

# Specifications Framework for Tests in an Outcome-based Language Program

Hoang Hong Trang<sup>\*</sup>, Nguyen Thi Chi, Duong Thu Mai

*Faculty of English Language Teacher Education, VNU University of Languages and International Studies, Pham Van Dong, Cau Giay, Hanoi, Vietnam*

Received 12 August 2016

Revised 24 September 2016; Accepted 22 November 2016

**Abstract:** Driven by the transformation of the language curriculum in the light of the competence-based approach, assessment activities serve as a tool both to measure students' achievement and to inform their learning progress. As such, it is a requirement that those activities be aligned with targeted competence, or learning outcomes. With broad understanding of outcomes, tests might also be considered as an outcome-based assessment tool, the quality of which can only be assured by a so-called "outcome-based" test spec. This paper, hence, presents various understandings of 'learning outcomes', and how testing can be adjusted to fit in with outcome-based assessment. Accordingly, different models of test specifications are reviewed and critiqued, followed by the proposal of a test specification model that is likely to facilitate outcome-based educational system.

*Keywords:* Outcome-based, testing, specification, tests.

## 1. Introduction

Curriculum restructuring at undergraduate level in Vietnam National University in 2012 led to the development of all new language curricular in University of Languages and International Studies, which is a subordinate university. At that turning point, Faculty of English Language Teacher Education decided to opt for the bifurcation of its language curricular into two parallel branches: Academic and Social English, both following the competence-based or outcome-based approach to curriculum design. As a result, new syllabi were written and materials composed, for both Academic English and Social English from

Intermediate to Advanced level of English proficiency, by lecturers of English in the faculty.

During the process of course design, classroom teachers, now as course developers, have confronted with several theoretical and practical difficulties, two of which were how to understand "outcome" and "outcome-based language education" and how to realize them in course materials as well as future teaching and assessment activities.

As for assessment tasks, since such formative assessment tools as portfolios, or performance-based assessment activities like presentations, debates, and forums have been well-used but not without fault, and due to the fact that students have to take a language "exit" test at the end of their tertiary education to be qualified for graduation, traditional paper tests

---

<sup>\*</sup>Corresponding author. Tel.: 84-965242384  
Email: hongtrangsp@gmail.com

have also been selected to provide a more accurate and comprehensive picture of an individual student's language proficiency.

A question arising then to the course designers was how a test in the outcome-based language program might be different from a "traditional" test, i.e. the test that had been composed and delivered so far. Differences if exist must be well and clearly presented in test specifications as test specifications, or test specs, are the blueprint for teachers to write a good quality test.

Hence, this research was conducted to find out the structure or components of a specification for "outcome-based" tests and features of a test specification that make it more "outcome-based". In particular, this research aims at answering two questions:

1. What are components of a specification for an "outcome-based" test?

2. What are the feature(s) of the test spec that can make it "outcome-based"?

To answer these questions, we started by investigating the literature of "outcomes", "outcome-based assessment", and "test specification" design. Later, through content analysis and practical experience as classroom test writers, we have come up with an eclectic model of test specification, with features that we believe can facilitate outcome-based language program.

## 2. Theoretical background

### 2.1. Outcome-based assessment and testing

Outcome-based assessment derives its name from outcome-based education approach, the practice of which can be traced back to the 1960s and 1970s [1]. However, it was not until a long time later that the theory of outcome-based education was discussed by a great number of scholars such as [2-6], to name but a few.

Outcome-based education (OBE) is commonly understood as an educational system in which teaching, assessment and learning are based on the intended results. In other words, by the end of the educational process students are expected to be able to achieve the predetermined learning outcomes. Outcome-based assessment, hence, refers to the procedure in which learners are assessed against those outcomes.

As suggested by the term, learning outcomes are key to the proper understanding and application of OBE in general and outcome-based assessment in particular. Even this concept of "outcomes" varies considerably among language experts. Generally, there can be two approaches to view "outcomes". In its narrow sense, "*outcomes are actions/performances that embody and reflect learner competence in using content, information, ideas and tools successfully*" [6: 13]. Purser [5: 5] also affirms:

Learning outcomes are important for recognition [...] The principal question asked of the student or the graduate will therefore no longer be 'what you do to obtain your degree?' but rather 'what can you do now that you have obtained your degree?' This approach is of relevance to the labour market and is certainly more flexible when taking into account issues of lifelong learning, non-traditional learning, and other forms of non-formal educational experiences.

