Investigating English Teachers and Content Instructors' Tests in the ESP Exams at Medical Universities

[1]Iman Alizadeh*
[2]Fereidoon Vahdany
[3]Seyedeh Shiva Modallalkar

Abstract

This study aimed to investigate the test types which English teachers and content instructors employ in the ESP exams. To do so, samples of the tests developed by the ESP teachers were collected. Moreover, semi-structured interviews were performed with both groups to gain insights into their attitudes towards testing. The results indicated that word-formation, translation, definition, multiple-choice, reading comprehension, and cloze test were the test types the teachers used. It was revealed that except for cloze tests, which were used only by the English teachers, other test types were used by both groups; the frequency of the test types, however, differed in the exams. It was also discovered that both groups of the teachers used multiple-choice tests most frequently. The results of the interviews showed that the English teachers preferred integrative and communicative tests, whereas the content instructors tended to use syllabus-based structuralist tests. The study concludes that the ESP teachers mostly favored pre-scientific and psychometric-structuralist approaches to testing and did not use communicative tests. The implications of the study pertain to the pivotal role of the ESP teachers' awareness and evaluation of the students' real needs for English as well as the teachers' testing literacy in ESP courses.

Keywords: English for Specific Purposes, English Teachers, Content Instructors, Test Types

1. Introduction

Recently, the rapid growth of colleges and universities all around the world, the large number of educational programs and courses offered in English, and the growing number of students undertaking their studies in English have contributed to the creation and expansion of English for Specific Purposes (ESP) courses. Like any other educational program, teaching and testing are viewed as the main components of the ESP courses. As for the teaching of ESP, Dudley-Evans and John (1998) assert, "ESP requires methodologies that are specialized or unique" (p. 305). Arguing for the specialty of teaching methods in ESP, Donesch (2012) says such a specialty is because of the learners' needs, their target situation, and the language used in this situation. In the Iranian context, there is no principled or documented pattern of teaching in the ESP courses (Rajabi, Kiany, & Maftoon, 2012) and there has been no check for the efficacy and impact of the ESP teaching methodologies (Fakharzadeh & Eslami Rasekh, 2009). The ESP teaching methodology has been suffering from a principled standardized approach and, as Fakharzadeh and Eslami Rasekh (2009) note, fashions have adversely affected ESP teaching methodology in the country. ESP teaching methodology in the Iranian context is mostly determined by ESP instructors' knowledge, experience, and teaching environments (Hosseini Massum, 2011). The primary goal of teaching ESP in the Iranian context has been developing students' reading comprehension ability (Atai &

[1] Assistant professor (Corresponding Author), iman_alizadeh87@yahoo.com; Cardiovascular Diseases Research Center, Heshmat Hospital, School of Medicine, Guilan University of Medical Sciences, Rasht, Iran. School of Paramedical Sciences, Guilan University of Medical Sciences, Rasht, Iran.

[2] Associate professor in TEFL, frvahdany@gmail.com; English Language Department, Payame-Noor University, Rasht, Iran.

[3] MA student, modallalkars@gmail.com; Cardiovascular Diseases Research Center, Heshmat Hospital, School of Medicine, Guilan University of Medical Sciences, Rasht, Iran. English Language Department, Payame Noor University, Rasht, Iran.

Tahririan, 2003; Ghaemi & Sarlak, 2015; Khany & Tarlani-Aliabadi, 2016; Soodmand Afshar & Movassagh, 2016). Other researchers have expanded the spectrum of the goals of the ESP courses, adding other language skills like translation and writing (Amiri, 2000; Khoramshahi, 2015). Saffarzadeh (1981), for instance, maintains, "Developing translation knowledge is a must for ESP learners" (p. 3).

Language testing has witnessed pre-scientific, psychometric-structuralist, psycholinguistic-sociolinguistic, and communicative language testing eras (Fulcher, 2000; Spolsky, 1975). Accordingly, there are different approaches to language testing including, traditional, structural, integrative, and communicative. The prescientific era of language testing, which is linked with the grammar-translation approaches to language teaching, is characterized by grammar and translation tests. This approach to language testing tends to employ tests showing the students' educational abilities; this approach to testing is reported to be unable to cover all aspects of students' learning (Spolsky, 1975). Traditional testing methods mainly focus on receptive language skills and fail to check the students' productive language skills. Translation tests in the ESP course can be regarded as an example of the traditional testing; such tests are quick to administer and score. In the structural approach to testing, the language ability of learners is assessed by testing language elements separately. This approach holds that language includes four skills of listening, speaking, reading, and writing and that the skills themselves have components like structure, vocabulary, and phonology/ orthography (Madsen, 1983; McNamara, 2000; Weir, 2005). One of the test types which is frequently used in this approach is discrete-point multiple-choice questions. It maintains that a language test should sample all four skills and as many linguistic discrete points as possible (Brown, 2004). In contrast to the structural approach to language testing, the integrative approach views that language skills should be tested as a whole. In this regard, Oller (1979) stresses that a unified set of interacting abilities constitute language competence, adding that such competence cannot be tested separately. One of the test types, which follows the principles of integrative testing, is the cloze test. Communicative language testing, according to Moller (1981), assesses students' ability to use language to communicate, receive, and understand ideas and information. Communicative language tests aid teachers in testing "students' ability to use the language in realistic content-specific situations and tasks (Bakhsh, 2016). This type of test centers on learners' knowledge of the language and the way they use it (Bakhsh, 2016).

Concerning the significance of ESP assessment, Dudley-Evans and John (1998) argue that assessment occupies a prominent place in ESP courses, giving an ESP teacher information on the effectiveness and quality of learning. It is, therefore, essential that the tests be developed in such a way that they can measure abilities and knowledge relevant to students' current academic field of study or his or her future employment (Hafen, 2015). On the other hand, the assessment should be done in a way that provides opportunities for both learners and teachers to get the necessary information (Leung, 2007). The assessment methods in ESP courses mostly rely on formative and summative approaches in most universities (Maarouf, 2013). When teachers' classroom assessments become an integral part of the instructional process and a central part of their efforts to help students learn, the benefits of assessment for both students and teachers will be boundless (Guskey, 2003).

