

Pilot 3: *St.Petersburg Tower, go-o-od evening, Lear Three Four Five Oh. . . is inbound uh visual three five right.*

ATC: *. . .Three Four Five, St.Petersburg Tower Three Five Right, good evening, clear to land, wind three...two zero at eight.*

Pilot 3: *Cleared to land three five right, ah Three Four Five, thank you.*

ATC: *Three Four Five turn left, taxi to the ramp, which one will you use?*

Pilot 3: *We're going to JetExec, Three Four Five to the ramp.*

ATC: *See you later. . .*

Pilot 3: *See ya*

Pilot 4: *St. Petersburg Tower this is ah One Five Two Delta Golf, we are ready for takeoff at three five right.*

ATC: *. . .Two Delta Golf right traffic, three five right cleared for takeoff.*

Pilot 4: *Cleared for trackout...ah...takeoff and right traffic, One Five Two Delta Golf. . .*

Source: [www.amfly.com/KPIE\\_tower.htm](http://www.amfly.com/KPIE_tower.htm)

3. Design a blueprint for an English test for air traffic controllers, based on your analysis in question 2. Follow the framework provided in Table 3.4 above.
4. The International Civil Aviation Organization has prescribed standards for proficiency in aviation English that all national aviation authorities must follow: [www.toefa.com/english/articles/ICAO\\_language\\_proficiency\\_rating\\_scale\\_TABLE.doc](http://www.toefa.com/english/articles/ICAO_language_proficiency_rating_scale_TABLE.doc). After consulting the standards, and perhaps adjusting the abilities to be measured in your blueprint, develop a set of guidelines for raters including information about how to distinguish the various levels for each rating category.
5. Download *Hot Potatoes* ([www.hotpot.uvic.ca](http://www.hotpot.uvic.ca)) and design a short vocabulary quiz, using *JQuiz*, based on your analysis of the air traffic control transcript above.

## 4

### Alternatives in assessment

#### 4.1 NORM-REFERENCED AND CRITERION-REFERENCED TESTS

A common application of language test scores is to use them as a basis for grouping learners for some purpose. Imagine you wish to divide your class of 40 into small vocabulary-learning work groups, and you'd like to have the people in each group to have roughly the same level of vocabulary knowledge. You might give a test based on the vocabulary they have been learning, the results of which are shown in Table 4.1, below.

You would like to have seven learners in each group, so you simply select the top seven scorers for Group 1, the next seven for Group 2, and so on, with the last group, made up of the lowest scorers, having five participants, which is fine since they are the ones requiring the most help with vocabulary. This is an example of a *norm-referenced* use of test scores: we compare test takers with each other and rank them according to their relative performance. The assumption is that the scores will be

Table 4.1  
Vocabulary test scores

Name	Score	Name	Score	Name	Score		
Chinwa	100	Lisa	87	Ekaterina	75	Diana	49
Karina	98	Grace	87	Erica	73	Tim	48
Elena	97	Ivon	87	Ruslan	71	Felicity	45
Meja	96	Kimberly	85	Jessica	67	Vida	42
Anne	95	Jinrong	83	Sarah	67	Xuan	42
Yoo Ree	94	Edna	83	Renee	63	Mercedes	41
Adolfo	92	Hye-won	78	Na	55	Katherine	39
James	91	Peifeng	78	Monica	52	Yang	39
Erik	91	Richmond	77	Eliisa	51	Hilary	37
Jing	91	Katia	76	Lei	51	Dan	31

spread out along a 'normal' bell-shaped curve, as shown in Figure 4.1 below, hence the term *norm-reference*.<sup>1</sup>