As such, "outcomes" refer to what learners "can do"; knowledge, skills, and attitudes are not outcomes themselves but contribute to the demonstration of competence or learning outcomes. Given this, alternative assessment method, rather than traditional methods like paper-and-pencil tests, would be preferable since they provide simulated conditions for learners to demonstrate their language abilities.

In contrast, outcomes could be more broadly defined to include not only performances but also knowledge, skills and

attitudes, as evidenced by learners' actions. This understanding of outcomes implies a more flexible approach to assessment. Outcome-based assessment, as a result, can even employ tests as long as it is possible to elicit desired knowledge and skills from test takers.

Outcome-based assessment is by no means a new assessment approach. It still utilizes good assessment practices that have been available in other educational systems. However, outcome-based assessment does possess certain unique features which distinguishes itself from other approaches to assessment [7]:

1. Outcome-based assessment appears to be **formative** rather than summative although it cannot be said that summative assessment does not exist in outcome-based education.

2. Outcome-based assessment is criterion-referenced, or more broadly, **standards-referenced**.

3. Outcome-based assessment must allow possibility of discriminating students across **different levels of achievement**, which should not be simply a pass-fail scale.

As incorporated in outcome-based assessment, a test, therefore, must comply with those above-mentioned principles, meaning that it has to be written with a clear set of expected, measurable outcomes in mind, which then allows differentiation among test-takers and a

test should be used to foster future learning instead of summarizing a learning process.

## 2.2. Popular test specification models

In order to produce a "good" test that is valid and reliable, test construction process plays the key role, in which a test specification (or test spec) is irreplaceable no matter how detailed it might be or which format it might adopt.

Test specifications "*are the design documents that show us how to construct a building, a machine, or a test*" [8: 127]. Put it another way, they detail the nature of the items and the reasons why they are used in the test. In this sense, specifications play a vital role in ensuring the clarity of test forms so that they can be duplicated across different test times [9: 8].

With regard to outcome-based education which operates along a set of predetermined outcomes, it is crucial that the link between assessment activities (including tests) and learning objectives be clearly shown, because without which, no inferences about students' level of achievement can be made. Therefore, test specifications seem to be even more important.

Fulcher [8] has summarized five types of specifications that may be included in a complete test spec:

Table 1. Types of specifications in a test

Specification type	Basic function
Item/task specifications (i.e. prompt attributes)	Provide description of the prompts or the items or tasks in the test (e.g. the type of input material, wording of the instructions, sample items, sample anti-items, etc.)
Evidence specification (i.e. response attributes)	Provides description of test takers' expected performance and scoring method
Test assembly specification	Provides details on how the whole test is developed (e.g. number of each item type, the target reliability and the minimum number of items needed to meet this target, etc.)
Presentation specification	Provides information on how the items and test material are presented to test takers (e.g. margin size, spacing, the place to put page numbers, etc.)
Delivery specification	Provides information on test administration, test security and timing (e.g. space between desks or computers in a test room, number of invigilators per number of test takers, what is (not) allowed to use during the test, etc.)

These five types of specifications can be realized in a real test specification under different labels (or components). Following are three different popular test specification formats for test writers, which have been put forward by notable language assessment experts.

The first format, proposed by Popham (1978, 1981), has gained much popularity among language specialists and educators for its simplicity and efficiency. This five-component spec includes:

- **general description:** description of the behaviour or skill to be assessed, the focus of assessment, the learning objective or goal taken from the syllabus, and any contextual or motivational constraints in the particular test setting
- **stimulus attributes:** (i.e. the prompt attribute) description of everything related to the test items or test tasks given to test takers, which makes clear the link between the tasks and the learning goals or objectives that they are trying to aim at.
- **response attributes:** description of examinees' expected response
- **sample item:** presents the actual look of an item/task
- **specification supplement:** provision of further details necessary for test developers to facilitate test construction work (e.g. a list of potential text types, a list of potential topics, etc.)

[9: 14]

The second model [10], the Bachman & Palmer one, has the following parts:

- **purpose:** an explicit statement of how the test item/task should be used.
- **definition of the construct:** a detailed description of the construct, or particular aspect of language ability, that is being tested. This includes the inferences that can be made from the test scores, which overlaps with the purpose of the test.

