory, abdominal, neurological, musculoskeletal and genitourinary problems - sleep patterns.
Family & social history	Ask about: - illnesses that run in the family - smoking, alcohol, drugs - life circumstances, who is there to support him - ask about work, living conditions.
Diagnosis & recommendations	Talk about the suspected diagnosis. Advise the patient: - about drug treatment - on how to change dietary habits - to quit harmful habits (alcohol drinking, etc.).
Follow up and closing the examination	- Summarize your recommendations. - Check understanding. - Schedule follow-up. - Close conversation.

Please see below a B2 level task sheet for Dentistry candidates:

Situation
Use the information presented in the table below to interact with the client in a professional way. Location: Dental surgery You: Dentist Patient: Mr./Mrs. Jackson, a 30-year-old patient presents with a complaint of regular bleeding of her gums and pain while brushing the teeth.

	TASK
Biodata	- Greet the patient and introduce yourself. - Ask about 4 pieces of biodata. - Signal that you move on to the chief concern.
Presenting complaints	- Clarify the nature of the complaint. - Enquire about the onset of the complaint. - Enquire about the duration of pain. - Find out more details about triggering and soothing factors. - Signal the end of history taking.
Past medical history	Ask the patient about previous or chronic diseases: - asthma, allergies - hypertension - stroke, hepatitis - jaundice, epilepsy.
Past dental history	- Find out about the last dental visit. - Ask about previous oral dental surgeries. - Ask about loss of teeth. - Explain the importance of regular visits.
Social and family history	Ask the patient about: - smoking / alcohol consumption - coffee, tea - complementary therapies used.
Reviewing systems	Ask the patient questions related to the different systems: - respiratory system - cardiovascular problems - gastrointestinal system.
Further steps	- Signal the end of the interview. - Summarise what you have learnt about the case. - Inform the patient about treatment options (calculus removal/scaling, paradontological treatment, oral hygiene). - Ask the patient if he/she has any further questions. - Close the interview and arrange an appointment.

Please see below a B2 level task sheet for Pharmacy candidates:

| **Situation** |
| **Use the information presented in the table below to interact with the client in a professional way.** |
| Location: Pharmacy
You: Pharmacist
Patient: Mr./Ms. Jackson, a 30-year-old patient presents with a complaint of burning, itchy eyes. |

	TASK
Reason for visit	- Greet the patient and describe the purpose of the interaction. - Find out what the patient needs (prescription/OTC medication). - Acquire information about previous usage and experience.
Presenting complaints	- Current problem and chief concern for visit. - Enquire about the onset of the problem, provoking, relieving factors. - Quality and severity. - Find out about the treatment received. - Understand the impact on life.
Past medical history	Ask the patient about: - previous diseases of the eye or allergies - applied treatment - effectiveness of treatment - chronic diseases - treatment for chronic diseases.
Medication history	Find out about the administered drugs (OTC or prescription drugs) - dosage and frequency - route (oral, topical etc.) - past medications and adherence (consistent with taking the medications) - adverse reactions and allergies.
Assistance/Instructions	- Explain to the patient what was prescribed (eye drops for the infection). - Find out what the patient knows about the product. - Give instructions on dosage (one drop), route of administration (drop into each eye), frequency (4 times a day), duration of therapy (5-7 days).
Side effects and drug interaction	- Inform the patient about drug interactions (no contact lenses). - Explain the advantages (relieve symptoms). - Explain the side effects (redness, itching, burning, stinging pain). - Explain the precautions (children, driving, pregnancy, storage, overdose, expiry).

Further steps and closure	- Offer available options (generic option, less expensive). - Summarize the regimen. - Check understanding, encourage questions. - Close the interview.

Please see below a B2 level task sheet for Nursing candidates:

| **Situation**
Use the information presented in the table below to interact with the client in a
professional way. |
|---|
| Location: District Hospital, General Surgery Department
You: Nurse trained in preoperative nursing
Patient: Mr./Mrs. Simpson, a 59-year-old patient who is scheduled to undergo open-heart surgery |

	TASK
Present complaints	Ask about: - current health and illness status - the patient's understanding of the surgery - previous experiences if any - specific concerns or feelings.
Informed consent/ Insurance	Explain the necessity of signing an informed consent for: - surgical intervention - a scientific study. Ask information about health insurance.
Preparing the patient for surgical intervention	Tell the patient about: - the medical examinations ordered by the physician and your involvement (assistance, schedule, etc.) - fasting/enema before surgery - bathing or cleaning/shaving the area to be operated on - advise on clothes to wear.
Postoperative care toward sustaining care	Talk about: - care at bedside after surgery - wound care - administration of drugs/pain management - maintaining fluid and electrolyte balance - bowel movements.

Visitors	Inform the patient: - about visiting hours - hospital rules on visiting patients - about food/flowers/drugs etc. introduced by family and friends - hospital pass for a family member for permanent monitoring after surgery.
Dietary advice	Talk about: - adequate nutritional support after surgery - restrictions to family members to bring food in the ward after surgery - dietary plans based on the patient's needs after surgery.
Discharge from the hospital and further advice	Inform about: - discharge planning (approx. when/ how to leave the hospital, etc.). Make recommendations for home care: - palliative care, nursing associations providing care - wound care, positions in bed - personal hygiene.

Task 3 at level B2

Task 3 at level B2 is somewhat more complex than at level B1. At level B2, visual stimuli, in addition to requirement of extracting facts and giving descriptions, should also require explanation and interpretation of these facts as well as some mental activities like comparing, contrasting, making generalisations and inferences. The graph that follows below is an example:

Distribution of Cancer Incidence by Cancer Type and Sex Massachusetts, 1997-2001
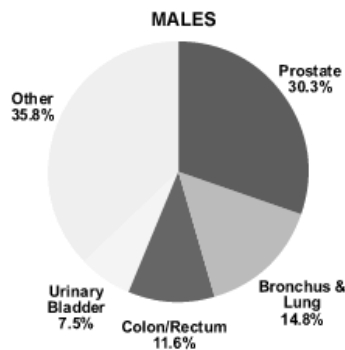
Statistics are from Cancer Incidence and Mortality
in Massachusetts, 1997-2001: Statexwide Report

4.3.3   IMETS SP items at level C1

Task 1, level C1

At first glance SP Task 1 at level C1 is very similar to Tasks 1 at levels B1 and B2. The list of 10 topic areas remains the same for this level (for the list see 4.3.1 above). There is a crucial difference, however, in implementing the task, which requires test developers to write questions on each topic area with C1 requirements for speech in mind. In addition to producing oral narratives and descriptions containing factual, global and implied information, the major requirements characteristic for level C1 EMP speech include expository and argumentative language. Therefore, up to five questions should be written for each topic area, which provide opportunities for testing expository and argumentative language use.