1.1. ESP in the Iranian Context

ESP programs were developed in Iran by the Iranian Ministry of Science, Research, and Technology in the form of specialized English programs (Eslami, 2010). At the Iranian medical universities, these programs are offered as a two or three-credit course at different education levels. In the explanation of the objectives of the ESP programs, it is said that they aim to prepare medical students to meet their English language needs in their future studies and at their workplace. In Iran, where the official language is Persian (Farsi), English is spoken as a foreign language. English is the language of the original texts medical sciences students use and is the language in which specialized language courses are offered at the universities. Students usually find the ESP course difficult to learn. In the Iranian medical education system, both instructors specializing in Teaching English as a Foreign Language (TEFL) and course specialists teach ESP courses at the universities

and assess students' achievement in the courses. According to Mostafaei Alaei, and Ershadi (2017), ESP teachers in Iran are TEFL teachers as the forerunner, subject specialists as the most popular ones by the learners, non-TEFL non-subject-specialists, or the rare group of ESP teachers professionally trained to teach ESP courses. They also divide Iranian ESP learners into two broad groups of university students majoring in a subject field other than the English language and vocational learners taking up in-service ESP courses.

The existing literature indicates that there is no study in the Iranian context investigating the test types ESP teachers use to assess the students' achievement in the ESP courses at medical universities. Moreover, the differences between the testing techniques and methods content instructors and English teachers use at the medical universities, and their attitudes towards ESP testing have remained unexplored. Discovering the ESP teachers' test types and attitudes towards testing will enrich the literature on ESP assessment and shed light on whether the ESP teachers' testing methods can meet the ESP goals or not. Therefore, the present study aimed to investigate the test types employed by English teachers and content teachers in the ESP courses at the medical universities in Guilan Province. It also aimed to discover the ESP teachers' attitudes towards testing. The objectives of the present study urged the researchers to formulate the following research questions:

Research Question One: What test types do Iranian English teachers and content instructors use in their exams in the medical ESP courses?

Research Question Two: Is there any significant difference in the frequency of the test types used by Iranian English teachers and content instructors?

Research Question Three: What testing approaches do Iranian English teachers and content instructors follow in their assessments in the ESP courses?

2. Literature Review

The body of literature in the area of ESP shows that many studies in the Iranian context have targeted ESP teaching (Atai, Babaii, & Nili-Ahmadabadi, 2018; Hayati, 2008; Sherkatolabbasi & Mahdavi-Zafarghandi, 2012). The studies which have been in the area of ESP testing and assessment have investigated the assessment of learners' achievement in the ESP courses (Ajideh, 2012; Alibakhshi, Kiani, & Akbari, 2010), the authenticity of the tests (Alibakhshi, Kiani, & Akbari, 2010), the differences between ESP and EGP testing (Ajideh, 2012), the students' perceptions of ESP tests (Moattarian & Tahririan, 2014) and ESP testing practices by English teachers and content teachers (Latif & Shafipoor, 2013; Nezakatgoo & Behzadpoor, 2017; Taherkhani, 2019). These studies are reviewed below.

Nezakatgoo and Behzadpoor (2017) investigated the main challenges of teaching ESP in medical universities of Iran from the perspective of ESP stakeholders from two Iranian medical sciences universities. They classified the challenges into institution challenges, learner related challenges, and teacher-related challenges. The issue of ESP testing was among institution related and teacher-related challenges. The students in the study believed that the ESP tests they took were a medical knowledge test in English which focused only on the students' translation ability. The teachers participating in the study also believed that they did not take any course on language testing and mainly followed what other teachers did.

Taherkhani (2019) investigated the cognitions and practices of language teachers and content teachers at medical sciences universities in Iran. Concerning testing and assessment in the ESP courses, the findings of the study discouraged content teachers from teaching ESP courses as they were unfamiliar with principles of testing and the assessment performed by the content teachers was not valid. Along the same line of research, Latif and Shafipoor (2013) studied ESP final exam tests for the students of accounting at Islamic Azad University, Iran. The results showed that the selected tests in the study were neither standard nor designed under a validated standard criterion. They concluded that the tests were "developed independently and based on individual test developer's (instructor's) personal tastes". They argued that the problem in the testing methods "proves the

Chabahar Maritime University
Iranian Journal of English for Academic Purposes   ISSN: 2476-3187
IJEAP, 2020, 9(4), 112-128   (Previously Published under the Title: Maritime English Journal)

inconsistency of the syllabi applied in the ESP classes". They further added that the focus of the tests was on translation skills, decontextualized morphological knowledge, and a few grammatical points.

Alibakhshi, Kiani, and Akbari (2010) also investigated the authenticity characteristics of ESP/ English for Academic Purpose (EAP) tests administered as a part of the MA/MSc or Ph.D. entrance examination at Iranian universities using a mixed-method design study. To collect data, they used a semi-structured interview and a questionnaire. They also compared ESP test tasks with characteristics of target language use situations. The results showed that ESP/EAP tests administered at Iranian universities were not authentic and that a change in the purpose of ESP tests was needed. Ajideh (2012) researched the relationship between EGP and ESP tests in four medical fields of study in the Iranian context. The results indicated that there was no systematic relationship between the students' scores on EGP and ESP tests. The study concludes that the tests do not provide similar information in the medical field of studies.

Moattarian and Tahririan (2014) studied how assessment could be beneficial in identifying ESP learners' needs. They attempted to investigate the language needs of Iranian graduate students of tourism management based on their wants, lacks, and necessities. The participants in the study were graduate students, English instructors, subject-specific instructors, and experts in tourism management who completed a questionnaire and participated in semi-structured interviews. The study concludes that all language skills need to be emphasized in the ESP courses to satisfy the specific needs of tourism management graduate students and that having a good assessment method can help instructors reach the ESP goals.

The review of the related literature also shows that many researchers (Dhindsa, 2007; Linn & Miller, 2005; Herrera, 2007; Malone, 2013; Wiliam & Thompson, 2008) have examined the critical role of teachers in the assessment process; however, there is no systematic study investigating the test types content specialists and English teachers employ in the ESP courses. In the same line of argument, Alibakhshi, Ghand Ali, and Padiz (2011) report, "Although different approaches to materials development in ESP have been practiced in the world of ESP, no real innovation in testing methods has been made." Being informed about the relevant body of literature and the existing gaps in the field of ESP testing, this study set out with the aims of investigating the test types used by the English teachers and content teachers in the ESP courses and discovering the teachers' attitudes towards ESP testing.