- **setting:** a listing of the characteristics - physical location, participants, and time of administration - for the setting in which the test will take place.
- **time allotment:** the amount of time allowed for completing a particular set of items or a task on the test.
- **instructions:** a listing of the language to be used in the directions to the test takers for the particular item/task.
- **characteristics of the input and expected response:** essentially a description of what will be presented to the test takers (i.e., prompt attributes) and what they will be expected to do with it (i.e., response attributes)
- **scoring method:** a description of how the test taker response will be evaluated.

The last spec format to be reviewed is developed by Alderson et al. [11], who advocate the variation in format and content of a test spec depending on which audience it is targeting at. According to these experts, the audience of test specs can be categorized into test writers, test validators, and test users. Within the scope of this paper, only the spec format for test writers is discussed below:

- **general statement of purpose:** states the purpose of the test, that is, to diagnose students' strengths and weaknesses, to place students into suitable classes, to measure students' achievement after a course of study, and so on
- **test battery:** lists the components of the overall test (e.g., reading, writing, listening, speaking) with the time required to complete each
- **time allowed:** gives the time provided for the individual component being covered by the spec (e.g., reading)
- **test focus:** provides information about the general levels of proficiency the test is meant to cover, along with a description of the particular subskills or knowledge areas to be tested (e.g., skimming, scanning, getting the gist)

- **source of texts:** indicates where appropriate text material for the test tasks can be located (e.g., academic books, journals, newspaper articles relating to academic topics)
- **test tasks:** specifies the range of tasks to be used (e.g., relating this section to the subskills given in the “test focus” section)
- **item types:** specifies the range of item types and number of test items (e.g., forty items, twelve per passage, including identifying appropriate headings, matching, labeling diagrams)
- **rubrics:** indicates the form and content of the instructions given to the test takers.

Practically, most of the components of these frameworks are realized in public test spec of major English tests (e.g. IELTS, TOEFL iBT, and Cambridge First Certificate Exam). Some components which are witnessed in one test, and not in the others encompass: “response attribute” (Popham, 1981) (as cited in [9]), “source of text” [11], “definition of construct” [10], and “instruction” [11, 10, (Popham, 1981)]. Based on public information of these tests [12-14], their realization of spec components is summarized in the table below:

Table 2. How components of different spec models are used in major standardised tests

Popham (1981)	Bachman & Palmer (1996)	Alderson et al. (1995)	IELTS	TOEFL iBT	FCE
General description	Purpose	General statement of purpose	X	X	X
	Time allotment	Test battery	X	X	X
		Time allowed	X	X	X
		Test focus	X	X	X
	Setting				
	Definition of construct				X
Stimulus attributes	Instructions	Rubrics			
		Source of texts	X		
	Characteristics of the input ...	Test tasks	X	X	X
		Item types	X	X	X
Response attributes	... and expected response				X
Sample items			X	X	X
	Scoring method		X	X	X
Specification supplement					

### 3. A recommended specification framework for “outcome-based” tests

While Popham’s (1978, 1981) (as cited in [9]) test specification format emphasizes the importance of sample items by considering them as a separate component in a test specification, the two other formats do not

make sample items so explicit. For Alderson and colleagues, sample items are more necessary for teachers and learners or test takers than test writers because candidates need such essential information to familiarise themselves with the test prior to taking it. Moreover, the way Popham termed the first part of the spec General Description appears to be rather broad

and ambiguous although this section also takes test objectives as its core, just like the first section of the other models. Besides, although Popham's model does not specify how the test will be scored, it does include Specification Supplement, which provides room for flexibility in a test spec, thus, undoubtedly benefit test writers by providing further necessary information about test development. Regarding the spec model proposed by Bachman & Palmer [10], almost all components are similar to Popham's (but under different names), except for the lack of Sample Items and the inclusion of Definition of the Construct. Bachman & Palmer really want to make clear the tested language ability as a way to validate the coming test tasks and to strengthen the link between test tasks and test objectives. As for the model by Alderson et al. [11], the fact that neither response attributes specification nor setting (i.e. delivery specification) nor scoring method exists might derive from the authors' viewpoint of tailoring the test specification format for different target users. The aforementioned format aims at test writers, hence, such sections like response attributes, setting or scoring method can be omitted.