The topic areas should be presented by the test takers in a way that they express opinions, contrast advantages and drawbacks of certain measures, applying a critical approach to certain issues and discussing several aspects of some phenomena, entities and arrangements. In practical terms of test development this means that the questions to be written should be stimulating, and thought-provoking. The ideal question can easily be turned into a prompt for complex communicative task, which can be performed by the test taker without further explanation. For instance, a question like "*What do you think about the health education system in your country?*" is a form of question which encourages the candidate to perform an existing communicative task, and is a better way of achieving this than saying "*Discuss the advantages and drawbacks of the health education system in your country.*".

Task 2, C1

SP Task 2 at level C1 is totally different from those at the lower levels. It consists of a simulated interview between a health care professional and a client. Although it is a dialogue, because of the explanatory and argumentative functions required at level C1, the task taker is the principal performer assuming the role of the professional. The interlocutor does, however, play an active part in order to simulate the conversational characteristics of the test, and is required to ask 1 to 3 relevant questions. In order to ensure spontaneity of speech the exact questions will not be prescribed; in addition the test will focus on professionalism as it relates to the three groups targeted by IMETS.

The stimulus contains a brief summary of the situation (between 300 and 400 words), the test taker's task and ten task prompts. A sample sheet of the type given to candidates is provided below for General Medicine candidates:

Situation
Use the information presented in the table below to interact with the patient in a professional way.
Location: County Emergency Hospital, General Surgery Clinic
You: Surgeon on duty
Examiner: 21-year-old patient

FACT SHEET

A 21-year-old patient has been brought to the emergency room with severe burns on his/her face. The patient suffered burns when a hot steam pipe broke at his/her workplace (dry cleaner). The patient is in much pain, and the skin on his/her face is seriously damaged.

1. **Skin grafts** are applied after serious injuries when some of the body's skin is damaged. Surgical removal (excision or debridement) of the damaged skin is followed by skin grafting.
A thin layer is removed from a healthy part of the body (the donor section), or a full-thickness skin graft, which involves pitching and cutting the skin away from the donor section. A full thickness skin graft is risky, in terms of the body accepting the skin, yet it leaves only a scar line on the donor section, similar to a Caesarean section scar.
For full thickness skin grafts, the donor section will often heal much more quickly than the injury and is less painful than a partial thickness skin graft. Recovery time from skin grafting can be long. Patients should wear compression garments for several months and should be monitored for depression and anxiety endemic to long-term pain and loss of function.
The advantage is that patients can be admitted immediately to the hospital and the intervention is financially supported.

2. ReCell is a stand-alone, rapid, autologous cell harvesting, processing and delivery technology that enables surgeons and clinicians to treat skin defects using the patient's own cells in a regenerative process. This new technique is only available in private clinics and is quite expensive.
Developed as an 'off the shelf' kit, ReCell enables a thin split thickness biopsy, taken at the time of the procedure, to be processed into an immediate cell population for delivery onto the surface of the treatment area using a highly efficient, easy-to-use proprietary 'spray-on' application process. Tissue collection, cell segregation and preparation of the cell suspension takes approximately 20-30 minutes in total during which time the treatment area is prepared. Once processed, the cell suspension is available for immediate use and can cover a treatment area up to 80 times the area of the donor biopsy.
Advantages include:
· Minimisation of donor site size and depth with a concomitant reduction in complications, morbidity, and healing time.
· Improved wound healing time and scar quality.
· Repopulation of melanocytes to reduce hypopigmentation.
· On-site processing for immediate application.
· Increased viability through immediate harvest and application.
· Ability to be processed by a clinician without the need to involve laboratory staff.
Needs previous scheduling because there is a waiting list.

Act out the previously presented situation based on the following tasks.
Note: *You will be assessed on your ability to communicate, not on your reading comprehension or your medical knowledge in the field of medicine.*

TASKS
1. Ask the patient about the accident.
2. Explain to the patient the procedure of skin grafting.
3. Talk about the disadvantages of skin grafting.
4. Tell the patient about the advantages of skin grafting.
5. Explain the procedure of the new ReCell technology.
6. Explain the advantages of the new ReCell technology.
7. Talk about the disadvantages of the new technology.
8. Compare the two procedures.
9. Ask the patient if he/she can afford treatment at a private clinic.
10. Discuss the various options the patient has and issues such as his/her employer's responsibility.

A sample sheet of the type given to candidates is provided below for Dentistry candidates:

Situation
Use the information presented in the table below to interact with the patient in a professional way.
Location: Dental Clinic
You: Dentist
Examiner: 67-year-old patient

FACT SHEET
A 67-year-old patient, after missing teeth for a long time due to an accident, would like to receive new teeth in positions upper right 4 and 6 with the help of implants.

A **dental implant** is a surgical fixture that is placed into the jawbone and allowed to fuse with the bone over the span of a few months. The dental implant acts as a replacement for the root of a missing tooth. In turn, this "artificial tooth root" serves to hold a replacement tooth or bridge. Having a dental implant fused to the jawbone is the closest thing to mimicking a natural tooth because it stands on its own without affecting the nearby teeth and has great stability. Most dental implants are made of titanium, which allows them to integrate with bone without being recognized as a foreign object in our body.

When it comes to **tooth replacement**, generally, there are **three options:**
1. removable dental appliance (complete denture or partial denture),
2. fixed dental bridge (cemented), and
3. dental implant.

Dentures are the more affordable option for replacement teeth but are the least desirable because of the inconvenience of a removable appliance in the mouth. Furthermore, dentures can affect one's taste and sensory experience with food.