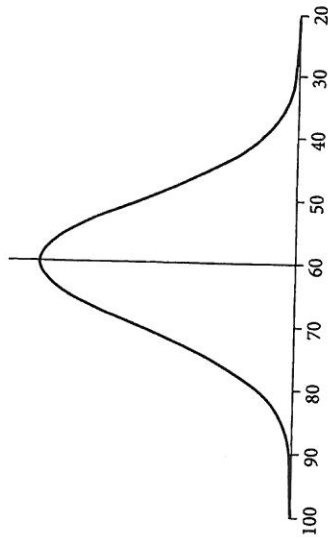


Figure 4.1 A normal curve

Notice that most of the test takers in Figure 4.1 scored between 70 and 80 on our hypothetical test, with correspondingly fewer scoring either very high or very low. Norm-referenced uses of test scores are most valuable when we need to make decisions about groups of learners based on an understanding of who has the 'most' of an ability, who has the next most, and so on. Often placement, selection and grading decisions are based on norm-referenced uses of test scores. Note that a goal of a norm-referenced test is to distinguish among the test takers in terms of the ability being assessed; therefore we try to design norm-referenced tests to measure a wide range of abilities so that the test takers will be spread out along the scoring scale, as shown in the example in Table 4.1 above.

Suppose, in a different case, that the educational district in which you teach had a statutory requirement that all learners of French know at least 90 per cent of the 1000 most common words in French by the end of the third year of study. For this purpose, you aren't really interested in who knows the most vocabulary, who knows the least, and so on, but rather you want to know whether all the learners in your class know *at least* 90 per cent of the most common 1000 words. To do this, you might extract a representative sample of 100 words from the larger list and test the students on those. You would hope that your class would all do very well on the test, scoring 90 per cent or above, as evidence that they all controlled the required vocabulary. The condition that all the test takers score at least 90 per cent on the test is thus the *criterion* in the term *criterion-referenced test*. The criterion will vary according to the test purpose: in a test to determine whether air traffic control trainees know the required terminology for responding to an emergency, for example, we might well want to ensure that they know *all* the obligatory emergency phraseology, not just

<sup>1</sup> In fact, the normal distribution of scores can only be assumed when we are testing a large number of people who represent a full range of whatever ability we wish to measure. We will discuss this in more detail in the next chapter.

most of it, and would thus set the criterion at 100 per cent. A criterion-referenced test is therefore designed not to measure a wide range of abilities, but to establish that test takers possess at least a *minimum* level of ability for some purpose, rather like setting a minimum standard for participation in the Olympics high jump: the bar is set at two metres, say, and aspirants must get over it in order to qualify for a trip to London in 2012.

It is possible, and in fact is often done, to use one and the same test for both norm-referenced and criterion-referenced purposes. We could use test scores to rank a group of applicants for admission to a programme of study that had 30 places for new students and simply take the top 30. Alternatively, we could decide to admit all applicants who scored above a certain level on the test, say 80 per cent, on the grounds that experience (or research) had indicated that those who possess at least that level of ability tend to do well in the programme while those scoring less tend not to succeed. Nevertheless, it is far better practice to use tests designed specifically for the purposes for which they are to be used so that test tasks can reflect either the range of knowledge assumed to be held by the target population, in the case of norm-referenced tests, or the minimum level of ability required for participation in a target language use situation, in the case of criterion-referenced tests.

#### 4.2. COMMUNICATIVE LANGUAGE TESTS

For the last 40 years now, one of the dominant themes in language testing has been *communicative testing*: assessing the ability to use language for communication in specific contexts, involving productive language either through meaningful input for the test taker to comprehend or interpret, or as meaningful output generated by the test taker. The communicative paradigm, as it is sometimes called, was developed in part in response to an earlier approach to language teaching and testing, the *structuralist* approach, which involved analysing the language into its component parts – phonemes, morphemes, syntactic forms, etc. – and assessing them separately, often without reference to context of use or communicative purpose. It was also a response to the Chomskyan paradigm of the 1970s which also focused on the rules for explaining grammatical sentences, ignoring performance or communicative issues. The philosophy behind the communicative approach is that even if a learner knows all the bits and pieces of a language – the sound system, the vocabulary, the grammar – she would still be incapable of communicating effectively. What the learner needs in addition to language knowledge is *communicative competence*, or the ability for language use (Hymes 1972), which involves judgement about what the grammar will allow one to say (e.g. how to apologise, complain, sympathise, congratulate, etc.) and about what is socially appropriate to say in a given situation (e.g. the use of title and family name as opposed to given name, etc.). Even given the relatively long history of communicative language testing, with its emphasis on eliciting language performance in relevant contexts of use, it is still not uncommon to find examples of tests that do not require contextualised production or comprehension and that test knowledge of isolated points of phonology or