Given the three models above with their strengths and weaknesses as well as their practical use in popular international language tests, an eclectic model that hopefully can tackle the requirements for a test in an outcome-based syllabus is formulated. Below are major components of the suggested spec together with an example. Such elements like title of the spec or spec version has already been trimmed.

- **Statement of purpose:** briefly describes the purpose of designing and using the test or the reason(s) why such a test is necessary, that is, for example, to check the progress of students (progress test), to evaluate what students have been able to achieve after the course (achievement test), to place students in suitable classes (placement test), and so on.

*Eg. This test is designed to measure students' achievement after fifteen weeks*

*learning and practising academic language and skills.*

- **Test objectives:** identifies the course objectives that the test is going to cover, that is, the tested knowledge, skills and abilities.

*Eg. Based on the course guide, the following listening sub-skills will be tested with varying degrees of significance:*

1. Realizing the purposes of different parts of a lecture	2 items
2. Realizing the relationships between parts of a lecture	2 items
3. Inferring a lecturer's opinion	3 items
4. Identifying specialized terms in a lecture	3 items
5. Following description of features / processes in a lecture	5 items
6. Following the main points made in a lecture	5 items
7. Identifying supporting detail in a lecture	10 items
<b>Total</b>	<b>30 items</b>

- **Test tasks:** overviews the possible types of tasks to be covered in the test in order to meet the test objectives, together with the number of questions and time allocation for each task. A list of possible task types for each test objective should be made in order to avoid test-oriented instruction.

*Eg.*

<i>Tested skills</i>	<i>Question/Task type</i>
<i>Can understand main idea of instructions</i>	<i>Gap-filling</i>
<i>Can identify details which are clearly stated</i>	<i>Matching</i>

- **Item specifications:** describes instructions, input materials, features of test items and sample items for each item type. Also this section should detail instructions on designing items that can differentiate different levels of students' achievement of test objectives.

For a language test, one way to design a test which can differentiate students' level is by incorporating both items at one-level lower and one-level higher than the target level of the test, besides the items at the target level. For tests including items at only the target level, the difficulty of test items can also be reflected in

the level of cognitive demands for student test-takers. To design questions with different cognitive demands, teachers can refer to the Bloom's taxonomy and/or the Structure of Observed Learning Outcomes (SOLO) for useful guidelines.

Part 1:

*Eg.*

- 6 instructions, guidelines, announcements
- Each instruction/ announcement includes around 50 to 90 words
- Instruction may also include 3 to 5 turns
- Speed: 120 - 190 words/minute
- Topic: varied on social themes - unit 1-8 (4A Course guide)

Level of input: C1

**Sample instruction and item:**

*PART 1*

*In this part, you will hear SIX short announcements or instructions. There is one question for each announcement or instruction. For each question, fill in the blank with NO MORE THAN 3 WORDS AND/OR A NUMBER.*

*Question 1. The flight VN701 to Lyon has been delayed due to \_\_\_\_\_.*

- **Response specifications:** this section is optional for an objective test with the selected response format whereas it is essential for a subjective test in which the students have to construct their own responses.

*Eg:*

- *Students write answers to Wh-questions using their words.*
- *The answers must be within a word limit (no more than 50 words).*
- *The answers must rely on some evidence extracted from the text.*
- *Accurate spelling and grammar are expected for a correct answer; however, quality of ideas should receive more weight.*

- **Test presentation:** specifies how to present the items and other input materials, for example, the margin size, the font type and size, spacing, and other formatting features.

*Eg. See "Scripts for test instructions" below*

- **Scoring method:** clarifies how to score objective item types or mark subjective responses.

*Eg. Each correct answer is awarded 1 pt. The whole test with 20 questions has the total mark of 20 pts, which is finally divided by two to convert to the 10-point grading scale.*

**Scripts for test instructions:**

[2] Section 1

Now turn to section one

Listen to ...

Then, answer Questions 1-10.

You will listen to the recording ONCE only.

First, you have some time to look at questions one to ten

[ Silence = 45 seconds (10 items) ]

Now listen carefully, and answer questions one to ten.

