QUESTION BANKING FOR ENGLISH COMMUNICATION SKILLS COURSES: A GUIDE

Mirna Abd El Malak, PhD Candidate
Arab Open University, Lebanon

Abstract

With the advent in computers and the Internet, item banking emerged as an effective approach to testing. One of the many responsibilities currently burdening educators is regularly building tests. A question bank can help reduce the workload of constantly developing tests. A carefully constructed question bank renders test building efficient and the evaluation of learning more structured and criteria based. The significance of item banking stems from the increased inclination towards standardization in educational institutions. Question banking is a step towards developing standards for testing, getting reliable data and statistics to evaluate the educational process, and making the necessary educational decisions. For this purpose, it is essential that educators be conscious of the various options available before launching an item bank so that their choice be sound and based on the advantages, the disadvantages, and the implementation of each option. This study presents three main possibilities to launch an item bank: locally developing a software program and constructing items accordingly, leasing or buying item bank software and locally constructing items to supply it with, or buying a readily made item bank and the accompanying software. Clear guidelines and illustrations are provided to help educators construct well-built question banks.

Keywords: Question Bank, Item Bank, Testing, Assessment, English as a Foreign Language, EFL, English as a Second Language, ESL

Introduction

Test construction is a process that requires considerable time and effort, and building new tests on semestrial or annual basis is a very time consuming task (Squires, 2003; McCann & Stanley, 2009). Assessment in the learning process is an essential step in yielding tangible evidence to determine the value and level of learning. It is "a tool that leads to a continuous cycle of improved student learning" (Washtenaw Community College, 2013). Assessment yields qualitative information on how well

students are meeting the learning outcomes or comparable data that help evaluate achievement and provide more effective instruction (Washtenaw Community College, 2013; Weaver, 2013). Therefore, it is the basis of several educational decisions such as evaluating an educational programme, assessing students' competence, or shaping the next stage of learning. Various types of tests exist. For example, achievement tests measure the extent to which the learning outcomes have been achieved. Aptitude tests measure ability. Success on aptitude tests can also help predict future performance based on past successes. Diagnostic tests provide data on students' difficulties and gaps in knowledge. Knowing how the wrong answers were obtained helps the teacher in planning remedy (Izard, 2005). However, for these educational decisions to be sound, tests should be carefully constructed in order to test the students' abilities in the most efficient manner. Item banking is a novel and vital approach to test design and administration. It is a desired option made possible through advancements in computers and the Internet (Squires, 2003; Weiss, 2011). A well-built question bank supports the assessment of learning and renders test construction an efficient and simple task (Parchure, 2006). This study aims to present the different options available for tutors or course coordinators to establish a question bank for the course they teach or coordinate. The study also aims to illustrate with vivid examples how to adapt each option to construct an item bank for English courses, in particular. Finally, guidelines on question bank construction are provided with a brief note on interpreting test scores.

        With the aim of providing quality education to students, educational institutions currently manifest an increased inclination towards standardization and quality control. Essential to quality control is developing standards for testing. Relying on a test bank is required to generate exams in an efficient manner. Thus, instructors, in general, and those of English communication skills courses, in particular, need to be aware of the various options available for them to launch an item bank. The choice needs to be made in light of the advantages, the disadvantages, and the implementation of each option.

        According to the current trends towards language learning, the instructor is supposed to play a minimal role in presenting the material. Nevertheless, nowadays the EFL instructor is burdened with additional new roles to be assumed in order to create an effective learning environment that ensures student engagement and interaction, necessary for language learning to take place. One of the responsibilities that overwhelm educators is the regular construction of tests to assess student performance and keep an eye on their progress and development during a course. Consequently, a question bank helps reduce the workload of constantly developing tests. Also, it is a

step towards standardizing the practices and getting reliable metadata and statistics to evaluate educational practices, material, or student achievement, the springboard to appropriate educational decisions.

The Arab Open University (AOU) is currently interested and engaged in building question banks for the various courses on offer. Nevertheless, for this endeavor to thrive, it has to be grounded in researching the available options for question banking to make an optimal choice in light of the advantages versus the disadvantages and to set clear guidelines on the implementation of each.

Features of A Question Bank

A question bank is a set of carefully composed and jointly calibrated items that provide an operational definition of one variable through developing, defining, and quantifying a single common theme (Wright and Stone, 1999). It is a catalogue of questions classified according to their difficulty and the content they measure. Therefore, a question bank establishes a database of items serving as a source for generating a test with items of specified characteristics randomly selected from the bank. Security is ensured as each test sitting can be replenished with a different set of items (Squires, 2003). A question bank is characterized by the efficiency of feeding it with questions that can be easily searched for and retrieved. For questions to be retrieved easily and for various measurement purposes, the parameters of a question are to be listed. The parameters include metadata such as the revision number or usage data such as the performance, difficulty level, and discrimination power of the item defined (Parchure, 2006).