syntax. As I will explain below, in the section on *discrete-point* and *integrative* tests, there may be a place for uncontextualised tests of grammar or vocabulary, depending upon the purpose of the assessment, but generally speaking, the communicative approach is the guiding principle for this book, and the framework of characteristics of language knowledge and strategic competence presented in Chapter 3 is based on a communicative paradigm.

#### 4.3 'GENERAL' AND 'SPECIFIC PURPOSE' LANGUAGE TESTS

For much of the history of language teaching and assessment, it has been common to distinguish between so-called *general* and *specific purpose* language courses and tests, the former referring to situations in which the purposes for learning cannot be specified with much certainty and the latter to those in which it can. Of course, it is true that no language course or test is developed for no purpose (in spite of humorous references to *TENOR*, for example – Teaching English for No Obvious Reason, cf. Abbott 1981). However, purposes range along a continuum from the somewhat vague, as in *French for cultural purposes*, for example, or *Spanish Conversation*, to the somewhat more definite *English for academic purposes* or *Business German*, or even more specific *Chinese for health workers*. In fact, it has been the case in recent years that the line between general purpose language tests and specific purpose language tests has become somewhat blurred. The procedure for test development outlined in Chapter 3 – defining the purpose of the test, conducting a preliminary investigation, collecting primary and secondary language data, analysing the target communicative tasks and language and developing test tasks that reflect the target tasks – should provide the basis for the development of a test at any point on the continuum of specificity, from rather general to highly specific. Moreover, as was discussed in Chapter 2, we use language test results to make inferences about language abilities with regard to some purpose, whether it might be making decisions about assigning grades in a language class, deciding whether students know enough Portuguese to participate in a Study Abroad programme in Rio de Janeiro, deciding in which level learners should be placed in a German course, or deciding whether air traffic controllers know standardised phraseology well enough to do their jobs. Though specific purpose language teaching and assessment will continue as distinct branches of applied linguistics, the old theoretical distinction between specific purpose language teaching as a 'training' activity, focused on providing a restricted linguistic code for a specific context, and general purpose language teaching as an 'educational' enterprise with a goal of providing learners with the ability to respond to a variety of unconfusable communicative situations (Widdowson 1983) is no longer viable.

#### 4.4 DISCRETE-POINT AND INTEGRATIVE TESTS

In reading and in talking with others about language testing, you may encounter references to *discrete-point* tests or *integrative* tests. *Discrete-point* refers to an

approach to testing which focuses on precise points of syntax, vocabulary, pronunciation, or morphology, often largely without reference to situation or purpose. For example, if we wish to test learners' knowledge of vocabulary we had just taught we might choose to do so by means of a multiple-choice test, as below:

Angry

- a. Alerta<sup>2</sup>
- b. Aburrido
- c. Jugueton
- d. Enojado
- e. Molesto

We might provide more context, to test verb form knowledge, as in the following example:

X: ¿Te gustó la comida?

- Y: Sí, \_\_\_\_\_ mucho.
- a. me gustó
  - b. te gusta
  - c. me gustan
  - d. me gustaron

Translation:

X: Did you like the food?

- Y: Yes \_\_\_\_\_ very much.
- a. I liked it
  - b. you like it
  - c. I like them
  - d. I liked them

This is still a discrete point task since it focuses on only one morphological point, and the context provided by the sentence doesn't really help the learner get to the correct answer: either she knows the correct form or she doesn't.

An *integrative* task, on the other hand, requires the learners to process a number of aspects of the language concurrently in order to arrive at a response, on the premise that the whole is more than the sum of the parts. For example, consider the task in Figure 4.2 below.