[ Recording 1 ]

**Test format:**

Font type: Times New Roman

Font size: 12

Spacing: Multiple - 1.12

Margin: Top 1.5 cm, Bottom 1.5 cm, Left: 2 cm, Right 1 cm, Header 0.5 cm, Footer 0.5 cm

- **Specification supplement:** this section provides further information that item writers may need in order to write an effective test, for example, how to identify levels of text difficulty (for reading passages), possible sources of texts, etc.

Eg.

*How to decide on the difficulty level of a recording:*

*Difficulty of a recording can be decided by the following factors:*

*- Speed of delivery: this can be calculated by dividing the number of words delivered by the length of the recording. For example: your recording lasts 5 minutes 20 seconds (or 320 seconds) and the script has 400 words. Then the delivery speed is 400:320, which equals 75 words per minute.*

The eight-component model presented above incorporates the most preferred features of the three models put forward by Popham (1981, 1994) (as cited in [9]), Bachman & Palmer [10] and Alderson et al. [11] respectively. The reason why there exists “the statement of purpose” section is that we want to clearly position the test in the course timeline to decide its general role and goal. This is

essential in outcome-based education as the goal of outcome-based assessment should be formative rather than summative. Moreover, with the identification of the general goal of the test, and later the course objectives that the test addresses, the test is more likely to be properly and closely aligned with outcome-based instruction. This also explains why “course objectives” are set as a separate component of our spec. More importantly, what distinguishes this spec model from the others of the same type is the content of “item specifications”. Besides common information about test items, this section is also supposed to do further by pointing out what and how test writers can write items that help reveal different levels of students’ achievement of course objectives. Instead of the usual pass-fail scales, more meaning about students’ language ability will be added to the scores of tests designed from this suggested spec.

Furthermore, instead of using the terms “prompt attributes” and “response attributes” as in Popham’s model, “item”, “response” and “specifications” have been utilized. In practice, classroom teachers in many cases are also test writers and the fact that test spec appears less



technical and more teacher-friendly will make it a less frightening experience for teachers when having to use it to develop a test. Also, “test presentation” is included to ensure consistency in format in case more than one person takes responsibility in test development process. “Scoring method” also exists for the purpose of clarity and convenience since test scores might need to be converted to match the grading system that is currently in use in each institution.

Additionally, “specification supplement” is added to facilitate teacher’s process of test design. This section is supposed to include anything that a teacher needs to know in order to develop the test, which has not been addressed in the previous sections.

Lastly, the most important feature that makes this test spec more “outcome-based” is the content of the item specifications, which should show test designers how to write items of different levels of difficulty. Consequentially, students, instead of receiving a “fail” or “pass” score, would know which level they are at and then possibly be shown (by their teachers or peers) what they should do in order to reach their target level of achievement.

#### 4. Conclusion

To sum up, paper-and-pencil tests can be clearly incorporated in continuous assessment highlighted in outcome-based education as long as the test purpose and contents bear all of its hallmarks, that says, formative and standards-referenced. In this sense, test specifications have a critical role to play in aligning tests with learning outcomes in outcome-based education. The three reviewed spec models have certain similarities and differences which are possibly beneficial in various contexts; however, tests designed in accordance with these models may not be explicitly viewed as “outcome-based” tests. Taking the attributes of outcome-based assessment, a more “outcome-based” test spec model has been proposed, components of which are developed upon the review and critique of

those available models as well as of different aspects in a number of favored international language tests. It is hoped that the proposed model facilitates test development process in an outcome-based English language program.

#### Acknowledgements

This research has been completed under the sponsorship of the University of Languages and International Studies (ULIS, VNU) in the project No N.15.07.

