

tests, b) democratization of testing through giving more attention to learners as one of the primary stake holders and c) a call for more learner autonomy and responsibility.

Notes:

1. This paper draws a fine distinction between *testing* on the one hand and *assessment* on the other.
2. Also known as Structural Approach

References

- Alderson, J. C., Banerjee, J. (2001). Language Testing and Assessment (Part I). *Language Teaching*, 34, 213-236
- _____ (2002). Language Testing and Assessment (part II). *Language Teaching*, 35, 79-113
- Alderson, J. C., Clapham, C. (1992). Applied Linguistics and Language Testing: A Case Study of ELTS Test. *Applied Linguistics*, 13(2), 149-167
- Bachman, F. L. (1991). What does Language Testing Have to Offer? *TESOL Quarterly*, 25(4), 671-704
- _____ Fundamental Considerations in Language Testing. Oxford University Press
- Bachman, F. L., Palmer, A. (2000). Language Testing in Practice. Oxford University Press
- Brown, D. J., Hudson, T. (1998). The Alternatives in Language Assessment. *TESOL Quarterly*, 32(4), 653-675
- Buck, G. (1997). The Testing of Listening in a Second Language. In C. Clapham & D. Corson, (Eds.), *Language Testing and Assessment* (Vol. 7, pp. 65-74) Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Eckstein, M. A., Noah, H. J. (1993). Secondary School Examinations: International Perspective on Policies and Practices. Yale University Press, New Haven.
- Farhady, H., Ja'farpur, A., Birjandi, P. (2006). Testing Language Skills: From Theory to Practice. Tehran: SAMT
- Fulcher, G. (2000). The 'Communicative' Legacy in Language Testing. *System*, 28, 483-497
- Hamp-Lyons, L. (2000). Social, Professional and Individual Responsibility in Language Testing. *System*, 28, 579-591
- Harris, D. (1991). Testing English as a Second Language. McGraw-Hill Book Company
- Heaton, B.J. (1990). Writing English Language Tests. Longman Inc., New York
- Lado, R. (1961). Language Testing. New York: McGraw-Hill
- Laufer, B., Nation, P. (1999). A Vocabulary-Size Test of Controlled Productive Ability. *Language Testing*, 16 (1), 33-51
- Morrow, K. (1979). Communicative Language Testing: Revolution of Evolution? In: Brumfit, C. K., Johnson, K. (Eds.). *The Communicative Approach to Language Teaching*. Oxford University Press, Oxford. 143-159
- Rea-Dickens, P. (1997). The Testing of Grammar. In C. Clapham & D. Corson, (Eds.), *Language Testing and Assessment* (Vol. 7, pp. 87-97) Dordrecht, The Netherlands: Kluwer Academic Publishers.
- _____ Fossilization or Evaluation: On the Case of Grammar Testing. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lurnley, T. F. McNamara & K. O'Loughlin (Eds.), *Experimenting With Uncertainty: Essays in Honor of Allan Davies* (Studies in Language Testing Series, Vol. 11, pp. 22-32). Cambridge: University of Cambridge Local Examinations Syndicate and Cambridge University Press.
- Ross, J. S. (2005). The Impact of Assessment Method on Foreign Language Proficiency Growth. *Applied Linguistics*, 26(3), 317-342
- Richards, C. J., Rogers, S.T. (2002). *Approaches and Methods in Language Teaching*. Cambridge University Press.
- Spolsky, B. (1989). 'Communicative Competence, Language Proficiency and Beyond.' *Applied Linguistics*, 10(2), 138-156
- Wall, D. (2000). The Impact of High-Stakes Testing on Teaching and Learning: Can This be Predicted or Controlled? *System*, 28, 499-509

assessment of language skills. As Rea-Dickens put it, the lack of explicit assessment of grammatical points does not mean that grammar is not assessed at all. Testing of grammar could be pursued implicitly while assessing speaking and writing in form of grammatical accuracy of the written or spoken discourse. Grammatical knowledge is also required for the successful accomplishment of reading tasks whereas knowledge of grammatical and rhetorical structures influences our understanding of the text's details.

● *Alternative Assessment*

Partly influenced by the communicative movement and partly under the influence of humanistic psychology, the 1990s witnessed a surge of interest in alternative modes of assessment among language teachers and testing practitioners. Alderson and Banerjee (2001, p. 228) delineate alternative assessment in the following manner:

Alternative assessment is usually taken to mean assessment procedures which are less formal than traditional testing, which are gathered over a period of time rather than being taken at one point in time, which are usually formative rather than summative in function, are often low-stakes in terms of consequences and are claimed to have beneficial washback effects.