3. Methodology

The present study followed a mixed-methods study design to answer the research questions, which addressed Iranian content teachers and English teachers' test types and attitudes toward testing in the medical ESP courses. The selection of a research design is based on the nature of the research problem or the issues being addressed. Mixed methods research is a methodology for conducting research that involves collecting, analyzing, and integrating quantitative (e.g., surveys) and qualitative (e.g., interviews) research. Quantitative data include close-ended information collected by instruments such as questionnaires or checklists. Qualitative data consists of open-ended information that the researcher usually gathers through interviews, focus groups, and observations (Creswell, 2013). The present study used a checklist and interviews as data collection instruments. The content of the sample tests by the ESP teachers was analyzed using the checklist and interviews were conducted to gain in-depth insights into the teachers' attitudes towards testing.

3.1. Participants

There were two groups of participants in the study: the raters and the ESP teachers. There were two raters (a male and a female) who analyzed the content of the ESP tests. They were language teachers and had a 10-year-experience in language teaching. Moreover, ten ESP teachers (5 English teachers and 5 content teachers) at Azad and State medical universities in Guilan Province participated in the study by providing the researchers with their original ESP tests. Census sampling, which is a method of statistical enumeration where all members of the population are

studied, was used in the study. As the number of professors teaching ESP courses at the universities was low, all professors teaching ESP courses at the universities were targeted. Both female and male teachers participated in the study, their age ranged from 34 to 50 and their teaching experience varied from five to 30 years. The teachers were teaching ESP courses in different fields of medicine.

3.2. Instruments

The instruments used in the present study were the final exam questions developed by ten ESP teachers (5 English teachers and 5 content teachers) at Azad and State medical universities in Guilan Province and a checklist. The final exam questions were developed by the ESP teachers themselves. As in the present study the researchers aimed to discover the test types the professors used in their real classes, the researchers asked the ESP teachers for their original tests. The original tests the ESP teachers used in their exams were the target of the present study; the tests could have been standardized or not. The checklist was made of four sections: instructor type, test type, frequency of the test type, and the testing approach. In the test type section, the type of the test in terms of whether it is a multiple-choice, word formation, true-false, matching, translation, and cloze passage was indicated. In the instructor type section, it was shown whether the test had been developed by a content instructor or an English teacher. The frequency of the test type in the sample tests of the teachers was also indicated. In the testing approach section, it was shown under which testing approaches of traditional, discrete-point, integrative, or communicative the test type falls. The checklist was developed after conducting a thorough review of the relevant literature, pooling up dominant themes and concepts in the field, consulting with key informants and experts in the field, reviewing and restructuring, piloting, and modifying. Five experts in the field verified the content validity of the checklist. To secure the reliability of the checklist, the level of agreement between the two raters' analyses was calculated using Cohen Kappa.

3.3. Procedure

The data were collected through a precise examination of the tests developed by the English teachers and content teachers during a semester. To do so, the tests given to the students at the end of the semester were collected from each of the two groups of the teachers. After collecting the test samples, the checklist was used to determine the type and frequency of the tests used by the English and content instructors in the ESP courses.

Semi-structured interviews were also performed with all ESP teachers to elicit their attitudes towards ESP testing. During the interviews, three main questions were asked from the participants: 1. *What language skills do you focus on when developing your ESP tests? 2. Based on what language testing constructs do you develop your ESP tests? , and 3. What test types do you use to develop your test items?.* The interviews normally started with some warm-up questions on the participants' experiences of developing tests and assessment of students' knowledge in the ESP courses. Moreover, probing questions like "what do you mean by saying…?", and "can you give an example?" were asked to delve further into the students' answers. In total, 10 face-to-face interviews were performed and each of them lasted between 10 to 15 minutes. The participants' responses were recorded and then transcribed by the researchers.

3.4. Data Analysis

In the data analysis phase, two English teachers checked and analyzed the sample tests using the checklist. To analyze the tests, the teachers first read and reviewed each test several times and after gaining a full understanding of the content of the tests, completed the checklist. The data collected using the checklist were analyzed using the Statistical Package for Social Sciences (SPSS) version 19. To analyze the data collected in the interviews, the conventional content analysis method which was based on Graneheim and Lundman's (2004) model was used. To do so, first, the entire text of each interview was taken as the unit of analysis. The text of the interviews was read several times to gain a clear understanding of the data. Meaning units were then identified and were subsequently condensed into codes.

4. Results

4.1. Inter-rater Reliability

Two raters analyzed the sample tests collected from the ESP teachers at the medical universities. Cohen Kappa was used to calculate the degree of agreement between the analyses of the two raters. The results of the test can be interpreted as follows: values ≤ 0 as indicating no agreement and 0.01– 0.20 as none to slight, 0.21–0.40 as fair, 0.41– 0.60 as moderate, 0.61–0.80 as substantial, and 0.81– 1.00 as almost perfect agreement (McHugh, 2012).

Table 1: Cohen Kappa Test to Check the Agreement of the Raters' Analyses

		Value	Asymptotic Standardized Error[a]	Approximate T[b]	Approximate Significance
Measure of Agreement	Kappa	.857	.138	3.873	.000
N of Valid Cases		10			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

The result of the inter-rater reliability analysis showed a Kappa = 0.857. This measure of agreement is statistically significant. Values of Kappa from 0.40 to 0.59 are moderate, from 0.60 to 0.79 are substantial, and 0.80 and above are outstanding (Landis & Koch, 1977). It can be concluded that there was a strong agreement between the two raters' analyses.

4.2. Research Questions One and Two

The primary goal of the study was to discover the test types which the content teachers and English teachers used in the assessment of the students' achievements in the ESP courses. The results indicated that word-formation, translation, definition, multiple-choice, reading comprehension, and cloze test were the test types the ESP teachers used in their exams (See appendix A for samples of the test types). The results are given below.