In order to respond correctly, readers must process information from the entire passage to understand that the main point is the procedure of making pemmican, although there is also information about who uses pemmican today, what can be put into it, and why it is still eaten today. Such a task requires *inferencing* ability as well as knowledge of syntax, morphology, and vocabulary. One could argue that this type

<sup>2</sup> The alternatives, in English, are *alert*, *bored*, *playful*, *angry* and *to be annoying*



Native Americans dried strips of meat, pounded it into a paste, and then mixed it with fat. Sometimes they added berries and sugar. Then they pressed it into small cakes. They called these cakes pemmican. Pemmican didn't spoil, and it provided lots of energy for people traveling or going hunting. Today explorers still carry and eat this food.

The passage mainly tells

- who uses pemmican today
- what can be put into pemmican
- how pemmican was prepared by Native Americans
- why people eat pemmican today

Source: Contributor Jeanine Matore, *Quintessential Instructional Archive* (2008)

Figure 4.2 Example integrative task

of task is in a sense easier to get correct for those who may possess a little less knowledge of specific points of grammar but who can combine several pieces of information from the text, as opposed to the previous two examples where knowledge of the answer was all or nothing.

Both approaches to language testing have their uses, depending upon test purpose, and both have their place in the tester's repertoire of techniques. It should be noted that the discrete-point/integrative distinction is not in fact an either/or dichotomy: tests can be placed on a continuum from clearly discrete-point to clearly integrative. It should be remembered, too, that the term *integrative* task, as used here, is distinct from, though related to, the term *integrated* task, as used in Chapter 3. There I was discussing the integration of input modalities, so that test takers had, for example, to read a passage and listen to a text on a similar topic, then write an essay summarising, comparing or contrasting the points from the two sources.

#### 4.5 FORMATIVE AND SUMMATIVE ASSESSMENT

Yet another distinction often made in language testing is that between *formative* and *summative* assessment. The purpose of formative assessment is to provide learners with information about their progress which they can use to guide their continuing learning, or to provide teachers with information which they can use to guide course development and lesson planning. Summative assessment, on the other hand, is usually carried out at the end of a course of study to measure achievement. The results of formative assessments are often provided to learners in the form of feedback on strengths and weaknesses in learning and information about what to do in order to make further progress or correct weak points. Summative assessment results are usually provided in the form of scores or grades, though increasingly, some type of corrective feedback may also be given. As is the case

with norm-referenced and discrete-point uses of tests, the same test may be used for both formative and summative purposes, depending upon the type and scope of feedback that is provided. However, as is usually the case as well, it is better to develop a test specifically for the purpose for which it is to be used: a test developed for formative assessment would no doubt reflect the methods and materials used for learning to a greater degree than would be necessary for a test developed for summative purposes.

#### 4.6 ALTERNATIVE APPROACHES TO ASSESSMENT

Since the 1990s there has been a certain amount of interest in what has been called 'alternative assessment', referring to a reaction to more 'traditional' assessments or tests (e.g. Herman, Aschbacher and Winters. 1992). Often in response to standardised testing imposed by administrators or outside authorities, proponents of alternative assessment argued for assessments that a) were extensions of normal classroom learning activities familiar to the learners, b) requiring learners to produce language in natural, real-world contexts, c) in a culturally sensitive way, d) scored or rated by human beings rather than computers, and e) focusing on higher level thinking skills and process rather than product. Today, nobody disagrees that these qualities are desirable, and in fact what were once alternatives to traditional assessment procedures are now seen as *alternatives in assessment* (Brown and Hudson 1998), part of the repertoire of techniques available for assessing the language abilities of learners, and subject to the same requirements discussed in Chapter 2: construct definition, authenticity, consistency, validity and evidence that the test results may be interpreted in the desired ways. In this section, I will discuss some of the alternatives in assessment, including the use of conferences, portfolios, peer-/self-assessment, task-based assessment, performance assessment, and 'dynamic assessment'.