Dental bridgework was the more common restorative option prior to the relatively recent shift to dental implant treatment. The main disadvantage to bridgework is the dependence on existing natural teeth for support. Implants are supported by bone only and do not affect surrounding natural teeth.

Deciding on **which option to choose depends on many factors.** Specifically for dental implants, these factors include
- location of missing tooth or teeth,
- quantity and quality of the jawbone where the dental implant is to be placed,
- health of the patient,
- cost, and patient preference.

A dental surgeon examines the area to be considered for the dental implant and makes a clinical assessment of whether the patient is a good candidate for a dental implant. There are great advantages to choosing a dental implant for tooth replacement over the other options. Dental implants are conservative in that missing teeth can be replaced without affecting or altering the adjacent teeth. Furthermore, because dental implants integrate into the bone structure, they are very stable and can have the look and feel of one's own natural teeth.

Source: https://www.medicinenet.com/dental_implants/article.htm

Act out the previously-presented situation based on the following tasks.
Note: *You will be assessed on your ability to communicate not on your reading comprehension or your medical knowledge in the field of medicine.*

TASKS

1. Ask the patient about the reason for the visit.
2. Ask about the missing teeth (causes, replacement).
3. Tell the patient about dental implants.
4. Explain the surgical procedure of dental implants to the patient.
5. Explain the outcomes of dental implant placement to the patient.
6. Tell the patient all the options available for tooth replacement.
7. Explain the disadvantages of dentures and bridgework.
8. Explain the most important factors of clinical assessment for dental implants.
9. Summarize the advantages of choosing dental implants.
10. Ask the patient about her concerns and assure her about the success of the procedure.

A sample sheet of the type given to candidates is provided below for Pharmacy candidates:

Situation
Use the information presented below to interact with the client in a professional way.
Location: Pharmacy
You: Pharmacist
Client: University student

FACT SHEET

Situation: You are a licensed pharmacist and work in a pharmacy where you sell natural remedies as well as other products. A student, who is in the middle of the examination period, comes with a prescription for *Ritalin* and he asks you to recommend some anti-stress drugs and stimulants. Fellow students have told him/her about drugs they usually use when they sit exams.

Many university students are relying on performance enhancing drugs to help them boost their averages. They often abuse prescription drugs such as anti-stress medications, antidepressants, or stimulants such as **Adderall**. *'Study drugs' like* **Ritalin**, *the first drug to treat attention deficit hyperactivity disorder, have been widely used by normal students hoping to be extra sharp while taking exams. Students hope their problems will fade with the dissolving of a pill, but it could even make things worse.*

*Naturopathic nutrients and botanical supplements that can **help reduce stress and maintain productivity**:*
- magnesium, lavender preparations, etc.
- **Wu Wei Zi** herbal tea. The Schizandra berries tea has been used in Chinese medicine for centuries.

Memory Supplements:
- **Ginkgo biloba** is a supplement that tends to improve blood flow in small vessels. Studies show benefits in improving mood, alertness, and mental ability in healthy people.
- **Mind Power Rx – a** mind supplement and memory pill has mixture of more than **15 herbal ingredients, vitamins, and amino acids** that improve mental alertness and wakefulness along with boosting mental stamina and helping maintain healthy memory.
- **Myco Formulas Memory** nourishes your nervous system with four **medicinal mushrooms** that have been shown to safely and sustainably promote healthy brain function.
- **Choline,** is an **organic, water-soluble compound**. It is neither a vitamin nor a mineral. However, it is often grouped with the vitamin B complex due to its similarities. It helps the normal development of the brain and memory enhancement.

Act out the previously presented situation based on the following tasks.
Note: *You will be assessed on your ability to communicate not on your reading comprehension or your knowledge in the field of pharmacy.*

TASKS

Talk about:

1. the prescription and why the student thinks he/she needs drugs when sitting examinations
2. the use of stimulants (energy drinks, coke, coffee, etc.) and ask about their use
3. combining stimulants with drugs
4. drugs college students often use to study (Adderall, Ritalin)
5. antidepressants and memory enhancement drugs used by healthy people
6. the side-effects of psychopharmaceuticals
7. overdose.

Suggest:

8. natural products that can reduce stress (magnesium, lavender, tea, any other product)
9. products that can enhance memory (Ginkgo biloba, Mind Power Rx, Myco Formulas Memory, Choline, any other products)
10. Discuss your professional opinion related to some of the above-mentioned issues and suggest consulting a psychologist.

A sample sheet of the type given to candidates is provided below for Nursing candidates:

Situation
Use the information presented below to interact with the patient in a professional way.
Location: Psychiatry Clinic of the County Emergency Hospital
You: Psychiatric nurse
Examiner: Patient

FACT SHEET

You are a registered psychiatric nurse, trained for nursing care in electroconvulsive therapy (ECT), working at the Psychiatry Clinic of the County Emergency Hospital. You have to prepare a 29-year-old patient for ECT. He lives with his elderly mother and he was brought into the hospital by family (brother/sister) after the neighbours had reported to the police that they saw him in the house, chasing his mother with a knife. He was immediately admitted to the psychiatric ward. Drug treatment failed to work. Now you are preparing the patient for ECT. The hospital would rather not take responsibility for the patient's valuables (jewellery, money, etc). Usually family members are asked to take them home.

ECT - Electroconvulsive therapy is a procedure, done under general anaesthesia, in which small electric currents are passed through the brain, intentionally triggering a brief seizure. ECT seems to cause changes in brain chemistry that can quickly reverse symptoms of certain mental health conditions. It is a minor surgical procedure that requires preoperative preparation and postoperative care. Personal items (prostheses, dentures, glasses, hearing aids, contact lenses, piercings, or earrings) should be removed before the procedure.

Although ECT is generally safe, **risks and side effects** may include:
- **Confusion**. Immediately after treatment which can last from a few minutes to several hours. Rarely, confusion may last several days or longer. Confusion is generally more noticeable in older adults.
- **Memory loss.** Some people have trouble remembering events that occurred right before treatment or in the weeks or months before treatment or, rarely, from previous years. These memory problems usually improve within a couple of months after treatment ends.
- **Physical side effects**. On the days of an ECT treatment, some people experience nausea, headache, jaw pain or muscle ache. These generally can be treated with medications.
- **Medical complications**. As with any type of medical procedure, especially one that involves anaesthesia, there are risks of medical complications. During ECT, heart rate and blood pressure increase, and in rare cases, that can lead to serious heart problems.