Table 2: The Test Types Used by English Teachers and Content Teachers

Test types	Content Instructors	English teachers	Test categories	Test Approach
Word-Formation	47	30	Discrete Point	Structuralist
Translation	18	33	Traditional (pre-scientific)	Pre-scientific
Definition	32	26	Discrete Point	Structuralist
Multiple-choice	81	51	Discrete Point	Structuralist
Reading	1	7	Integrative	Integrative
Cloze test	0	33	Integrative	Integrative
Total	179	180		

As the table shows, the English teachers used more integrative test types in their exams. On the other hand, the content instructors tended to use discrete point tests. The testing approaches the content teachers and the English teachers employed in their assessments are presented on the graph below.
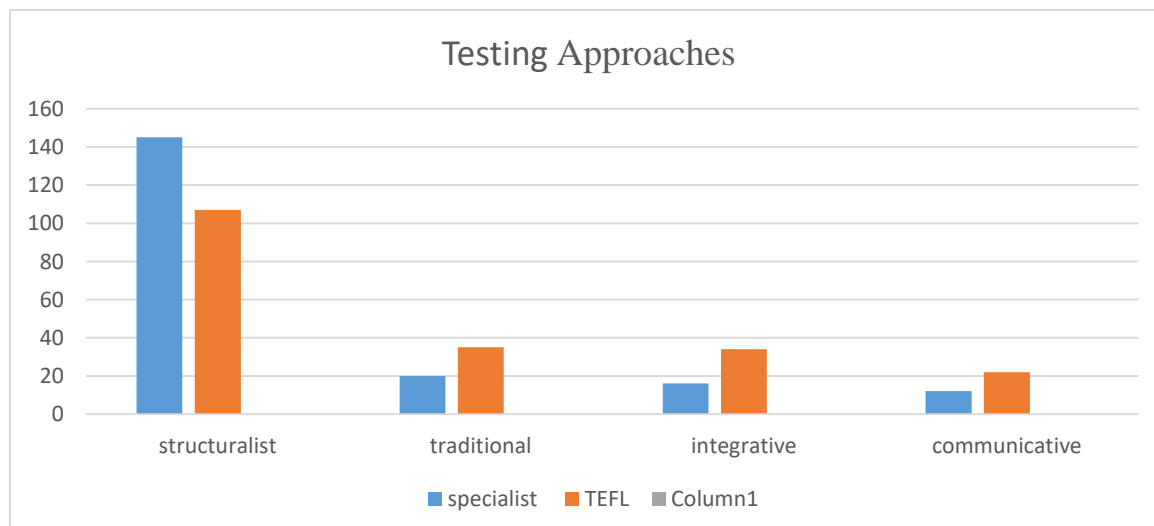
Figure 1: Testing Approaches by English Teachers and Content Teachers

As the figure shows, the content instructors employed the structuralist approach to testing more frequently. Translation, which is categorized as a traditional (pre-scientific) method of testing, was present in the tests of both the English teachers and the content teachers. As for the integrative tests, the results showed that the English teachers showed more preference for using these kinds of tests. To check whether the frequency of the test types differed in the exams of the English teachers and content teachers, the Chi-square test was run. The results of the test are given in Table 3 below.

Table 3: Differences in the test types used by the English and content teachers

Test type	Statistical test	Value	df	Asymp. Sig. (2-sided)
Word- formation	Pearson Chi-Square	4.667	4	.323
Translation	Pearson Chi-Square	6.000	5	.306
Definition	Pearson Chi-Square	4.667	5	.458
Multiple choice	Pearson Chi-Square	5.333	6	.502
Reading comprehension	Pearson Chi-Square	3.143	3	.370
Cloze passage	Pearson Chi-Square	10.000	3	.019

As can be understood from the table, the English teachers and content teachers differed in the use of cloze tests.

4.3. Results of the Interview Analysis

As was mentioned in the methods section, semi-structured interviews, which aimed to elicit the instructors' attitudes towards ESP testing, were also performed. All the participants in the study answered the three main questions of the interviews (1. *What language skills do you focus on when developing your ESP tests? 2. Based on what language testing constructs do you develop your ESP tests? , and 3. What test types do you use to develop your test items?).* The results of the analysis of the first question led to the discovery of such codes as "medical terms", "reading and translation tests", "four language skills", and "productive language skills". The same procedure was taken for all interviews. The analysis of the teachers' answers to the first question also showed that the two

groups of teachers had different views on the assessment of language skills. The content teachers asserted that focusing on medical terms should be one of the main components of the tests developed in the exams of the ESP courses. In this regard, one of the participants said:

Excerpt 1: "Medical students should acquire medical terminologies. This should be considered as an ultimate goal for them. Therefore, medical terms should always be one of the indispensable parts of ESP exams."

The content teachers also emphasized the role of reading and translation tests in the exams of the medical ESP courses. One of them said:

Excerpt 2: "Our students need to read many original texts in English and translate them from English into Persian. Therefore, reading and translation should be included in the ESP exams."

On the other hand, the English teachers believed that tests in the ESP courses should focus on all four language skills. One of the English teachers, for example, said:

Excerpt 3: "Teachers should test all four skills in their assessment procedures".

Emphasizing the role of productive language skills like speaking in the ESP courses, one of the English teachers said:

Excerpt 4: "Medical students need to be able to speak English. Therefore, their speaking skill should be part of the tests they are given".

Another question, which was asked during the interviews, was on the constructs based on which the ESP teachers developed their tests. The analysis of the interviews showed a sharp contrast between the perspectives of the two groups on the testing constructs and contents of the tests. The content instructors emphasized a syllabus-based testing construct. One of the instructors, for example, said:

Excerpt 5: "When developing tests in the ESP exams we should follow the educational objectives of ESP textbooks in the tests as well."

Another content specialist said:

Excerpt 6: "The ESP instructors should follow the syllabi designed and developed for the ESP textbooks and they try to use these the key points emphasized in syllabi in their exams".

The English teachers, however, insisted that in the ESP testing the instructors should distance themselves from traditional approaches to testing and develop test items with a focus on all language skills. One of the English teachers, as a case in point, said:

Excerpt 7: "In the tests of ESP course, the teachers should apply newer testing methods" that focus on "all four language skills."

Another English teacher said,

Excerpt 8: "There should be some changes in the ESP assessment procedures. Sticking to old and classic methods should be abandoned and recent testing trends like communicative ones should be sued as communicative tests are supposed to make students ready for real-life situations".

The English teachers also commented on the authenticity of the tests used in the exam. One of them, for instance, said:

Excerpt 9: "I think authentic materials are highly important for medical students. One of my purposes is to employ authentic materials for teaching and testing in the ESP courses."