Question banking is a process consisting of clearly defined steps to launch and sustain it. Theories such as the Classical Test Theory or the Item Response Theory, known as the Latent Trait Theory, provide guidelines on establishing and maintaining a question bank. The Item Response Theory is concerned with designing, analyzing, and scoring of assessment instruments to measure variables such as abilities and attitudes. The theory is called Item Response due to its focus on the difficulty of each item, rather than the difficulty of the test, in general, and its treatment of the item difficulty as information in organizing items. The underlying assumption of the theory is that mathematical relationships between the item response and the ability/attitude constitute the mathematical function of a person and item parameters (Rudner, 2001; Baker, 2001). According to the Item Response Theory, Baker (2001) specifies three item parameters. The first is the discrimination parameter, called *a*. Parameter *a* indicates how well an item measures the students' ability. It's calculated by comparing a student's performance on the item to their performance on the whole test. Second is the difficulty parameter, referred to as *b*. This parameter designates the

difficulty of a question in relation to the student ability. Specifically, difficulty is inversely proportionate to ability. Finally, the guessing parameter, parameter c, is the probability of guessing the correct answer without any knowledge of the subject. Items with high c parameter are rejected (Parchure, 2006; FCAT, 2005). Another indicator of difficulty is the p-value, the proportion of students who answered an item correctly (FCAT, 2005).

Importance of A Question Bank

An item bank has significant advantages for test development (Rudner, 1998). As creating new tests regularly is time consuming, a question bank establishes a well-organized question database. Search parameters such as topic, keywords, and revision number allow easy access to questions (Parchure, 2006). As such a well-organized and carefully built bank is at the heart of constructing 'the best possible test for any assessment purpose' (Wright and Stone, 1999, p.109). For example, achievement, aptitude, or entrance exams can be produced from the same bank with minimal time and energy (Parchure, 2006).

A question bank allows for a wide variety of tests, and their results can be compared. It is possible to control the number of items and the level of difficulty to be included in the different test versions, which vary in relation to the assessment goals (Wright and Stone, 1999). Thus, monitoring the student performance is facilitated, and necessary action can be planned accordingly. Along the same lines, monitoring the item performance across varying tests is rendered possible (Parchure, 2006). Moreover, a question bank allows the construction of equivalent but different tests to be taken by different test takers on the same sitting or by the same person at different time intervals (Squires, 2003).

The Possibilities for An Item Bank

To start an item bank, there exist possibilities varying from a readymade item bank that can produce a multitude of test forms to programs that rely on algorithms for generation of items and widely varied tests to certain specifications. Different examiners have specific needs and concerns. For instance, users may be licensing bureaus managing and producing many exams every year or class teachers in need of a contained item bank. Individual course tutors would be more interested in programs that support item-writing and operate on an existing computer system. Three main options exist for a tutor or an educational institution (school, institute, or university) to establish an item bank for the course being taught (Ward, 1994). Educators may opt to locally develop a software program and construct items to replenish it with. It is possible, otherwise, to lease or buy

item bank software that can be fed with items that the course administrator constructs. The final option is buying a readily made item bank and the software needed for it.

1. Local development of software programs and items

       A drawback of locally developed software programs is that they require access to necessary advanced computer know-how, which is usually beyond that of many test developers. However, the advantage of this option is that it permits test developers to tailor the programs to the specific test features. The use of a single type of questions and a good word processor can reduce much of the disadvantages of this option. Computer skilled users can input one item per file via a word processor. Then, they prepare an item map and format tests through a data management program such as D-Base (Ward, 1994).

       The English Language Unit (ELU) is an offshoot of the Faculty of Language Studies mainly concerned with developing the students' proficiency level in English to desired standards which should leave students confident about their English proficiency. On one hand, the short-term goal of the ELU courses is to prepare AOU students for the requirements of their major courses. The ELU mainly aims to enable students to comfortably deal with their major courses and ensures that their English proficiency contributes to their academic success at the University. Specifically, by the end of the courses, students ought to be equipped with essential academic skills. First, they should have developed their reading comprehension skills to an extent that allows them to read and understand the books and assigned material at an acceptable rate or speed. Second, they should have developed the necessary fluency that enables them to write a coherent essay to express their thoughts. Moreover, the courses must help them to draw on appropriate diction and sentence structure that is comprehensible with minimal errors in language and mechanics. At the same time, students should be prepared to conduct small scale research. That is, they are expected to use their own language to paraphrase and summarize and hence avoid plagiarism, which is penalized by the University. Third, in class, students should be able to listen to and understand the tutors' explanation which is carried out in English. They should be able to take notes too. Finally, ELU students should have at least minimal oral fluency. They should be capable of asking questions, expressing their opinions orally, and discussing issues with the tutor and classmates.

       To help achieve these goals, the ELU offers two sets of courses. The first set consists of two courses which fall into the category of general requirements: English Communication Skills I and English Communication Skills II. Each of these courses is a 3 credit course and is offered on

semestrial basis. These courses have been part of the general requirements ever since AOU started admissions in Fall 2002-2003. The second set is the English Orientation Program (EOP) courses, or intensive courses, which are zero credit courses and are not part of the university requirements. As AOU adopts the open policy of education, no minimal proficiency level in English is set as a condition for admission. Instead, admitted students sit for a placement test which helps place them at an appropriate course level. The ELU courses are prerequisites for major courses.