Act out the previously presented situation based on the following tasks.
Note: *You will be assessed on your ability to communicate not on your reading comprehension or your medical knowledge.*

TASKS

1. Explain why ECT is necessary.

Inform the patient about:
2. the procedure of ECT
3. side effects
4. medical complications
5. informed consent
6. safekeeping of valuables.
7. Prepare the patient before the procedure (bathroom, etc.)
8. Tell about the removal of personal items.
9. Reassure and support the patient.
10. Ask if the patient understood everything and if he has any further questions.

Task 3, C1

SP Task 3 is seemingly similar to the task at level B2. The communicative skills needed to solve this task, however, are much more challenging. In harmony with the task at level B2, the test taker is expected to give a detailed analysis of a graph, table or figure containing information in the target language but also, the test taker should be able to elaborate on the topic represented by the visual stimulus. Because of this additional task, some Task 3, level B2 visual stimuli might also be appropriate even at level C1 such as the one presented in 4.3.2. The graph that follows below is an example:

Describe and explain the following diagram. Present the detailed information in 5-7 minutes.

**Region-Specific Incidence Age-Standardized Rates
by Sex for Cancers of the Lung in 2018**
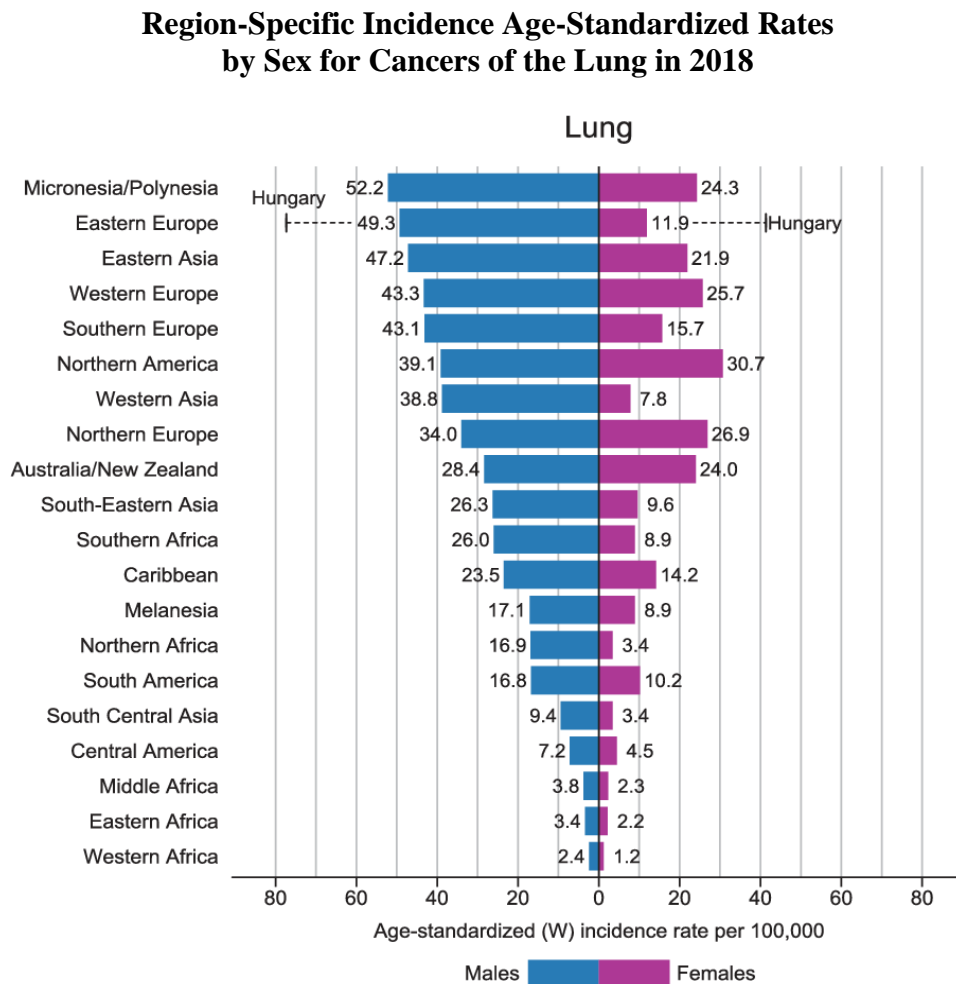
Source: GLOBOCAN 2018

5. WRITING

5.1 Background

Writing can be viewed as a process or a product. For teaching purposes, many prefer to view it as a process, while for the purposes of testing it will be seen as a product. In the IMETS testing system, attention is centered only on official writing and academic writing. Official letters in a professional context and written medical case reports seem to be a relevant choice, as both types of writing are part of healthcare professionals' routine activity. Another reason for choosing these two types of writing is practicality, i.e. the time necessary for creating samples within these two fields is relatively short, which is a major requirement for testing.

5.2 The construct of writing skills

Writing skills include a wide range of sub-skills, which are mainly physical or physiological. As such they are considered irrelevant for language testing purposes. What is important to make clear is that by writing we do not understand the physical part of it. In IMETS writing is an activity that results in an intellectual product readily available for inspection and evaluation. Criteria of evaluation of any written product including IMETS official letters and case report presentations in writing include the following traits:

1 Ideas & Content

2 Conventions & Organization

3 Word Choice & Voice

4 Sentence Fluency

It is ideas and content that is specific about communicating professional content. This content should not contradict the background knowledge of an informed lay person, however, for reasons listed above, it cannot be judged from a scientific point of view. It can be evaluated from the aspect of coherence. As far as the organization of the written product is concerned, it is important to follow some kind of logic. These formal written performances should be presented in accordance with conventions, i.e. using paragraphs to express each idea separately, cohesion markers and other formal attributes characteristic of formal letters. The formal voice should be used in these writings and this is associated partially but not exclusively with word choice. This also involves using an emotionless, impersonal, impartial and objective style and the knowledge of how to fulfil communicative functions, such as expressing suspicion, degrees of certainty, probability, possibility and necessity. Furthermore, it involves the use of these

functions in expository texts for argumentation. Sentence fluency is mainly an aesthetic nature, contributing to the smoothness of the text, a feature very difficult to define precisely. Finally, in addition to the implications mentioned above, conventions also include spelling, grammar rules, rules of decency and formality, formats, text types and genres. We can say that conventions are intertwined with all the other five traits of writing and this is especially true for EMP writing at level C1, where the degree of independence of language use is expected to be nearly 100 percent. The traits listed above will be revisited in the Manual for IMETS Evaluators, where they can serve as the basis for elaborating descriptive bands for assesment.