Another question asked during the interviews was on the test types the ESP teachers used to develop the test items. Explaining about the test types they use in their exams, the content instructors mainly reported using multiple choice-items, true-false items, and translation questions. Some of them also reported a miscellany of test types in their exams. One of the content teachers, for instance, said:

Chabahar Maritime University
Iranian Journal of English for Academic Purposes   ISSN: 2476-3187
IJEAP, 2020, 9(4), 112-128   (Previously Published under the Title: Maritime English Journal)

Excerpt 10: "I normally develop multiple-choice items and true-false questions in my exams in the ESP course. Using multiple-choice and true-false tests are easy for both teachers and students".

As for using a miscellany of items in their exams, a content specialist said:

Excerpt 11: "I try to use a variety of test forms in her questions to get a better understanding of the students' learning".

Many content instructors also said that they used translation questions in their exams. One of them said:

Excerpt 12: "I ask students to translate a part of a medical textbook as class projects and consider it as part of their score."

Compared to the content instructors who assigned translations projects to the students, the English instructors stated that they used productive language tasks to check their performance in the ESP courses. One of the English teachers, for instance, said:

Excerpt 13: "I usually ask the students to have lectures and oral presentation as class projects and consider them when giving the students' final score".

One of the English teachers, however, emphasized that the mainstream test items and methods at the medical universities do not allow for testing the "students' communicative skills and competences". One of the English teachers said:

Excerpt 14: "The ESP instructors should use items which test the students' ability in using language skills interactively like cloze tests."

Another English teacher said:

Excerpt 15: "The current teaching and testing methods are not sufficient for testing students' language skills interactively and communicatively".

5. Discussion

This study aimed to discover the test types used by the content teachers and English teachers in the ESP exams at the medical universities in Guilan Province and to investigate the teachers' attitudes towards ESP testing. The results indicated differences between English teachers and content instructors in terms of the test types and attitudes toward ESP testing. The findings showed that word-formation, translation, definition, multiple-choice, reading, and cloze tests were the test types which both groups of the ESP teachers used in their exams. The frequency of the test types, however, differed in the exams of the ESP teachers. The findings of the study showed that neither of the two groups tested all language skills. It was also discovered that the English teachers' and content teachers' tests mainly centered on traditional and structural approaches to testing, English teachers tended to employ more integrative tests and neither of the two groups used communicative tests, which involves the use of performance tests of speaking, writing, and listening. It was also noticed that multiple-choice tests were the most frequent test type in the final exams of both English and content teachers. It was also found that the content teachers tended more to test the students' knowledge of medical terminology using discrete point tests.

Although the English teachers used integrative tests like cloze test and reading comprehension more frequently than the content teachers, like the content teachers, they did not employ communicative tests. The findings of the study correspond with the results of the study by Latif and Shafipoor (2013) who reported that ESP final exam tests were "developed independently and based on individual test developer's (instructor's) personal tastes" and the focus of the tests was on translation skills, decontextualized morphological knowledge tests, and a few structure tests. Along the same line of discussion, Neznakatgoo and Behzadpoor (2017), quoting the participants in their study, reported that "ESP tests evaluate the students' translation proficiency and the other language skills such as reading comprehension, writing, and speaking proficiency are not evaluated at all." In the same vein, Moattarian and Tahririan (2014) stated that all four language skills needed

Chabahar Maritime University
Iranian Journal of English for Academic Purposes   ISSN: 2476-3187
IJEAP, 2020, 9(4), 112-128   (Previously Published under the Title: Maritime English Journal)

to be emphasized in ESP courses to satisfy the specific needs of ESP students. Additionally, applying an appropriate testing method can aid instructors in achieving ESP goals. In our study, English instructors favored "applying newer testing methods" that focus on "all four language skills". Arguing for the necessity of employing new testing methods in the ESP courses, Alibakhshi, Ghand Ali, and Padiz (2011) stress that a "new approach to testing is needed to account for the recent development in ESP materials. And also, it is concluded that instead of traditional tests, the new approaches to tests especially portfolio assessment should be practiced." Discussing the benefits of using new assessment methods for students, Kostrytska and Shvets (2014) state that teachers' using of new strategies for assessment helps students take control of their success and accept responsibility for their learning.

One of the interesting findings of the study was that despite not employing speaking and listening tests in their final exams, unlike content teachers, the English teachers expressed in the interviews that they should use all language skills in the ESP courses. One of the English teachers, for instance, said, *"Teachers should test all four skills in their assessment procedures";* another one said, *"Medical students need to be able to speak English. Therefore, their speaking skill should be part of the tests they are given".* The reason for not employing such tests in their final exams despite believing that such tests are necessary for the ESP exams could be institutional or organizational issues. Neznakatgoo and Behzadpoor (2017), for example, argue that universities need to set the objectives of offering ESP courses, specify the language skills which the students need to learn and provide the ESP teachers with guidelines for assessing students' ability in the ESP course. They list unavailability of proper assessment guidelines and large classroom size as the institutional factors which negatively affect ESP courses at medical universities.

The results also showed that the English teachers tended to employ integrative tests, while the content instructors mostly favored discrete point tests. One of the reasons for such a difference could be the training the two groups have received in their academic studies and education. English teachers develop a good body of knowledge of testing and teaching in their field of study, whereas content teachers' scarcely study issues like teaching and testing in their specialties and education. Quoting one of the content teachers in their study, Neznakatgoo and Behzadpoor (2017) reported, "I know nothing about testing methods. I just follow the other ESP teachers particularly my teacher who taught ESP courses at the university from which I graduated". It seems that instructors' knowledge of testing English language skills have been overlooked in the ESP courses. Neznakatgoo and Behzadpoor (2017) questioned the ESP teachers testing skills, saying, "ESP teachers are not fully familiar with the principles of language testing. Almost all ESP tests consist of translation tasks."

As discussed before, the assessment of the four language skills of listening, speaking, reading, and writing has been emphasized in the literature, which requires the instructors' knowledge of testing these skills in different situations. Alibakhshi, Ghand Ali, and Padiz (2011) also maintain that "although different approaches to materials development in ESP have been practiced in the world of ESP, no real innovation in testing methods has been made." The findings also indicated that the English instructors favored communicative tests even though the textbooks and syllabus of the ESP courses lack activities and tasks requiring the development of communicative skills. The content instructors believed that they should test only what was in the textbooks and the test types used in the books; whereas English teachers preferred stepping beyond the limits of the textbooks and using test types assessing the students' real-world needs. One of the reasons for the ESP instructors' failure to use communicative tests in their exams could be the set purposes of ESP courses in the Iranian context. The main purpose of developing ESP courses in the Iranian context has been working on the students' reading comprehension and translation skills (Atai & Tahririan, 2003; Ghaemi & Sarlak, 2015; Khoramshahi, 2015). As developing communicative language skills has not been incorporated in the curricula of the ESP programs in the Iranian context as a goal, the ESP instructors accordingly pay less attention to them and do not develop and use such tests in their exams. Arguing for the effects assessment techniques can have in satisfying ESP learners' needs, Moattarian and Tahririan (2014) stress that all language skills need

to be emphasized in the ESP courses and that having a good assessment method can help instructors reach the ESP goals.