Before beginning the discussion, however, recall that in Chapter 1 I noted a distinction between *evaluations*, *measurement*, and *tests* which is relevant here. Evaluation is the general term, and we may evaluate learners' abilities with or without tests or measurements. Measurement involves assigning a number or a place on a scale to a performance, while a test involves the elicitation of a language performance in a standardised way then assigning a number to it related to a standard scale. Some of the alternatives discussed in this section may be construed as tests if they are used in a more standardised fashion for the purpose of assigning grades, for example, as *measurements* if they are used in a formative assessment programme, or as *evaluations* focusing on informing individualised instruction. We might think of these terms as points on a continuum, with evaluations at one end and tests at the other, but it is probably not wise to get too involved in terminological distinctions since the important issue is maintaining fairness, consistency, and validity while engaging learners in tasks that are relevant to the purpose of the assessment.

#### 4.6.1 Conference assessments

Conference assessments usually involve one-on-one meetings between a learner and the instructor for the purpose of reviewing learning generally or to give feedback on a specific assignment and to discuss strategies for revision. The emphasis is on learning processes rather than products though a written or spoken product is often the focus of attention. Frequently conferences do not involve grading at all but are nevertheless useful in providing a basis for evaluating learner performance as well as assisting learners to develop self-awareness and critical skills. Over the course of an academic term, a series of conferences can provide a fairly clear picture of learner progress, perhaps focusing on successive drafts of written assignments, for example, or on discussions of reading assignments. The drawbacks to this form of assessment are that individual conferences are time consuming and can be rather subjective, making consistency across time and from individual to individual difficult to achieve.

#### 4.6.2 Portfolio assessment

Collections of learner work, whether samples of writing, audio or video recordings of speech samples or spoken interactions, or listening or reading logs and responses, have long been a part of the assessment of the learning process and of learning progress (Weigle 2002). Most typically we think of portfolios as involving writing assessment, since examples of students' essays and other written assignments lend themselves most readily to inclusion in a portfolio, but as indicated above, other forms of learner work can easily be included in both physical and electronic portfolios. Audio and video portfolios are particularly useful since they can represent records of learners' communicative interactions over time; blogs and email exchanges can also provide evidence of interactional progress. Portfolios can be used for a range of assessment purposes, including providing instructors with diagnostic information about what learners need to work on to inform class planning, helping learners become more self-critical and effective learners, providing evidence of progress to administrators and other stakeholders, and evaluating learner progress for grading purposes. Often portfolios are assembled by the learners themselves, with guidance from the instructor, according to guidelines established in consultation between the instructor and learner, or collectively by the instructors in a programme or by school administrators. Criteria for assessing portfolios may focus on both the collection as a whole (e.g. what it shows about the learner's self-reflection and metacognitive awareness and the scope and quality of the work included) and on individual entries (e.g. achievement of rhetorical/communicative purpose, completeness, and control of grammar, mechanics and style).

Three types of portfolios have been described, generally related to the assessment purpose (Herman et al. 1996):

- *Showcase* portfolios: examples of a learner's best work, usually for purposes of grading or external accountability. Often this work will be that produced later in the course rather than near the beginning.

- *Progress* portfolios: examples of successive drafts or other evidence of growth during the course, for the purpose of evaluating individual learner progress. Clearly, work included in a progress portfolio will span the duration of the course, representing different stages of learning and a variety of assignments and tasks.
- *Working* portfolios: examples of work illustrating mastery of learning goals or units of study, for purposes of programme and curricular assessment. Again, this type of portfolio might contain work spanning the entire course.

The essential components of portfolio assessment include Collection, Reflection, and Selection (Hamp-Lyons and Condon 2000). The goal of the portfolio is to provide evidence of a learner's writing ability that is rich in scope, so the collection of language samples must be more than simply a stack of papers or recordings. Often a reflective essay by the student will be included, commenting on the various examples of her language ability: why they were chosen and what each indicates about the learner's progress in acquiring and using the language. The examples might be selected to represent work conducted in the language class, assignments in other academic classes, or in situations outside the classroom altogether. As Brown and Hudson (1998) suggest, the portfolio can enhance the assessment process by involving both the instructor and the learner, giving the instructor a clearer picture of learner accomplishment by providing examples of authentically produced language, and expanding the range of work assessed over time.