5.3   Requirement for the tests produced at IMETS WR skills testing

Before addressing the individual tasks at each of the three skills levels, it seems necessary to summarize some views concerning academic writing. As has already been pointed out, academic writing is expository and / or argumentative. Generally speaking, academic writing should be precise, objective, impersonal, formal or semi-formal. Surprisingly these features can be best realized by relying on other people's views. Due to its expository or argumentative character, good academic writing is a real or simulated discussion, challenging or provoking real or imaginary people who have alternative views on a subject. The 'They say' vs. 'I say' method should always be in the background, as it constitutes the basis for many kinds of research. This method may give academic writing a vivid, many times even enjoyable characteristic.

5.3.1   IMETS WR items at level B1

Task 1, level B1

Task 1 at level B1 is a letter writing task. The required product is an official letter concerning factual information on biomedical topics. The stimulus for the test takers includes 10 prompts such as the name of the sender and the addressee, the date, the main communicative purpose and six facts to write about. The test takers are required to write an official letter 130-150 words in length, which may be either a congress application, an invitation or an acknowledgement letter. The formal features of the official letter i.e. sender's and addressee's names, address, date, closing remark and signature do not count in the length, however their presence, arrangement and correctness do. The assessment will be made taking the traits of writing mentioned above, into consideration. Naturally these traits will be adjusted to the level of

professional language use to be tested. Detailed descriptive bands for assessors will be provided in the Manual for Evaluators.

The test developers should use one of the following genres at level B1:

- job application
- congress application
- invitation
- acknowledgement

Major requirements for writing the instructions and prompts include the following:

1) Both the instructions and the prompts should be very brief and to the point.
2) The prompts are not supposed to be full sentences or long pieces of texts.
3) The prompts should be hints consisting of words or phrases, which are preferably nominalised, to avoid word-for-word utilisation by the test taker.

The sample letter, which should also be prepared as part of test development, is not meant to be the sole solution to this task. It is necessary for test developers to check how their own ideas can be put into practice. This sample answer also serves as an example of acceptable solution, should anyone make inquires after the test.

To illustrate the above, a sample letter writing task and a sample letter are provided below.

Candidate's sheet

Level B1, Letter Writing

Instruction: Write an official letter in 130-150 words. The letter you are asked to write is a congress application in which you are required to touch upon the following ten prompts:

1) Sender's name: John/Jane Brown (23 Charles Street, Edinburgh, EH8 9AD)
2) Addressee's name: Whiston Hospital (25 Kinley Road, Dublin 8, Ireland)
3) Date: 10/04/2012
4) Communicative purpose: Participation and giving presentation in the conference 'The Impact of the Global Financial Crisis on Health Care'
5) Info about the conference: January 2012, journal of Irish Healthcare
6) Application, title of presentation: Measuring effectiveness at the Child Clinic of Edinburgh University.
7) Attachment: abstract
8) Request: feedback about acceptance of presentation
9) Request: official invitation + application form
10) Inquiry: payment

Sample letter

23 Charles Street,
Edinburgh,
EH8 9AD
Scotland

Whiston Hospital
25 Kinley Road
Dublin 8
Ireland

10/04/2009

Dear Sir/Madam,

I am writing in response to an advertisement placed in the January issue of 'Irish Healthcare', which calls for attendees and presenters at the 'The Impact of the Global Financial Crisis on Health Care' conference, to be held in Dublin, Ireland from15-17 October 2009.

I would like to register as both an attendee as well as a presenter. The title of my presentation is: 'Measuring Effectiveness at the Child Clinic of Edinburgh University'. I would appreciate it if you could inform me regarding the acceptance of the presentation at your earliest convenience. You will find the abstract for the presentation enclosed.

I also would be grateful if you would send to me, via e-mail, both the official invitation as well as the registration form. My address is: jbrown@hotmail.com.
I would also like to know what the preferred payment registration method is: bank transfer or internet bank transfer.

Looking forward to your reply,

Yours faithfully,

John/Jane Brown   149 words

Task 2, B1

Task 2 at level B1 requires the test takers to present a medical case report in writing. The purpose of this task is to give an opportunity for testing skills of descriptive academic writing. At this level, explanation or interpretation are not required. Detailed criteria for assessment will be given in the Manual for Evaluators.

The test developers should select an authentic case report and, based on this, they should give ten prompts requiring factual interpretation about the symptoms, the examinations and

investigations performed, about their results, the diagnosis, the treatment, and the evaluation of the treatment. Copyright issues must be taken into account.

Test developers should also provide a sample solution in 130-150 words.

5.3.2   IMETS WR items at level B2

Task 1, level B2

Task 1 at level B2 is similar in form to that at level B1. First, the sender' data and addressee, the date of writing and the communicative purpose to be achieved in the official letter, are given. Following this there should be six facts, at least three of which are pragmatically oriented, e.g. they express different degrees of certainty, probability and necessity. Another difference is the length of the expected letter, which, at this level should be 180-200 words. The third difference is in the genres, which at this level include the following:

- congress application
- support letter
- complaint
- referral
- cover letter

A sample letter should also be provided by the test developers for reasons detailed at level B1.

To illustrate the above, a sample letter writing task and a sample letter are provided below.