6. Conclusion and Implications

This study aimed to discover the test types employed by the English teachers and content teachers in the ESP exams at medical universities and investigate the ESP teachers' attitudes towards testing. The results showed that the teachers used word-formation, translation, definition, cloze test, multiple-choice, and reading comprehension test types. It was also discovered that both the content teachers and the English teachers used multiple-choice tests most frequently, the content teachers tended to use more word-formation and definition tests, and the English teachers mostly employed cloze tests and translation tests. Although the English teachers used integrative tests like cloze test and reading comprehension more frequently than the content teachers, like the content teachers, they did not use communicative tests. It was also revealed that neither of the two groups tested all language skills. From the findings, it can be said that the ESP teachers mostly favored pre-scientific and psychometric-structuralist approaches to testing (Spolsky, 1975) and did not use communicative tests (Fulcher, 2000).

The results of the interviews also indicated that the content teachers and English teachers had different attitudes towards testing; the English teachers tended to use integrative and communicative tests with a focus on the speaking skill, whereas the content teachers favored discrete point tests of medical terms. The content teachers also tended to test whatever was incorporated in the ESP textbooks; the English teachers, however, preferred to step beyond the limits of the ESP textbooks. The contrast between the ESP instructors' views on testing can be, as Latif and Shafipoor (2013) conclude, the fact that the individual test developer's tastes determine the type of tests they use. Moreover, the ESP instructors' choice of test types is influenced by their awareness and evaluation of students' needs for English as well as their testing literacy. Therefore, the study suggests that in-service training programs or workshops on assessment and testing techniques be organized for ESP teachers to sharpen their knowledge about testing ESP. Popham (2004) describes as "professional suicide" teachers' lack of appropriate training in assessment (p. 82). Popham (2009) also warns that the inadequacy of assessment knowledge may "cripple the quality of education" (p. 43). Likewise, López and Bernal (2009) note that tests have different uses for teachers equipped with assessment training and those without assessment training, stressing that the former use tests to improve teaching and learning, while the latter use tests to give grades.

Training courses on ESP assessment for ESP instructors can be beneficial especially for content instructors as assessment courses are not incorporated in the education programs of most of the medical fields at medical universities. The study also suggests that both content teachers and English teachers should use communicative tests in the ESP exams. The rationale for using communicative tests is that medical sciences students learn English to meet their practical language needs in their future studies and at their workplace. As communicative testing methods help teachers test "students' ability to use the language in realistic content-specific situations and tasks", their knowledge of the language and the way they use it (Bakhsh, 2016), it seems that such tests are well suited for ESP exams. This study had some limitations that can be addressed in future studies. First, the present study did not address the ESP teachers' assessment literacy systematically. Future studies can use available instruments to investigate English teachers' and content teachers' assessment literacy. Second, the present study did not target the effects of teacher variables on the teachers' tests in the ESP courses. Future studies can research such variables as teachers' gender, age, and teaching experience. Moreover, the present study focused only on the teachers' tests and did not investigate the process of test development by the content teachers and English teachers in the ESP courses. Future studies can target the test development measures taken by the content teachers and English teachers in the ESP courses.

References

Ajideh, P. (2012). EGP or ESP Test for Medical Fields of Study. *Journal of English Language Teaching and Learning, 3*(7), 19-37.

Alibakhshi, G., Ghand Ali, H., & Padiz, D. (2011). Teaching and Testing ESP at Iranian Universities: A Critical View. *Journal of Language Teaching and Research, 2*(6), 1346-1352.

Alibakhshi, G., Kiani, G., R., & Akbari, R. (2010). Authenticity in ESP/EAP Selection Tests Administered at Iranian Universities. *Asian ESP Journal 6* (2), 64-92.

Amiri, M. (2000). *A study on the English language programs at the B.A. level at Tehran universities* (Unpublished master's thesis). Allameh Tabatabai University, Tehran, Iran.

Atai, M., Babaii, E., & Nili-Ahmadabadi, M. (2018). A Critical Appraisal of University EAP programs in Iran: Revisiting the Status of EAP Textbooks and Instruction. *Language Horizons, 2*(1), 31-52.

Atai, M. R., & Tahririan, M. H. (2003). *Assessment of the status of ESP in the current Iranian higher educational system.* Proceedings of LSP: Communication, Culture, and Knowledge Conference. University of Surrey, Guilford, UK.

Bakhsh, S., A. (2016). Testing Communicative Language Skills of the Speaking Test in EFL Classrooms at King Abdulaziz University. *International Journal of Educational Investigations, 3*(6), 112-120.

Brown, D. (2004). *Language assessment Principles and Classroom Practices*. Francisco State University Press.

Cassady, J., & Gridley, B. E. (2005). The effects of online formative and summative assessment on test anxiety and performance. *Journal of Technology, Learning, and Assessment*, *4*(1), 390-421.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approach.* Sage Publications, Incorporated.

Dhindsa, H., Omar, K., & Waldrip, B. (2007). Upper Secondary Bruneian Science Students' Perceptions of Assessment. *International Journal of Science Education*, *29*(10), 1281-1280.

Donesch, E. (2012). English for specific purpose: what does it mean and why is it different from teaching general English? *The Journal of ESL Teachers and Learners, 1*(1), 9-14.

Dudley-Evans, T., & St John, M. (1998). *Developments in ESP a Multi-Disciplinary Approach* Cambridge: Cambridge University Press.

Eslami, Z. (2010). Teachers' voice vs. students' voice: a needs analysis approach to English for Academic Purposes (EAP) in Iran. *English Language Teaching*, *3*(1), 3-11.

Fakharzadeh, M., & Eslami Rasekh, A. (2009). Why's of pro-first language use arguments in ESP context. *English for Specific Purposes World, 8*(5), 1- 10.