At the ELU at the Lebanon Branch of AOU, a database for the Placement Test (PT) was created through the help of the programming staff and was put to use since July 2009. As a first step, the PT blue print was decided on. As the ELU courses emphasize the written skills, the PT is divided into four sections that reflect the written skills of reading, writing, vocabulary, and usage. The PT consists of 100 items divided into: 30 grammar items, 20 vocabulary items, 40 reading, and 10 writing items. Each section or category was further branched into subcategories that sample the content on the specific skill and different difficulty levels. For instance, under the category of 'Grammar', the topics that students would encounter across the ELU courses were sampled; thus reflecting the range of difficulty of the different levels of the ELU courses that are spiral in nature. The same applied to the 'Vocabulary' section that sampled the lexical words students are likely to have learned by the end of each course. As for the 'Reading' section, students would be presented with four reading passages. To standardize the difficulty level, it was decided that TOEFL type passages be used on the test. As a step towards further standardization, it was decided that only ten questions would follow each passage. Thus, some questions were added and others deleted. Finally, the passages that were compiled were subcategorized into four types according to difficulty: high, mid to high, mid to low, and low. The difficulty level of each level was determined by at least three raters who have sufficient experience in the field.

After the categories and subcategories of the test were decided on, the test blueprint or set up was created accordingly. A single type of questions, which is the multiple choice questions, formed the test items. Item developers were given sufficient guidelines for developing the question and the choices for each item to control for difficulty level and quality. For each item, the test takers are presented with four choices, three plausible and one correct. From each item at least 6 versions were created. Finally, all the questions were revised and checked to ensure clarity, simplicity in wording, and freedom from bias. Moreover, the choices were revisited to make sure that all are plausible but one is correct.

The test items were constructed to reflect certain specifications and were fed into an Excel sheet. Both the Excel sheet containing the question

bank and the exam set up were placed on a web application (.net). The results are the interface of the web application. The computer randomly chooses a version of each item every time a student sits for the test. As such, an infinite number of versions would be created by the computer, and there would be no chance that any two students taking the PT would receive the same version.

2. Purchase of a bank of items and the software to operate it

An advantage of purchasing a readymade item bank is that it is an easy option to avoid the responsibility and demanding work of developing items and the appropriate software. Usually these programs are simple to operate and the items are carefully constructed. However, it is essential that the items be checked against the objectives or blueprint of the test to be constructed. Moreover, there are limits that constraint customizing the test. For example, the bank may consist of one type of items (Ward, 1994). Many book publishers have developed item banks for achievement test to accompany certain books or an item bank for a diagnostic test to go with a specific series of books.

Oxford University Press has developed the Oxford Online Placement Test (OOPT) to place students of varying English proficiency levels in corresponding course levels. The OOPT has been used since July 2011 by the ELU to place newly admitted students to AOU in the appropriate ELU course. The OOPT and the accompanying series are governed by the same purpose and underlying philosophy to teaching English. As the series aims to develop the overall communicative proficiency of students, the test does not assess the students' academic skills. This could be a major drawback of the test when adopted by institutions to place students in courses intended to develop their academic skills. That is, the OOPT does not focus on the formal written academic skills. Instead, the items test the conversational language. The language is situational and spoken which contains idiomatic expressions and phrasal verbs rarely used in written language. The test consists of two parts. The first part tests usage through knowledge of grammar, of meaning of vocabulary words from grammatical form or from context upon reading diversified texts in different disciplines, and of the structural components of a sentence. The second part assesses the student's ability to listen to dialogue, to grasp the literal, intended, and implied meaning, and to subsequently answer questions. The student is allowed to listen to questions on this section twice (Oxford University Press, 2012).

An advantage of the OOPT is that it is a computer adaptive test. Computer adaptive tests are shorter with fewer items. More difficult items would follow a correct answer. This process continues until the test taker starts providing wrong answers. Thus, the ability would be determined

(Squire, 2003). On the OOPT, students select the initial level. Then, they are presented with around 45 items, approximately 30 testing language usage and 15 testing listening comprehension, to complete in 50 to 80 minutes, according to the individual student's speed in answering questions. The time allowed for the test is enough for all students, regardless of their level in English, to complete the test. As they go further in the test, the test engine intelligence component adapts the question levels based on the student performance. Accordingly, the grade weight allocated to each question changes (Oxford University Press, 2012).

3. Purchase or leasing of item banking software with local development of items

A point of strength of this option is that it allows utmost flexibility and minimal computer knowledge. Various item banking programs for use by university instructors and classroom teachers are available to be bought or leased. As such, considerable effort is needed to select a computer software package with characteristics that match the user's needs. It would be unwise to select a program with a multitude of options to be used to generate exams for a course. It is preferable that the program to be chosen to do the desired work be simple to operate and manage. Some institutions opt for a software package that is compatible with the existing hardware. Item entry and formatting are to be adapted to suit the capacity of the selected program.