Candidate's sheet

Level B2, Letter Writing

Instruction: Write an official letter in 180-200 words. The letter you are asked to write is a referral in which you are required to touch upon the following ten prompts:

1) Sender's name: Dr. Paul/Paula Johnson (Family Health Practice, 59 Rutland Road, Southall UB1 2UR, UK)
2) Addressee's name: Prof. Peter Forbes, head of Dept. of Gastroenterology (University College Hospital, 235 Euston Road, London NW1 2BU)
3) Date: 04/04/2009
4) Communicative purpose: Referral letter for Mary Murphy to Prof. Peter Forbes
5) 1st presenting 01/03/2009, symptoms: constant vagueness, general malaise
6) Info on patient: 45 yr., 170 cm 58 kg, weight loss: 5kg within 2 months, smoking (-), alcohol (-)
7) tests performed: routine blood, result: anaemia, therapy: iron tablets
8) Patient presented again (1 week ago); condition: worsened; symptoms: pale, fatigue, strong palpitation; test ordered: abdominal ultrasound (negative), stool: Weber test (negative)
9) Required: gastroscopy and colonoscopy to exclude malignancies
10) Request: advice, surgical treatment

Sample letter

<div style="border:1px solid">

Family Health Practice
59 Rutland Rd
Southall UB1 2UR
UK

Prof. Peter Forbes
Head of Department
Department of Gastroenterology
University College Hospital
235 Euston Road
London, NW1 2BU
UK   04/04/2009

Dear Professor Forbes,

My name is Paul/Paula Johnson, I am writing to you as a family doctor working in Southall. I would like to refer my patient, Mary Murphy to you and would be grateful if you could perform further examinations.

She first presented to my office on 1 March, 2009 complaining of constant tiredness and malaise. The patient is 45, she weighs 58 kgs and is 170 cms tall. She has lost 5 kgs in the last two months, and neither smokes nor drinks alcohol. On the day of presentation I conducted a routine blood test that revealed anaemia and I prescribed her iron tablets.

However, her condition worsened and she came back to see me a week ago. On evaluation I found pallor, weakness and heavy palpitation. I ordered an abdominal ultrasound scan and Weber stool examination, both of which proved to be negative.

I would like to ask you to perform further examinations, including both a gastroscopy as well as a colonoscopy in order to exclude any possible malignancies.

I would welcome your assessment and advice on further treatment.

Thank you for your help in advance.

Yours sincerely,

Paul/ Paula Johnson, MD
Family Doctor   183 words

</div>

Task 2, level B2

Task 2 at level B2, similarly to that at level B1, requires the test taker to present a medical case in writing. The difference is that, in addition to giving factual information, the test taker should be able to give implied interpretation on the basis of the 13 prompts provided. Another difference is in the length of the product, which should be 220-240 words.

The test developers should select an authentic case report and, based on this, they should give 13 prompts requiring both factual and implied interpretation about the symptoms, the examinations and investigations performed, about their results, the diagnosis, the treatment, and the evaluation of the treatment. Copyright issues must be taken into account.

Test developers should also provide a sample solution in 220-240 words.

5.3.3 IMETS WR items at level C1

Task 1, level C1

Task 1 at level C1 is similar in form to that at level B2. First, the sender' data and addressee, the date of writing and the communicative purpose to be achieved in the official letter, are given. Following this there should be six facts, at least six of which are pragmatically oriented, e.g. they express different degrees of certainty, probability, necessity and polite expression of disagreement. Another difference is the length of the expected letter, which, at this level should be 220-250 words. The third difference is in the genres, which at this level include the following:

- letter to the editor
- response to professional issues
- case report
- congress application
- support letter
- complaint
- referral
- answering reviewers comments

A sample letter should also be provided by the test developers for reasons detailed above.

A sample writing task with its solution is presented below

Candidate's sheet

Level C1, Letter Writing

Instruction: Write an official letter in 220-250 words. The letter you are asked to write is a response to a professional issue in which you are required to touch upon the following ten prompts:

1) Sender's name: dr. Márton/Mária Szabó (Department of Orthopaedics Surgery, University of Szeged, 6 Semmelweis Street, Szeged, Hungary, H-7625)
2) Addressee's name: dr. Steven Dombay (Birmingham Hospital, Birmingham, B152 TG, UK)
3) Date: 10 April, 2009
4) Communicative purpose: Reflect on a colleague's opinion
5) Summarize dr. Dombay' view:link between CVD and Ca supplementation in postmenopausal women.
6) Point out: research supports the idea: link between CVD and osteoporosis(OP).
7) Call the attention: OP drugs – atrial fibrillation – thrombus, embolus, myocardial infarction
8) Point out: the need for confirmation
9) Argue for: healthy life style, diet, exercise
10) Offer: further discussion of the matter

Dr. Steven Dombay	University of Szeged
Birmingham Hospital	Department of Orthopaedics
Birmingham	Semmelweis u. 6
B 152 TG	6725 Szeged
UK	Hungary

10 April, 2009

Dear Dr. Dombay,

Thank you for your letter of 10 March asking for my viewpoint concerning the potential link between Ca-supplementation and increasing risk of CVD in postmenopausal women.
You have brought up a very important issue as both osteoporosis and CVD are major health burdens affecting millions of people globally, especially elderly women.

I would like to point out that the growing body of research supports a direct association between CVD and osteoporosis. The latest studies show that people taking osteoporosis drugs to prevent the occurrence of osteoporosis were more likely to develop atrial fibrillation. It should be kept in mind that atrial fibrillation can be serious and can lead to the development of thrombus, embolus or myocardial infarction.

Nevertheless, the pathophysiological connection between the atherosclerotic and the osteoporotic processes needs further elucidation and warrants additional research in this area.

It is unequivocal that clinicians should be very cautious when choosing treatment for osteoporosis and weigh the risks against the benefit of decreased fracture risk.

In addition to all the above mentioned factors, we should emphasize the importance of healthy life style, diet and, especially, the role of adequate physical exercise.

If you have further questions, do not hesitate to contact me.

Yours sincerely,

Dr. Márton Szabó

Task 2, level C1

Task 2 at level C1, similarly to that at level B2, requires the test taker to present a medical case in writing. The difference is that, in addition to giving factual and implied information, the test taker should be able to give argumentative interpretation on the basis of the 15 prompts provided. Another difference is in the length of the product, which should be 300-320 words.

The test developers should select an authentic case report and, based on this, they should give 15 prompts requiring both factual, implied and argumentative interpretation about the symptoms, the examinations and investigations performed, about their results, the diagnosis, the treatment, and the evaluation of the treatment. Copyright issues must be taken into account.