#### References

- [1] Kennedy, D., Hyland, Á. & Ryan, N. Writing and using learning outcomes: A practical guide. Retrieved on June, 26<sup>th</sup> 2016 from [http://www.tcd.ie/teaching-learning/academic-development/assets/pdf/Kennedy\\_Writing\\_and\\_Using\\_Learning\\_Outcomes.pdf](http://www.tcd.ie/teaching-learning/academic-development/assets/pdf/Kennedy_Writing_and_Using_Learning_Outcomes.pdf).
- [2] Harden, R. M. (2007). Outcome-based Education: The future is today. *Medical Teacher*, Vol. 29.
- [3] Kenney, N., Desmarais, S. (n.d.) A guide to developing and assessing learning outcomes at the University of Guelph.
- [4] Malan, S. P. T. (2000). The ‘new paradigm’ of outcomes-based education in perspective. *Tydskrif vir Gesinsekologie en Verbruikerswetenskap*, Vol. 28.
- [5] Pulser, L. (2002). Recognition in the European higher education area: An agenda for 2010. To be reported at the international seminar of recognition issues in the Bologna process, Lisboa, Fundação Calouste Gulbenkian, 11 – 12 April.
- [6] Spady, W. G. (1994). Outcome-based education: Critical issues and answers. U.S.A: American Association of School Administrators.
- [7] Killen, R. (2000). Standards-referenced assessment: Linking outcomes, assessment and reporting. Keynote address to be presented at the Annual Conference of the Association for the Study of Evaluation in Education in Southern Africa, Port Elizabeth, South Africa, 26-29 September.
- [8] Fulcher, G. (2010). *Practical Language Testing*. Great Britain: Hodder Education.
- [9] Davidson, F. & Lynch, B. K. (2002). *Testcraft: A teacher’s guide to writing and using language test specifications*. Canada: Yale University Press.

- [10] Bachman, L. F. & Palmer, A. S. (1996). Language testing in practice: Designing and Developing useful language tests. Hong Kong: OUP.
- [11] Alderson et al. (1995). Language test construction and evaluation. UK: CUP.
- [12] Information for candidates: introducing IELTS to test takers. Retrieved on June 30<sup>th</sup> 2016 from [www.IELTS.org](http://www.IELTS.org).
- [13] ETS. TOEFL iBT Research Insight: TOEFL iBT test framework and test development. Series 1, Vol. 1. Retrieved on July 1<sup>st</sup> 2016 from [https://www.ets.org/s/toefl/pdf/toefl\\_ibt\\_research\\_insight.pdf](https://www.ets.org/s/toefl/pdf/toefl_ibt_research_insight.pdf).
- [14] Cambridge English Language Assessment. (2015). Cambridge English: First handbook for teachers. Retrieved on July 1<sup>st</sup> 2016 from <http://www.cambridgeenglish.org/images/cambridge-english-first-for-schools-handbook-2015.pdf>.

## Bảng đặc tả kỹ thuật cho bài kiểm tra trong khóa học ngôn ngữ theo định hướng chuẩn đầu ra

Hoàng Hồng Trang, Nguyễn Thị Chi, Dương Thu Mai

*Khoa Sư phạm tiếng Anh, Trường Đại học Ngoại ngữ, ĐHQGHN,  
Phạm Văn Đồng, Cầu Giấy, Hà Nội, Việt Nam*

**Tóm tắt:** Cùng với việc chuyển đổi chương trình học ngôn ngữ theo định hướng chuẩn đầu ra, các hoạt động kiểm tra đánh giá đóng vai trò như một công cụ vừa để đo mức độ hoàn thành của người học, vừa để cung cấp thông tin về tiến bộ học tập của họ. Do đó, những hoạt động kiểm tra đánh giá này phải thống nhất với các mục tiêu đã được đề ra của khóa học. Nếu hiểu mục tiêu khóa học theo nghĩa rộng, thì bài kiểm tra cũng có thể được coi là một công cụ đánh giá dựa trên chuẩn đầu ra, và vì lẽ đó, chất lượng của nó chỉ có thể được đảm bảo thông qua một bảng đặc tả kỹ thuật, tạm gọi là “Bảng đặc tả kỹ thuật cho bài kiểm tra dựa trên chuẩn đầu ra”. Mục tiêu của bài viết này là trình bày những cách hiểu khác nhau về “kết quả học tập” hay “chuẩn đầu ra” và làm thế nào mà bài kiểm tra có thể được điều chỉnh cho phù hợp với đường hướng kiểm tra đánh giá dựa trên chuẩn đầu ra này. Từ đó, những mô hình khác nhau của Bảng đặc tả kỹ thuật cho bài kiểm tra đã được xem xét và phê bình, làm cơ sở để xây dựng một mô hình gợi ý cho Bảng đặc tả kỹ thuật của bài kiểm tra theo hướng dựa trên chuẩn đầu ra của khóa học.

*Từ khóa:* Chuẩn đầu ra, kiểm tra, bảng đặc tả kỹ thuật, kiểm tra đánh giá.