#### 4.6.3 Self- and peer-assessments

Closely related to portfolio assessment in terms of the involvement of learners in reflecting on the goals of language learning are self-assessment and peer-assessment. One of the main benefits of the self- and peer-assessment process is providing learners with enhanced awareness of learning goals and criteria for judging the quality of their own learning. A possible outcome is the enhancement of self-motivation and learner autonomy. Such assessment can work in at least two ways: learners can evaluate their own products, such as written essays, audio samples of their speech, or videos of their interactions in class or in communicative situations outside the class, or learners can be provided with 'can-do' checklists for them to rate their abilities in various communicative scenarios. It is essential that learners be given practice in self-assessment, both before judging their own work and before judging that of their fellow learners. Learners can be given an example of the type of language product to be assessed and work in groups to decide what aspects should be assessed and what the criteria might be. Alternatively, they can be given a set of guidelines by the instructor and practise using them to evaluate language samples. As they become more familiar with the criteria provided, they can be encouraged to produce their own. An example of a simple self-assessment guide is shown in Figure 4.3 below.

There are drawbacks to both peer- and self-assessment. Accuracy varies with the complexity of the skill or performance being assessed. Rating vocabulary use

1. In the past few lessons I have studied/practiced/worked on:  
(Students fill in a relevant topic or skill area [communicative function, grammatical point, cultural aspect] relevant to their cases.)
2. How well do you think you deal with the topic you listed under question 1?  
(Students assess their performance or understanding using a scale ranging from *not at all* to *thoroughly*.)
3. How is the topic you listed under question 1 relevant to your own needs and goals?  
(Students respond by using a scale ranging from *not important* to *extremely important*.)
4. Questions 1, 2, and 3 are repeated for other skills or topics as appropriate.
5. Summarising the past few lessons, I feel that I have learned:  
(Students rate their learning on a scale ranging from *nothing at all* to *a lot*.)
6. Thinking about the answers to questions 3 and 5, I realize that I need to change my learning approach or priorities in the following way(s):
7. The areas I need to work on more are:
8. The areas in which I feel I'm doing well are:
9. In the next few lessons, I am interested in learning about:

Source: National Capital Language Resource Center (NCLRC). (2008) *The essentials of language teaching*. Retrieved 25 September 2009 from <http://nclrc.org/essentials/assessing/languagepop.htm>.

Figure 4.3 Example language use self assessment form

or verb tenses, for example, may be more amenable to peer-assessment than the evaluating of the rhetorical effectiveness of an oral presentation. Self-assessment in particular must be used with caution, especially in situations where learners have a vested interest in obtaining higher ratings. For example, when learners are asked to rate their own performances for purposes of a classroom discussion only, the assessments tend to be relatively low or even underestimated; however, when the assessment may result in a shortening of a course of study or exemption from study altogether, learners tend to rate themselves higher.

#### 4.6.4 Task-based and performance assessment

The motivation for task-based assessment or performance assessment<sup>3</sup> is similar to that for the other alternatives discussed – the assessment of language ability in

<sup>3</sup> The terms *task-based assessment* and *performance assessment* have somewhat different origins and histories in the testing field, but here they are used as synonyms.

settings that lend themselves to authenticity and the complexity that is normally found in language use outside the classroom. Task-based assessments range from the more traditional assigned essay or speaking task to more elaborate problem-solving or group discussion tasks and the complex integrated reading, listening, writing tasks discussed in Chapter 3. There is more to task-based assessment, though, than merely setting up realistic tasks or communicative situations and letting learners go to work on them. Designing complex tasks, evaluating performance on them, and making valid interpretations of the results require careful planning. Assessment tasks demand not simply instructions to 'do' a task, but also provision of a clear communicative goal, a reason for engaging in the task related to a purpose outside that of mere display of language skill. Determining and stating the nature of the language ability required to perform the task is also a challenge, as is developing a rating scale for the measurement of those components of ability.