Test developers should also provide a sample solution in 300-320 words.

6   VALIDATION OF THE TEST DEVELOPMENT PROCEDURE

6.1   IMETS TEST DEVELOPERS

Test developers for IMETS testing system are to be (1) EMP instructors and testers having at least three years of experience in the field or (2) native speakers of English involved in the health care profession.

Test developers for the IMETS testing system are required to take part in the IMETS Test Developers Basic Training which is composed of the following four parts:

1. familiarization with the Common European Framework of Reference for Languages and its Companion Volume
2. familiarization with the Descriptors developed for the IMETS testing system
3. familiarization with the Test Specifications developed for the IMETS testing system (TS)
4. familiarization with the Manual for IMETS Test Developers and Item Writers (MTD)

Test developers for the IMETS testing system are required to take part in a IMETS Test Developers' Training Session at least once a year. These further training events are organised by the Chief Examiners for the language being tested. Following either a direct or multiplicator training methodology these further training events are meant to discuss (1) the experience

gained from previous test development and evaluation and (2) plans for future test development and evaluation.

The Main Testing Centre (It has not yet been decided at which partner's institute, this will be located) will issue a 'IMETS Test Developer' certificate containing:

 (1) A certificate stating that the recipient has satisfactorily completed the IMETS Test Developers' Basic Training course.

 (2) a certificate stating that the recipient has satisfactorily completed the IMETS Test Developers' Further Training course.

In both cases, the certificated will be signed and dated by a principal(s) of the institution.

6.2   TEST VALIDATION

As has been mentioned, validity is a key concept in EMP test development and assessment. *"A test assessment procedure can be said to have validity to the degree that it can be demonstrated that what is actually assessed (the construct) is what, in the context concerned, should be assessed, and that the information gained is an accurate representation of the proficiency of the candidates concerned" (CEFR).*

Validation of the IMETS tests is undertaken in the following way:

(1) internal validation within the test developing teams, i.e. consulting the first draft with 2 other exam developers

(2) social validation (6.2.1.)

(3) native speaker validation: checking linguistic correctness by an English native speaker (responsibility of the test developing team)

(4) post-implementation technical validation

(5) internal and external validation by pre-testing and traditional statistical analysis (see point 6.3.2) and giving feedback by post-testing (see point 6.3.3)

(6) post-implementation validation by 2 external experts

The following feedback form is to be applied for each validation step:

…………………………………………………. **VALIDATION**
FEEDBACK FORM

Test item:	Decision:	CEFR level:	Justification:
1			
2			
3			
4			
5			
6			
7 ...			

Overall: **A**: acceptable   **M**: to be modified   **D:** to be discarded

This project has been funded with support from the European Commission. This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

6.2.1   Social validation

Social validation includes a feasibility assessment of the edited test variety by a physician who is familiar with the CEFR and the IMETS descriptors and test specifications.

6.2.2 Internal and external validation by pre-testing and traditional statistical analysis

Internal as well as external validation is performed by pre-testing and analysing the results by accepted methods of statistical analysis. Internal validation is carried out by pre-testing the IMETS test varieties developed for a given skills level. For external validation purposes, which is undertaken using the same pre-testing procedure, some external EMP RC and LC tests, e.g. PROFEX EMP tests will be incorporated. These external EMP tests will be calibrated to the appropriate CEFR levels.

It is a pre-requisite to keep the testing materials confidential. In order to provide for test security, candidates taking part in pre-testing shall not know that they are involved in pre-testing real IMETS tests and they are not allowed to take any notes, pictures or copies of the pre-tested materials.

For the purposes of pre-testing, 30 candidates shall be involved at each level. It is important that potential test takers shall not take part in pre-testing. It is also crucial that these candidates' language competence shall meet the requirements of the pre-tested CEFR level. A candidate may be included in the pre-tests if:

1) the candidate's EMP teacher evaluates his/her EMP competence appropriate to the CEFR level pre-tested or
2) the candidate holds an EMP certificate at the CEFR level pre-tested.

The Chief Examiners are responsible for coordinating pre-testing. Each Examination Centre in each partner institution in the IMETS project, shall be involved and be responsible for conducting pre-testing according to IMETS regulations. The rules and regulations for conducting pre-testing shall be in accordance with that of a real exam (for details see the Manual for Organising Exam Centres).

The procedure of pre-testing is provided below:

1) The Chief Examiners make it clear which Examination Centre and each partner institution in the IMETS project, will be involved in the pre-testing, and give details of exactly how many candidates will take part.
2) Using the master copy of the given IMETS test and an external CEFR calibrated EMP RC and LC test (e.g. PROFEX), the Chief Examiners will prepare the pre-tested materials, which shall not contain any IMETS sign or logo.

3) The Chief Examiners will send the pre-testing materials to the appropriate Examination Centre in a confidential way.

4) The Examination Centres will organise and conduct pre-testing in accordance with IMETS rules and regulations.

5) The IMETS test evaluators at the Examination Centre will evaluate the pre-tests following the guidelines described in the IMETS Manual for Test Evaluators.

6) The exam coordinators at the Examination Centre will enter the data of the evaluated pre-tests in the IMETS statistical analysis software and send it to the Chief Examiners by E-mail.

7) The Chief Examiners will perform traditional test analysis and evaluate the results.

8) On the basis of the results, the Chief Examiners will prepare a feedback for the test developers and send it to them in a confidential way.

9) Providing feedback exclusively about the score achieved on the pre-test for the pre-test takers is up to the decision of each Examination Centre.

6.2.3 Post-testing

The main goal of post-testing is to gain a reliable picture about the tests used in real test situations. This feed-back can provide useful information for further test development and for further trainings for test developers and evaluators. For practical reasons, it is not allowed to modifying the test scores based on the post-test results in IMETS.. The procedure of post-testing is given below:

1) The exam coordinators at the Main Testing Centre will select 30 tests at each level in a way that they select 10 tests, where the test takers achieved an outstanding total score, 10 tests, where the test takers achieved an average total score and 10 tests, where the test takers achieved poor results.

2) The exam coordinators at the Main Testing Centre will enter the data of the selected tests in the IMETS statistical analysis software and send it to the Chief Examiners by E-mail.