One of the most widely spread software packages to establish student websites and an online course is Moodle, also known as Learning Management System (LMS), Virtual Learning Environment (VLE), or Course Management System (CMS). Moodle attempts to provide to educators the best tools to promote and manage learning. The social constructionist framework of education is the underlying educational philosophy that governs the design of this global development project (Moodle).

According to the social constructivist approach, learning; as manifested by the intellectual aptitude, cognitive strategies, motor skills, and dispositions people develop; is the result of the intentional interaction that takes place between people themselves or between them and the world around them. All the possible types of effective learning environments are required to support learners. Virtual spaces in distance education should mirror the real world by accounting for the social nature of learning. To foster deep learning, distance education environments are to ensure high levels of engagement and interaction among the users (Palloff & Pratt, 1999). Effective virtual worlds for learning create multiple means for the meaningful interaction between the creators (instructors) and consumers (students) of the knowledge that emerges (Bronack, Riedl, & Tashner, 2006).

The flexibility that characterizes the use of this educational tool is behind the high popularity of Moodle among educators around the world in creating online dynamic websites for students.

Moodle is an Open Source software. To work, it needs to be installed on a web server somewhere, either on one of your own computers or one at a web hosting company. It operates under the GNU General Public License, a free license for software packages. Basically, this means Moodle is copyrighted though it is free to download, and registration is voluntary. Another feature of Moodle is that it can be deployed to hundreds of thousands of students and can be used by different kinds of educators or educational institutions. It can be used as a platform for fully online courses or simply to support face-to-face tuition. The tools available such as the forums, databases, and wikis can be used to build rich collaborative communities of learning around a certain course, deliver content, and assess learning (Moodle).

Creating a question bank can be supported by Moodle. Question bank, a tool to create, edit, store, and import/export questions, allows for a pool of questions to be stored under different topics and subtopics to be later used to generate a quiz. Questions for a quiz can be randomly selected from the pool or directly added to the quiz. The questions can be either uploaded from a new file or imported/exported across the courses on Moodle. Questions under a category can also be managed by adding new questions, previewing, deleting, and changing the order of questions (Moodle).

To create a question bank, the course administrator has to access the *Question Bank* section from the *Settings* block by clicking on it. The four links to manage the *Question Bank* (*Questions, Categories, Import,* and *Export*) would expand under the *Settings* block.

The first step is to add categories by clicking the *Categories* link to open the *Edit categories* screen. This screen enables the user to add the categories of the test. It includes three f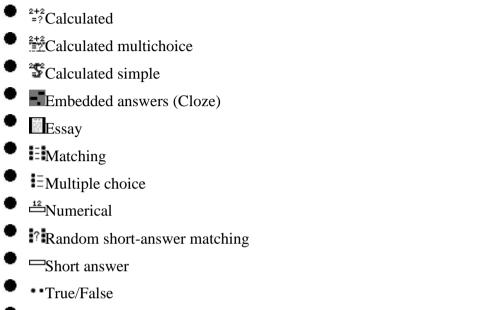ields: *Parent category, Name*, and *Category info*. It is essential that each category be given a *Name*. Then, it is advisable to describe each category by filling out the description in the *Category info* field. Each category can be listed under a *Parent category*. Once this information is filled out, the *Add category* icon has to be clicked to submit and save this information. Editing can always be possible by clicking on the category to be changed.

The second step is adding questions to the categories that were created for the desired test. To do so, the *Questions* link under the *Question Bank* in the *Settings* has to be clicked. The course administrator has to choose a category from the drop-down menu to add the question to. Then, click *Create a new question*. At this point, a secondary screen entitled

'Choose a question type to add' would appear revealing the following 12 different question types that can be possibly used.

Choose a question type to add
Select a question type to see its description.

- ² ₊ ²=? Calculated
- ²⁺² Calculated multichoice
- Calculated simple
- Embedded answers (Cloze)
- Essay
- Matching
- Multiple choice
- ¹² Numerical
- Random short-answer matching
- Short answer
- True/False

- DescriptionThis is not actually a question. Instead it is a way to add some instructions, rubric or other content to the activity. This is similar to the way that labels can be used to add content to the course page.

Fig. 1: Steps for Creating a Question Bank on Moodle

Settings

Question Bank

Step 1: to add categories   *Step 2: to add questions*

Categories

Questions

Edit categories(Parent category, Name, Category info)

Choose a category

Create a new question

Add category   *Choose a question type to add*

As the *Calculated, Calculated multichoice*, *Calculated simple,* and *Numerical* types of questions are more pertinent to mathematical tests than

English language tests; they will not be dealt with currently. However, a detailed description of how to construct the other types of items will be provided accompanied with a discussion of the options available for each. First, to add any question category, the *Question name*, *Question text*, and *Default mark* have to be typed in. In all these different types of questions, the students, after having submitted their answers, can be provided with feedback through an option called *General feedback* where information such as the answer and its justification would appear.