There is also the problem of deciding whether successful task completion is an essential aspect of the performance or not. On the surface of it, this would seem to be a given – of course the learner would have to successfully accomplish the task or the point of the exercise would be lost. However, it may be the case that success in terms of completing a task is simply outside the learner's control. Consider a task involving giving instructions for connecting and operating a computer projector, for example. The learner might be able to produce a very clear set of instructions, including describing which cables should be attached to which input and output ports, and so on. The reasonable outcome of such a task might be that the computer will display a PowerPoint® show on the projection screen. However, suppose that the person being given the instructions is simply 'all thumbs' and cannot carry out the instructions; or that the projector bulb is faulty and will not work; or that the computer settings are not correct for that particular projector. Which of us has not observed such problems at a conference presentation, or worse, experienced them ourselves? The point is that performance tasks are indeed complex and outcomes are not so predictable as might be the case with more traditional test tasks. The criteria for judging the performance thus should perhaps focus more on the process of the performance rather than on a 'correct' outcome. This means that the rating criteria will need to be fairly explicit about the components of a successful performance, and this will require considerable trialling and pretesting. Training raters other than oneself is also a challenging aspect of task-based assessment and a clear set of guidelines will have to be produced.

Finally, while a particular performance task may reflect faithfully the features of a task of interest in the real world, and successful performance on the assessment task can be interpreted as an indication that the learner will be able to perform the same task outside the classroom, what about other tasks in the same real-world context? Just because the learner has demonstrated an ability to take a telephone message and relay the contents to someone else in an email, can we interpret that to mean that he will be able to do the reverse and convert an email text into a spoken message and relay Or that he will be able to interact with a live visitor to the office and relay the visitor's message to someone else in a handwritten memo? The generalisability of performance tasks is a fundamental problem with task-based assessment, and evidence must be provided that we can interpret performance on assessment tasks in the ways intended.



Below is an example of a task designed for English language learners in Hong Kong. According to the designers, the task is intended to assess the following abilities:

- to generate ideas relevant to the experiential content of the task (tourism);
- to demonstrate a degree of mastery over the 'tourist brochure' genre;

#### *Process writing task*

##### *Overview of task*

In groups, learners design a brochure for the Hong Kong Tourism Board describing 4 attractions in Hong Kong which would appeal to young people of their own age.

##### *Task guidelines for learners*

#### Writing a Tourist Brochure

Imagine the Hong Kong Tourism Board has asked your class to design a brochure that would be of interest to young visitors of your own age. In groups of 4, design the brochure describing FOUR sites suitable to young people of your age coming to Hong Kong. Complete this task by following the steps below:

##### *Step 1: Group Task.*

Discuss in your groups which sites young people would want to visit in Hong Kong. Choose one site each to investigate. For homework find out as much as you can about the site, where the site is, when it is open, what one can see/do there, what the facilities are, how one gets there, etc. Bring this information to the next class.

##### *Step 2: Group Task.*

Exchange information with your group members. Tell them about the site you have found out about. Then decide how you are going to present the information in your brochure, what order you want to put your sites in, what illustrations you need, what title you want to give the brochure, etc.

##### *Step 3: Individual Task.*

Write a description about your chosen site (120 words). Remember to say why it is interesting. Proofread it carefully, then hand it to your teacher.

##### *Step 4: Group Task.*

In your groups edit your work based on your teacher's comments. Then put together your brochure. Your brochures will be assessed on the following basis:

- Task fulfillment: would your selected sites appeal to young people?
- Accuracy of language and information provided: is the brochure written in good English? Is the information provided accurate?
- Attractiveness of final written submission: is your brochure really attractive? Can you make it more appealing?

- to draft ideas, and then to revise and polish initial written efforts incorporating feedback and suggestions from fellow learners and the teacher; and
- to clarify and develop ideas by making revisions to own written texts through personal reflection and talk with others.

The task allows for the assessment of each individual writing performance as well as the group effort at producing the brochure.