3) The Chief Examiners will perform traditional test analysis and evaluate the results.

4) On the basis of the results, the Chief Examiners will prepare a feedback for the test developers and evaluators and send it to the Main Testing Centre Office, which will be responsible for its archiving and distribution.

7. PROCEDURE OF IMETS TEST DEVELOPMENT

All testing material shall be in accordance with the description of the TS and the MTD and free of political, racial or sexual discrimination and personal right violence issues. The procedure of test development is provided below:

1) Searching suitable textual materials
2) Text mapping
3) Selecting the appropriate task format
4) Writing test items
5) Implementation of the paper on the online platform

REFERENCES

1. Bachman, Lyle (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

2. Bachman, L. F. (1991) What Does Language Testing Have to Offer? TESOL QUARTERLY, Vol. 25,No. 4.

3. Baker, M. (1988). Sub-technical vocabulary and the ESP teacher: an analysis of some rhetorical items in medical journal articles. *Reading in a Foreign Language, Vol. 4 No. 2. pp. 91-105*

4. G. Brown and G. Yule, *Discourse analysis*. Cambridge: Cambridge University Press, 1983.

5. Canale, M.; Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*

6. Council of Europe. Common European Framework of Reference for Languages; 1, 177. Available at http://www.coe.int/t/dg4/linguistic/cadre_en.asp., 2008

7. Council of Europe (2009). Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual. Strasbourg: Language Policy Division. Available at http://www.coe.int/t/dg4/linguistic/Source/ManualRevision-proofread-FINAL_en.pdf

8. Clapham, C. (1996). The development of IELTS: a study of the effect of background knowledge on reading comprehension. Cambridge: Cambridge University Press.

9. Davies, A. (1990). Principles of language testing. Oxford: Blackwell.

10. Douglas, D. 2000. *Assessing Languages for Specific Purposes*. Cambridge: Cambridge University Press.

11. Geertz, C. (1973). *Interpretation of Cultures.* New York: Basic Books

12. Henning, G. (1987). *Guide to Language Testing*. Heinle.

13. *http://archpedi.jamanetwork.com/article.aspx?articleid=1216765*

14. *http://czhaiers.en.made-in-china.com/product/OoCnXtLEZvhg/China-New-Circular-Stapler-for-Rectal-Prolapse-and-Hemorrhoids-YG-32-34-.ht*

15. *http://dictionary.reference.com*

16. *http://journals.lww.com/emnews/Fulltext/2011/11000/Letter_to_the_Editor__Reducing_Medication_Errors.15.a*spx

17. *http://jscr.co.uk/2012/05/urologic-surgery/malignant-mesothelioma-of-the-tunica-vaginalis/*

18. *http://outcomes.cancer.gov/areas/pcc/ communication/ pcc_monograph.pdf*

19. *http://thegrantdoctor.com/samples/sample-proposal-aids-hiv-prevention-grant/Sample-Proposal-AIDS-HIV-Prevention4.html*

20. *http://www.careers.monash.edu.au/assets/docs/sample-pharmacy-intern-cover-letter.pdf*

21. *http://www.healthoneems.com/Agency_Specific_Protocols/NSMC_2012_System_Protocols_1.pdf*

22. *http://www.jobs2me.net/bbs/Topic.aspx?BoardID=53&TopicID=2105*

23. *http://www.lidgetgreen.org/index.php/raising-the-validity-bar/testing-listening-comprehension/the-listening-comprehension-construct*

24. *http://www.merckmanuals.com/professional/clinical_pharmacology/concepts_in_pharmacotherapy/drug_interactions.html*

25. *http://www.nature.com/ajh/journal/v25/n8/full/ajh201251a.html*

26. *http://www.ncbi.nlm.nih.gov/books/NBK2216/*

27. *http://www.omic.com/resources/risk_man/forms.cfm*

28. *http://www.patient.co.uk/health/Coal-Tar-for-Psoriasis.htm*

29. *http://www.rxrecruiters.com/rx-recruiters-pharmacy-employers-job-board.htm*

30. *http://www.tesol-taiwan.org/index.php?title =Language_testing*

31. *http://www.usmlemaster.com/board/gobyPDS/upLoadFiles/Harvard%C6%ED%C1%FD%C7%D1%20%C6%FB.pdf)*

32. *http://www.writinghelp-central.com/letter-of-application.html*

33. Hughes, a (1989) *Testing for Language Teachers*. Cambridge and New York. Cambridge University Press.

34. Minsky, M. (1975). A Framework for Representing Knowledge. In: Winston, P. H. *The Phylosophy of Computer Vision.* New York: McGraw H

35. Minsky, M. (1985). The Society of Mind. New York. NY: Simon & Schuster.

36. David Nunan (2003) Practical English Language Teaching. McGraw-Hill Companies,Incorporated

37. John W. Oller (1979) Language tests at school: a pragmatic approach *Applied linguistics and language study.* Longman.

38. Rebek-Nagy, G., *et.al.* (2012). Manual for sTANDEM test developers and item writers. Holdacska Ltd.

39. Rummelhart, D. (1980). Schemata: the building blocks of cognition. In: Guthrie, J.T. (Ed). (1981). *Comprehension and Teaching: Research Reviews*. USA: International Reading Association.

40. Schank, R. C. and Abelson, R. (1977). *Scripts, Plans, Goals, and Understanding*. Hillsdale: Erlbaum Associetes.

41. Swales, J. (1981). Aspects of Article Introductions. *Aston ESP Research Report No.1.* Language Studies Unit, University of Birmingham, Birmingham, Anglia.

42. Swales, J. (1985). A genre-based approach to language across the curriculum. Paper presented at the RELC Seminar on Language Across the Curriculum at SEAMEO Regional Language Centre, Singapore. Tickoo, M. L. (szerk.) (1986). *Language Across the Curriculum*. Singapore, SEAMEO Regional Language Centre.

43. Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings.* Cambridge: Cambridge University Press.

44. Thomas, Sand Hawes, T. P. (1994). Reporting Verbs in Medical Journal Articles. *ESP Journal. Vol 13. No. 2. pp. 129-147.*

45. Widdowson, Henry G. (1997) "EIL, ESL, EFL: global Issues and local Interests". World Englishes 16/1. 146-53