Also, for the *Embedded answers (Cloze), Multiple Choice,* and *Short answer* type of questions; each choice can be allocated a certain percentage of points. In addition, in the above mentioned types of questions and in the *Matching*, students can be allowed to attempt a question more than once, and the percentage of grade given can be manipulated through an option called *Penalty for each incorrect try* under a function called *Settings for multiple tries*. In the *Mutiple choice* and *Short answer* types, feedback on each choice that appears to the student when they have made the choice can be typed in. The development of the student answer along the tries indicates how helpful the feedback provided is. The options that enable the student to attempt the questions more than once and to receive feedback on the answers they provide make the quiz on Moodle a learning experience.

To create the *Embedded answers (Cloze)* or the *Random short-answer matching* types of questions, HTML coding is required. HTML allows the user to create all types of questions such as multiple choice and short answer. Fill-in-the blank type of questions fall under the category of *Embedded answers (Cloze)*. Also, it becomes possible to include for each choice or answer necessary feedback which will appear after the test taker has made his choice to provide a rationale of why that particular choice is wrong or right. Below are two examples on using HTML coding to create a multiple choice and a short answer question.

Example 1:
Match the following cities with the correct state:
* San Francisco: {1:MULTICHOICE:=California#OK~Arizona#Wrong}

Example 2:
The capital of France is {1:SHORTANSWER:%100%Paris#Congratulations!
~%50%Marseille#No, that is the second largest city in France (after
Paris).~*#Wrong answer. The capital of France is Paris, of course.}.

To create *Matching* questions, each question and its correct answer have to be typed in. As a setting for multiple tries, penalty for each incorrect try can be added. Moreover, two *Hints* can be added allowing to *Clear the incorrect responses* and/or *Show the number of correct responses*. The *Combined Feedback For any correct response, For any partially correct*

response, or *For any incorrect response* permits the test builder to add feedback that would appear to the test taker after making any incorrect response, any partially correct response, or any correct response in case there are more than one,. The students will see all the questions, and can choose the right answer from a drop-down menu including the answers to all the questions.

The *Multiple Choice* type would results in questions similar to the previous category where each question is followed by several answers from which one or more correct answer can be chosen. The course administrator allows one or multiple correct answers from a drop-down menu. Also, the administrator can click on a box to shuffle the choices so that the order of the answers would be randomly shuffled for each attempt. In addition, the choices can be numbered using lower case alphabet, Capital letters, Arabic numerals, Roman numerals, or no numbering at all. The number of choices provided can be controlled. As a setting for multiple tries, penalty for each incorrect try can be added. Moreover, the options of adding *Hints* as well as *Combined Feedback* discussed for the *Matching* questions are possible for the *Multiple Choice*. The students will see the *Multiple Choice* question followed by the choices and can click a radio button next to the right answer.

For *Short answer* type of questions, *Case sensitivity* can be enabled by choosing from a drop-down menu *'Yes, case must match'* or *'No, case is unimportant'*. Answer 1, 2, 3 etc.. can be added. This option is essential for this type of questions as variations for the answer might be possible and acceptable. Thus, each variation can be typed in, and the percentage of how correct it is considered can be assigned. As mentioned earlier, penalty for each incorrect try can be added. Moreover, the option of adding *Hints* as well as *Combined Feedback* discussed above also applies to *Short answer* questions. The students can fill in an answer of up to around 80 characters.

To add *Essay* questions, three parameters for the answer can be specified each from a drop-down menu: *Response format* to enhance the font and format of a text answer (HTML editor, HTML editor with file picker, Plain text, and Plain text monospaced font), *Input box size* (8 options ranging between 5 and 40 lines), and *Allow attachments* (No, 1, 2, 3, unlimited). Also, the answer key for any *Essay* question can be added to guide grading under *Information for graders*.

Finally, for the *True/False* type, the correct answer whether False or True can be chosen from a drop-down menu. *Feedback for the response 'True'* and *Feedback for the response 'False'* can be added. The students will see each statement followed by the two choices and can click a radio button next to the right answer.

Table 1: Summary of Features for the Different Question Types on Moodle

	Embedded answers (Cloze)	Essays	Matching	Multiple Choice	Random Short answer Matching	Short answer	True/ False
General Feedback *	X	X	X	X	X	X	X
Allocating a percentage of points per question	X			X		X	
Penalty for each incorrect try **	X		X	X		X	
Feedback ***	X		X	X	X	X	X
HTML coding required	X				X		
Answer key to guide graders		X					
One or multiple correct answers				X		X	
Hints/ Shuffling choices				X			

*Answers & justification
** A question can be attempted several times and the percentage of grade given can be manipulated
***Feedback can be added to appear to the test taker after choosing their answer

To create a quiz, the test builder has to go to Moodle and add it as an activity. Questions can be chosen manually by selecting a category of questions from a drop-down menu and adding the desired questions from each category. Otherwise, the quiz questions can be randomly added from the categories that are already created. The layout of the quiz can be managed in terms of the number and order in which the questions will appear. The dates of opening and closing the quiz to students should be specified. For all the question categories, students can be provided with options to *check* the answer and get feedback on each item before the final submission. Otherwise, the feedback can be deferred till the end of the quiz.