#### 4.6.5 Dynamic assessment

A relatively recent alternative in language assessment, known as *Dynamic Assessment*, is based on a Vygotskian concept that it is not sufficient to assess what a learner can do by him/herself at the present moment as a result of past learning, but rather is necessary to assess what the learner will be capable of doing in the future, as a result of mediation by a teacher/assessor. The approach intentionally blurs any distinction between teaching and assessing, viewing the two activities as sides of the same coin (Poehner and Lantolf 2005). One method<sup>4</sup> of dynamic assessment consists, minimally, of an initial assessment, an intervention or *mediation*, followed by a second assessment to obtain a *difference score*, reflecting the change resulting from the mediation. For example, researchers in Israel engaged in a dynamic assessment procedure to measure the *learning potential* of a group of young adults learning to read English as a foreign language (Kozulin and Garb 2002). The learners were given a standard test of reading comprehension in which they had not only to answer the multiple choice questions but also indicate the clues in the passage and the test question that helped them arrive at each response. The mediation focused on helping the learners identify the knowledge required to answer each question and on strategies for finding clues in the texts. The post-test was based on different reading passages but required similar types of information and strategies to answer questions that were of the same length and level of difficulty as those in the pre-test. The results suggested that some learners did indeed benefit from the intervention and were able to apply the strategies acquired to the new reading passages, while other learners were less amenable to the intervention and scored the same or lower on the post-test. The researchers devised a formula based on the difference between the pre- and post-test scores to calculate a *learning potential score* for each learner. This information would enable teachers to provide instruction directed at the particular needs of groups of learners: those with average pre-test scores but relatively low learning potential could be given instruction in reading strategies, for example, while those with average pre-test scores and higher learning potential could be given more independent reading activities with more challenging material<sup>5</sup>. A major

<sup>4</sup> Poehner and Lantolf (2003) distinguish between *interventionist* and *interactionist* approaches to dynamic assessment. The procedure discussed here represents the *interventionist* approach, which is closer in concept to view of assessment taken in this book.

<sup>5</sup> Such research as this raises many interesting questions, such as how one controls the many variables in such complex intervention, which are beyond the scope of this book.

drawback to dynamic assessment is that the mediation process can be time consuming, although since mediation is in fact teaching, perhaps this drawback is largely mitigated. The development of dynamic assessment programmes, and research into their effectiveness, is fairly new in the language testing field and its impact on mainstream language assessment remains to be seen. However, there is no doubt that it is an exciting concept that has the potential for providing new perspectives on learners and learning.

#### 4.6.6 Summary

In this section we've considered a number of alternative approaches to language assessment which are intended to elicit richer, more meaningful assessment data, usually requiring learners to perform some kind of goal-oriented task, invoking a relevant context, and reflecting everyday classroom/learning activities. The objective is to assess not only the products of learning, but also the processes and higher level thinking skills. There is a temptation to believe that because these alternatives in assessment are closely related either to classroom activities or to real-world language use activities they are automatically authentic and valid; that because they usually involve multiple tasks, drawing on a range of abilities, and are rated by experienced, trained professionals, they are automatically reliable. It is important even with, and perhaps *especially* with, these assessment alternatives that we make a validity argument, providing evidence that our assessments are authentic, valid, and consistent, as discussed in Chapter 2. It is particularly important with such complex procedures as task-based, performance or dynamic assessments that we define as precisely as possible the ability construct to be measured so that the inferences we make about learners' levels of ability and the decisions made on that basis are as fair and useful as we can make them.

#### 4.7 CONCLUSION

In this chapter I have discussed a number of related perspectives on and approaches to language testing. Depending upon the test purpose, the kinds of inferences we wish to make about learners' language abilities, and the types of decisions we might need to make in light of these inferences, different combinations of perspectives and approaches might be relevant. None of them, in and of itself, constitutes bad assessment practice, just as none of them is the one and only best way to assess language ability. Having decided upon an approach and developed the assessment according to procedures outlined in Chapter 3, we need to consider how to provide evidence that our test is working the way we intend. One type of evidence that is commonplace in the language testing field is of a statistical nature, owing to the conceptualisation of language testing as measurement, and, whether we wish to conduct our own statistical analyses of test performance or to understand the analyses others have carried out, an introduction to some statistical concepts and techniques will be of value. That is the topic of the next chapter.