Fulcher, G. (2000). The 'communicative' legacy in language testing. *System, 28*(4), 483–497.

Ghaemi, F., & Sarlak, H. (2015). A critical appraisal of ESP status in Iran. *IJLLALW, 9*(1), 262-276.

Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: Concepts, procedures, and measures to achieve trustworthiness. *Nurse Education Today*, *24*, 105-112.

Guskey T. R. (2003). How Classroom Assessments Improve Learning. [Electronic Version]. *Educational Leadership*, 6-11.

Hafen, C. A., Hamre, B. K., Allen, J. P., Bell, C. A., Gitomer, D. H., & Pianta, R. C. (2015). Teaching through interactions in secondary school classrooms: Revisiting the factor structure and practical application of the classroom assessment scoring system-secondary. *The Journal of Early Adolescence*, *35*(5), 651-680.

Hayati, A. M. (2008). Teaching English for Special Purposes in Iran: Problems and suggestions. *Arts and Humanities in Higher Education*, *7*(2), 149-164.

Chabahar Maritime University
Iranian Journal of English for Academic Purposes   ISSN: 2476-3187
IJEAP, 2020, 9(4), 112-128   (Previously Published under the Title: Maritime English Journal)

Herrera, S. Murry, K., & Cabral, R. (2007). Assessment accommodations for classroom teachers of culturally and linguistically diverse students. *English for Specific Purposes World*, *40*(1), 98-121.

Hosseini Massum, M. (2011). The Role of general background in the success of ESP courses: A case study in Iranian universities. *Literacy Information & Computer Education Journal* (LICEJ), *2*(3), 424-433.

Khany, R., & Tarlani-Aliabadi, H. (2016). Studying power relations in an academic setting: Teachers' and students' perceptions of EAP classes in Iran. *Journal of English for Academic Purposes*, *21*, 72-85.

Khoramshahi, E. (2015*). A needs analysis study on the curriculum of simultaneous interpretation major in an applied-scientific comprehensive university* (Unpublished master's thesis). Islamic Azad University, Saveh-Science and Research Branch, Saveh, Iran.

Kostrytska, M., & Shvets, S. (2014). Assessment strategies in the ESP course as a way of motivating student learning. *Assessment for Learning in Higher Education*, *6*(2), 59-71.

Landis & Koch, (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159-174.

Latif, F., & Shafipoor, M. (2013). Critico-analytic Study of ESP Final Exam Tests for Students of Accounting in Iranian Universities. *Theory and Practice in Language Studies*, *3*(10), 1790-1795.

Leung, K. (2007). Asian social psychology: Achievements, threats, and opportunities. *Asian Journal of Social Psychology*, *10* (1), 8-15.

Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Prentice-Hall.

López, A., & Bernal, R. (2009). Language testing in Colombia: a call for more teacher education and teacher training in language assessment. Profile: *Issues in Teachers' Professional Development*, *11*(2), 55-70.

Maarouf, N. (2013). *The Importance of Continuous Assessment in Improving ESP Students' Performance* (Unpublished thesis). Kasdi Merbah Ouargla University.

Madsen, H. S. (1983). *Techniques in Testing*. Oxford: Oxford University Press.

Malone, M. E. (2013). The essentials of assessment literacy: contrasts between testers and users. *Language Testing*, *30*(3), 329-344.

McHugh, M.L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276-82.

McNamara, T. (2000). *Oxford Introductions to Language Study: Language Testing*. (H. G. Widdowson, Ed.) Oxford, New York: Oxford University Press.

Moattarian, A., & Tahririan, M. (2014). Language Needs of Graduate Students and ESP Courses: The Case of Tourism Management in Iran. *RALs, 5*(2), 134-153.

Moller, A. (1981). Reaction to the Morrow paper. In J. C. Alderson & A. Hughes (Eds.), *Issues in Language testing*: ELT documents 111 (pp. 39-45). London: The British Council.

Mostafaei Alaei, M., & Ershadi, A. (2017). ESP Program in Iran: A Stakeholder-based Evaluation of the Program's Goal, Methodology, and Textbook. *Issues in Language Teaching*, 5(2), 306-279.

Nezakatgoo, B., & Behzadpoor, F. (2017). Challenges in Teaching ESP at Medical Universities of Iran from ESP Stakeholders' Perspectives. *Iranian Journal of Applied Language Studies*, *9*(2), 59-82.

Oller, J. M. (1984). Communication theory and testing: what and how. Paper presented at 1984 TOEFL Invitational Conference, Henry Chauncey Conference Center. Princeton, New Jersey.

Popham, W. J. (2004). Why assessment illiteracy is professional suicide. *Educational Leadership, 62*(1), 82-83.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, *48*, 4-11.

Rajabi, P., Kiany, G. R., & Maftoon, P. (2012). ESP in-service teacher training programs: Do they change Iranian teachers' beliefs, classroom practices, and students' achievements? *Ibérica, 24*, 261-282.

Saffarzadeh, T. (1981). *An introduction to the English books published by SAMT*. Tehran: SAMT Publications.

Sherkatolabbasi, M., & Mahdavi, A. (2012). Evaluation of ESP Teachers in Different Contexts of Iranian Universities. *International Journal of Applied Linguistics English Literature*, *1*(2), 198-205.

Soodmand Afshar, H., & Movassagh, H. (2016). EAP education in Iran: Where does the problem lie? Where are we heading? *Journal of English for Academic Purposes*, *22*, 132-151.

Spolsky, B. (1975). Language Testing: Art or Science? In Moller, Alan. 1981, In ELT Documents 111- *Issues in Language Testing*. London: The British Council.

Taherkhani, R. (2019). A Nationwide Study of Iranian Language Teachers' and Content Teachers' Cognitions and Practices of Collaborative EAP Teaching. *Iranian Journal of Language Teaching Research,* 7(2), 121-139.

Weir, C. J. (2005). *Language Testing and Validation: An Evidence-based Approach*. Palgrave Macmillan.

Wiliam, D., &Thompson, M. (2008). *Integrating assessment with learning: What will it take to make it work? The future of assessment: Shaping teaching and learning*. New York: Lawrence Erlbaum Associates.

Appendix A: Examples of ESP instructors' test types

The ESP instructors at the medical universities used different test types in their exams. The samples presented and discussed below are taken from the actual tests we could collect from the instructors. The instructors voluntarily provided the researchers with copies of their final exams. Examples of the test items in the exams of the instructors are given below.