Once the students submit and finish the test, they receive feedback. Summary of attempts and of previous attempts are provided. Information such as the date of each attempt and the grade on each attempt appears. Also, students can review the feedback on each attempt. They can even reattempt the quiz. The number of attempts can be decided on by the test administrator from the settings of the quiz which can be chosen from a drop-down menu.

Guidelines on Building An Item Bank
Whether the English language instructor chooses to construct the question bank using readymade or locally-developed software, a test blueprint needs to be determined and test items need to be developed according to clear guidelines that ensure the quality of the test.

Quality Issues

The quality of the test results is crucial in drawing plausible conclusions and making appropriate decisions. Results would be trustworthy when the test is valid and reliable. The content validity of a test refers to the extent to which a test reflects the content of the curriculum (Izard, 2005). A test with high content validity closely matches the learning outcomes. A properly developed competency model specifying the skills needed for a task/question is the springboard for establishing validity. The competency model is considered the blueprint of a test (Squires, 2003). Reliability refers to the consistency of measurement (Izard, 2005). A reliable score is 'an accurate estimate of a student's true achievement' (FCAT, 2005, p. 58). An estimate bears errors, but when enough items are well written to reflect the content, are bias free, range in difficulty, and are positively correlated to success on the test; the test reliability will be high (FCAT, 2005).

Test Blueprint

'The blueprint is usually an outline or a hierarchical structure that delineates the structure and sub-domains of the primary domain, frequently with additional levels of specificity' (Weiss, 2011, pp. 4-5). The first step is to build the specifications of the question bank. That is to define the curriculum area and determine which items explicate it (Wright and Stone, 1999). Items/ questions that are not appropriate for the curriculum would render comparison of data insignificant and the conclusions drawn invalid (Izard, 2005). The scholastic variable is to be detailed enough to specify the organization of items by difficulty level along one main line of scholastic growth (Wright and Stone, 1999). Items are developed not only for content coverage but also for difficulty level. Items testing a particular topic and with a specified difficulty level form a group. Each group consists of 6 to 10 items. When there are many difficulty levels, the various versions would be more comparable in difficulty. Also, items should be written to reflect the development of the learner from being a novice to becoming a master of the skill represented by the topic. On the other hand, if all the items on a test are of a comparable difficulty level, the test will have a low discriminatory power as it would identify around two groups of achievers whereas a test with items ranging in difficulty levels would yield different groups of achievers with useful information on each (FCAT, 2005). Experts are aware of the skills, knowledge, and mistakes that mark each stage (Squires, 2003).

It is quite impossible to test all the material taught. The achievements of interest are to be selected, and each intended achievement is to be tested by questions or items that require students to perform tasks (Izard, 2005). Items written for a topic should represent all the possible items that could be written for a topic. The competency model guides item writers into producing a comprehensive and accurate sample of items for a particular

topic (Squires, 2003). The individual questions are not reliable indicators of success. Nevertheless, a combination of items into a test helps in establishing a pattern of success, which is a more reliable indicator as it is based on multiple sources of evidence collected by the individual items. Each question/ item should be independent in terms of the information required to answer it (Izard, 2005).

The material to be assessed, the assessment strategy, and reporting methods are decided on by the curriculum objectives, the importance associated to the different course sections, and the audience in need of the assessment data. Selecting representative tasks is not simple. It would be desirable if item/question writing is not idiosyncratic, representing a single person's limited view of the topic to be assessed, but rather the result of interaction among colleagues. In fact, topics selected for assessment are considered important by examiners and are highlighted in education. Thus, classroom instruction is significantly affected by assessment (Izard, 2005).

Content analysis and test blueprints provide a summary of the content to be covered by the curriculum including the significant sections and sub-divisions. A test blueprint specifies the test rather than the curriculum coverage. It includes details such as the test title, purpose, time, place, administrator, graders, target students, whether students can consult material, balance of questions, aspects of curriculum covered by the test, types of tasks, the use to be made of the data collected by the test, and the criterion level (passing score) set (Izard, 2005).

Item writing

Writing test questions with specific features demands considerable expertise and knowledge about the curriculum and how students learn. Items have to be developed according to specifications previously set. To maintain content validity, item writes are provided with clear instruction to access specific benchmarks, and later these items are reviewed for their connection to these benchmarks (FCAT, 2005). Item writers should measure what is important rather than what is easy to measure so as not to distort the assessment process or convey inappropriate information about the curriculum (Izard, 2005).

Questions on a test have to range in difficulty so that the test results allow for discrimination among abilities. Easier questions/items provide evidence of students' learning while the difficult reveal the areas where distinguished students excel. A test that doesn't permit low performers to demonstrate learning would have what is called a 'floor' effect whereas a test with a 'ceiling' effect is one that doesn't permit high achievers to demonstrate excellence in performance (Izard, 2005). Test items of moderate difficulty are the ones that elicit rich information as they are neither too easy

to be answered correctly by all students nor too difficult to be answered incorrectly by all students (FCAT, 2005).

Well written items don't lend themselves to guessing, ambiguous meanings, or tricking the test takers. Instead, they are easily comprehended and measure the intended skills/abilities. Otherwise, incorrect answers would no longer signify lack of knowledge or competence at question, but may be the result of other factors such as a limited reading ability (Squires, 2003).

It is unwise on objective items to use two choice items as the possibility of gaining a high score randomly is higher. Instead, there must be three to five distracters so that the probability of getting a high score through guessing would be minimized (Izard, 2005). Also, all the distracters have to be plausible but only one correct answer exists. Otherwise, if distracters are easily eliminated, the probability of students guessing the right answer would be higher (FCAT, 2005).The options have to be arranged in some logical order or from easiest to most difficult. Patterns of correct answers have to be avoided (Izard, 2005).

Pilot Testing

Evaluating and updating the item bank is an essential measure to maintain the quality of the bank and the integrity of the testing system. Item banking allows for smooth maintenance and revisions without disrupting the testing process. Pilot items can be added to the bank without including them in the scoring. After several administrations, statistical analysis would indicate how well an item is doing (Squires, 2003). Pilot versions usually include already accepted items in addition to the ones being piloted for the first time. Linking new and old items is referred to as anchoring (Izard, 2005).

To be able to compare test results, it is essential that the questions used in one version/ occasion be comparable but not the same as those on other versions/occasions. One way to ensure that two questions are comparable in difficulty is to administer these questions to students of a comparable achievement level and the students' responses provide evidence of comparison or difficulty. Comparable questions will exhibit a similar range of difficulty, will yield similar performance by subgroups such as males or females, and will have similar discrimination patterns over the range of achievement. For this reason, it is important to pilot the test on a group of students similar to those the test is aimed for. Variables of the pilot group such as number, gender ratio, diversity of age/schooling levels, size of rooms, and availability of invigilators are all to be considered (Izard, 2005).

Therefore, an item should first appear on a test for piloting. The pilot gathers information about each item such as its difficulty and ability to discriminate among students' performance. Moreover, pilot tests help gather information about how clear an item is and how students react to it. If the

pilot indicates that it is statistically sound, it would be retained in the bank. Data gathered from administering the final version of the test provide feedback on the individual items/question and function as quality control checks (Izard, 2005; FCAT, 2005). Data include the date of each administration. In addition, for open ended question, the distribution of scores on the item and the reliability of the scoring are to be considered. As for multiple choice questions, the difficulty (b or p-value), percentage of selecting each option, and discrimination index are to be noted (Ward, 1994).

Committee Review

Finally, items should be reviewed by a panel to check if the items have a clear task, are expressed in the simplest possible language, are fair for assessment at this level, are worded appropriately for the educational level, don't have unintended clues to the right answer, have a single clear answer, are the type appropriate to the required information, do not offend, are not biased, and are representative of the behavior to be assessed (Izard, 2005; FCAT, 2005). As such, items that don't meet the above mentioned criteria should be deleted. For instance, items which confuse, offend, or do not engage the student fail to gather appropriate information. Questions/ items that have been used over and again should be retired from use as candidates would have prior knowledge of them and thus such questions would no longer be testing achievement or aptitude (Izard, 2005). Moreover, items that become obsolete or that prove to be poor should be replaced (Squires, 2003).

Statistical Analysis

Statistical analyses are usually generated after the field or pilot test and after the operational administration to describe the quality of fit to the model. It corroborates the quality of the separate items to ensure that performance on a specific item conforms to the expectations. Also, statistical analyses verify the validity of the test, in general. Maintaining consistent content and difficulty is at the heart of achieving comparable results from different sittings. The results of a test that does not sufficiently sample the set of benchmarks would not be accurate (FCAT, 2005).

Test Construction

After individual items are out together to construct a test, some issues need to be taken into consideration. First, reviewers ensure that the items and the test as a whole meet previously set design and psychometric criteria (FCAT, 2005). The final test should be consistent with the blueprint. Moreover, the easiest questions have to be presented first to encourage candidates to proceed and to make sure that weaker students provide adequate evidence of their achievement and skills. At the end, the duration of a test should be long enough for most students to attempt most items so that sufficient sample of performance would be gathered (Izard, 2005).

Scoring And Interpreting Test scores

Scoring of multiple choice items is consistent and rapid. Essay questions, on the other hand, require graders who are skilled in evaluating essays in the appropriate content area, are time consuming, and require consistency among markers, which is difficult to achieve (Izard, 2005). Committee members determine the competencies required for a student to answer a question and which scores match the different levels of performance. Scores should be an accurate reflection of students' abilities (FCAT, 2005).

Assessment data should be read and interpreted in the context in which they were collected. Variables such as the teaching, the duration of teaching, class attendance, the questions asked, and the expected answers constitute the educational context. Scores should point out how well the learning outcomes have been met (Izard, 2005).