1. Word formation

Word formation in the form of multiple choices, true-false and matching items, which was categorized as discrete point tests, was used by the ESP instructors. This type of test was used in the English for Specific Academic purposes and Medical Terminology courses. Both content teachers and English teachers used this type of test; the Content-specialist instructors, however, tended to use this form of tests more than the English instructors. Completing word-formation tests requires the recognition of suffixes, roots, and prefixes. Below, an example of this type of test is given. The students are asked to write the meaning of the prefixes and suffixes and give an example of the given suffixes or prefixes.

An example of a word-formation test developed by an English teacher in a medical terminology course:

Instructions: W*rite the meaning and example of the prefixes and suffixes in the table below.*

	Root	Meaning	Example
0	-g –therapy		
	Pyre-		
	-scope		
	Paleo-(Paleo-)		
	Xero -		

The instructors also utilized multiple-choice tests to measure the students' understanding of word-formation. An example of this type of test is given below:

An example of a word-formation test developed by a content instructor in an ESP course:

Instructions: Which is a compound word?

a) hysterectomy b) gastritis c) nephrosis d) cholecystitis

The instructors employed a variety of test types to check the students' knowledge of prefixes and suffixes. Making negatives was a type of test that the English instructors used in the final exam to check students' word-formation knowledge. An example of this type of test is given below:

An example of a word-formation test developed by an English teacher in an ESP course:

Instructions: Add a prefix to form the negative of the following words:

compatible

Infect

conscious

2. Translation

Translation was another test type that both the English teachers and the content specialists used. It was discovered that this type of test was used more by English teachers. The translations were from English into Persian and vice versa. It is worth mentioning that the English teachers presented longer texts in comparison to the content specialist. This kind of test was categorized as a traditional (pre-

126

scientific) test. Most of the content instructors used English-to-Persian forms of translation; while the English teachers used Persian-to- English translations as well. Examples of the translation tests are given below:

An example of a translation test developed by a content instructor in an ESP course:

Instructions: Please translate the following sentence into Persian.

Directional terms describe the position of structures relative to other structures or locations in the body.

Instructions: Please translate the following sentence into English.

سطح تاجی (سطح پیشانی) :یک صفحه ی عمودی است که از سمتی به سمت دیگر میرود وبدن یا هر یک از اجزا آن را به قسمت های قدامی و خلفی تقسیم مینماید

3. Definition

The definition of medical terms and matching items were also used by both the content instructors and the English teachers. This type of test was particularly used by the content specialist. These tests are categorized as discrete point tests (Brown,2004). Two examples of this type of test are given below.

An example of a definition (matching) test developed by a content teacher in an ESP course:

Instructions: Write the opposite of each of the following words:

   Sinistromanual _____

   Exogenous _____

An example of a definition (matching) test developed by an English teacher in an ESP course:

Instructions: Match the following terms and Write the appropriate letter to the left of each number

. aspiration a. decreased rate and depth of breathing

. hypopnea b. accidental inhalation of foreign material into the lungs

. apnea   c. a substance that reduces surface tension

. surfactant d. a measure of how easily the lungs expand

. compliance   e. cessation of breathing

4. Multiple-choice

Multiple-choice tests were also used by English teachers and content specialists. The content specialists, however, used it more frequently. Multiple-choice tests are classified as discrete point tests (Madsen, 1983; McNamara, 2000; Weir, 2005). These tests mainly targeted vocabulary and medical terms formation.

An example of a multiple-choice test developed by a content specialist in an ESP course:

Instructions: choose the correct answer.

Agriculture workers are often at risk of certain cancers related to………..

A. Chemical use   B. Agricultural machinery   C. prolonged sun exposure   D. A+C

An example of a multiple-choice test developed by an English teacher in an ESP course:

Instructions: choose the correct answer.

The plural of foramen is……...a) foramenices   b) foramenes   c) foraminata   d) foramina

5. Reading comprehension

Reading comprehension tests were also used by both English teachers and content instructors. This type of test was, however, more popular with the English instructors. They used both short and long texts in their samples. An Example of this type of test is given below.

An example of a reading comprehension test developed by an English teacher in an ESP course:

Instructions: Read the passage and answer the questions.

Just a few years ago, scientists did not know phytochemicals existed. But today they are the new frontier in cancer-prevention research. This pioneering science couldn`t have hit at a better time. People are more confused than ever about the link between diet and health: margarine is healthier than butter (or not): oat bran will save you (or won`t): a little alcohol will keep heart attacks at bay (but give you breast cancer). Just the effects of the popular vitamins known as antioxidants delivered a decidedly pessimistic massage. "We should have a moratorium on unsubstantiated health claims for antioxidants and cancer", says Dr. Julie Buring of Brigham and Women`s Hospital in Boston. Amid all the debate, phytochemicals offer the next great hope for a magic pill, one that would go beyond vitamins.

1- It can be inferred from the paragraph that <u>the pioneering science</u> (line 2) refers to the ……..

a) discovery of vitamins b) discovery of phytochemicals

c) Link between diet and health d) new knowledge about antioxidants

2-The writers state that the discovery has ……………..

a) Happened at the best possible time

b) Made people confident in previous understanding

c) Resulted in people`s trust in scientific findings

d) Been the most influential in health

6. Cloze tests

The last test type identified in the instructors' sample tests was cloze tests. This test type was used more by English teachers. Relatively, short texts were adopted for this purpose and both new vocabularies and grammatical points were among these tests. Cloze tests are considered integrative ones (Oller, 1984). An example of the cloze tests is given below.

Instructions: Fill in the blanks in the following passage with the words given. One word is extra.

Closure   recovery   Stable   debridement   healing   Principles

*   Methods of wound closure have advanced over the last decade with the addition of newer techniques but the fundamentals……1…….. of wound management remain unchanged and must be understood to achieve a ……2…….. wound that can be successfully closed. This article reviews the basic science of wound ……3…….. and the clinical principles of wound management concerning timing and assessment of the wound, and finally, the principles of irrigation and ………4….. that one must understand to obtain a stable wound before any soft tissue ……5…….. Once a stable wound has been achieved, some of the newer advancements in wound dressings and soft tissue coverage techniques that are discussed in this article can be applied to any open wound to achieve a stable durable wound closure that will ultimately lead to improved form and function.*