Conclusion

Test banking emerged from the need for paper and pencil tests that first saw the light in the 1920s and governed the educational and other applications of testing throughout the majority of the 20th century. Question banking was a tedious process. It wasn't until the 1930s that the possibility of machine scanning answer sheets to calculate a score was accessible and only for large testing programs. After World War II, technology influenced several dimensions of life including testing. The introduction of the personal computer revolutionized test development, analysis, and delivery. Through this labor-saving device, word processing software could be developed and adapted for testing purposes (Weiss, 2011).

As the testing process evolved, several possibilities for constructing an item bank emerged. For an item bank for a language course or a program to be started, there exist three main options that have to be carefully weighed in light of several parameters before making a decision. Educators may develop, with the help of an IT technician, a software program for their course and feed it with items that they can locally build. Although this option permits the construction of tests with specific desired characteristics, it requires the effort and time to well- construct items as well as the advanced computer skills or resorting to a computer programmer. Another option is to buy a readymade item bank and the software needed for it. Going for this option, the educator is freed from the responsibility of developing test items or the required software programmes, but the objectives must be checked to ensure the suitability of the test to the user's objectives. Also, it is possible to lease or buy item bank software which they have to construct items for. This possibility provides the flexibility of using different question types and reduces the burden of developing computer software.

Different test administrators and educational institutions might find it wise to resort to each of these possibilities depending on their needs and circumstances. However, for the test results to be reliable and allow for valid conclusions to be drawn, a test blueprint has to be established first. Moreover, test items have to be built according to clear guidelines and specifications set by the course administrator so that the test would end up testing what it is intended to test. Test piloting and constant revisions are recommended to guarantee fairness of assessment. The current paper discussed each option and provided a guide and specific examples on constructing an item bank for   courses, especially of English Communications Skills, to make the testing process not only less labor-intensive process but also more efficient and accurate. Finally, it is suggested that future research focus on the effects of adopting a question bank on different educational variables such as the instructors, students, classroom teaching, or the educational process, in general.

Acknowledgement

References
Bronack, S., Riedl, R. & Tashner, J. (2006). Learning in the Zone: A social constructivist framework for distance education in a 3-dimensional virtual world. *Interactive Learning Environments, 14* (3), 219-232.
Izard, J. (2005). Overview of test construction. A module prepared for UNESCO International Institute for Educational Planning. Retrieved o March 18, 2013 from http://unesdoc.unesco.org/images/0021/002145/214553E.pdfhttp://www.iiep.unesco.org/fileadmin/user_upload/Cap_Dev_Training/Training_Materials/Quality/Qu_Mod6.pdf
Florida Department of Education (2005). *FCAT Handbook—A Resource for Educators*. Retrieved on March 18 from http://fcat.fldoe.org/handbk/complete.pdf
Baker, F. (2001). *The Basics of Item Response Theory.* ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD. Available online: http://echo.edres.org:8080/irt/baker/
MacCann, R. & Stanley, G. (2009). Item Banking with Embedded Standards. *Practical Assessment, Research & Evaluation, 14* (7), 1-8.
Moodle. Retrieved on March 20, 2013 from https://moodle.org/about/

Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., & Valesquez, R. (2004). Psychological testing on the internet: New problems, old issues. *American Psychologist, 59*, 150-162.

Oxford University Press (2012). Oxford Online Placement Test. Retrieved March 20, 2013 from http://www.oxfordenglishtesting.com/defaultmr.aspx?id=3048

Palloff, R., & Pratt, K. (1999). *Building learning communities in cyberspace*. San Franciso, CA:

Josey-Bass.

Parchure, M. (2006). Question Banking. A paper presented in *Seminar for Teachers on Demystifying e-learning.* Retrieved on March 18, 2013 from http://www.cdacmumbai.in/design/corporate_site/override/pdfdoc/Question_ Banking.pdf

Rudner, L. (2001). Item Response Theory. Retrieved on July 13, 2013 from http://echo.edres.org:8080/irt/

Rudner, L. (1998). Item banking. *Practical Assessment, Research & Evaluation*, 6 (4). Retrieved on March 20, 2013 from http://PAREonline.net/getvn.asp?v=6&n=4

Squire, P. (2003). An item Bank Approach to Testing. A paper prepared for Applied Skills & Knowledge, LLC. Retrieved on March 17, 2013 from http://unesdoc.unesco.org/images/0021/002145/214553E.pdf

Ward, A. (1994). Guidelines for the Development of Item Banks. *Educational Measurement: Issues and Practice*, 13 (1), 34-39.

Washtenaw Community College (2013). Assessment and Its Importance. Retrieved on August 15, 2013 from http://www4.wccnet.edu/departments/curriculum/assessment.php?levelone=i mportance

Weaver, B. (2013). The Importance of Assessment. Retrieved on August 15, 2013 from http://www.scholastic.com/teachers/article/importance-assessment

Weiss, D. J. (2011) Item Banking, Test Development, and Test Delivery. In Geisinger, K. F. *The APA Handbook on Testing and Assessment*. Washington DC: American Psychological Assossiation.

Wright, B. & Stone, M. (1999). *Measurement Essentials* (2nd ed). Delaware: Wide range Inc.