Among alternative assessment procedures, self-assessment, peer assessment, teacher observation, portfolios, logs, journals, videotapes, audiotapes, and conferences are a few to be mentioned. Although such procedures may be time consuming and more difficult to administer and score compared to *traditional* test methods, e. g., true-false, matching, multiple choice, fill-in, short answer questions, etc., their claimed advantages turn them into an appealing option for language testers and teachers. Brown and Hudson (1998, p. 654) list twelve positive characteristics of

alternative assessment resulting in their preferred status over traditional assessment methods:

1. require students to perform, create, produce or do something;
2. use real world context or simulations;
3. are non-intrusive in that they extend the day-to-day classroom activities;
4. allow students to be assessed in what they normally do in class every day;
5. use tasks that represent meaningful instructional activities;
6. focus on processes as well as products;
7. tap into higher level thinking and problem-solving skills;
8. provide information about both the strengths and weaknesses of students;
9. are multiculturally sensitive when properly administered;
10. ensure that people, not machines, do scoring, using human judgment;
11. encourage open disclosure of standards and rating criteria; and
12. call upon teachers to perform new instructional and assessment roles.

As a result of the increased interest in the alternative procedures within the last decade, the conventional teacher-dominated summative assessment methods have gradually lost their appeal in favor of the more flexible learner-centered formative assessments. Consequently, "In many second and foreign language instructional contexts, assessment practices have increasingly moved away from objective mastery testing of instructional syllabus content to ongoing assessment of the effort and contribution learners make to the process of learning" (Ross, 2001; p. 318). In brief, the genesis of this new obsession with alternative procedures and formative assessment could be traced to a) the higher significance attributed to face validity of language

- What an Authentic Language Test Entails?

In order to justify the use of language tests, we need to be able to demonstrate that performance on language tests corresponds to language use in specific domains other than language test itself. One aspect of demonstrating this pertains to the correspondence between the characteristics of TLU [Target Language Use] tasks and those of the test tasks. ... We define authenticity as the degree of correspondence of the characteristics of a given language test task to the features of a TLU tasks (Bachman & Palmer, 2000; p. 23)

In effect, authenticity is a feature of language tests which enables us to envisage test takers' performance with regard to their future target domains on the basis of how well they perform on language test items. In the other words, for language test tasks to be called authentic, they should represent a true sample of the actual tasks test taker will be required to perform in TLU situations. This correspondence of test tasks and TLU tasks is not an all-or-nothing phenomenon and is better to be viewed as a relative construct which varies along a continuum from highly authentic to highly non-authentic. In this sense, the common dichotomy of language tests into *either authentic or non-authentic* can not accurately reflect the complexities of the issue and may culminate in oversimplification of the concept. This is demonstrated in the following figure:



In somewhat a more detailed taxonomy, Bachman (1991) distinguishes between *situational* and *interactional* authenticity. Situational authenticity demands a match of test features with those of target tasks while interactional authenticity pertains to the degree of test takers' language

knowledge involvement in accomplishing test tasks. Based on this categorization, it is possible for a language test to be a) high on situational authenticity while having little degrees of interactional authenticity, b) having high interactional authenticity but with little situational authenticity, c) high on both situational and interactional authenticity and d) having little situational and interactional authenticity.

- What is Meant by Performance Orientation and Real Life Outcomes?

"The assumption underlying the performance tests advocated and developed in the 1980s was that the observation of behavior that mirrored 'real-world communication' would lead to scores indicating whether the learner could perform in the real world. This 'real world' involves interaction, unpredictability, and the integration of skills" (Fulcher, 2000; p. 489). For a language test to be interaction-based there should be an unpredictable interface between the participants with an integration of both productive and receptive language skills at work in the meantime (Morrow, 1979). Moreover, communicative language tests should be scored on the basis of their real world outcomes, i. e., whether the learner was able to achieve a certain communicative effect (Fulcher, 2000). This means that what really counts is test takers' success in accomplishing *real life tasks* through the use of language and not merely accuracy of his language.

Under such influences, testing of grammar experienced somewhat of a decline in this period. For Rea-Dickens (1997, 2001) the current lack of interest in direct testing of grammar could be traced back to the communicative movement of 1980s which shifted the locus of attention from teaching and testing of grammar to the more communicative aspects of language use. As a result, grammar came to be assessed implicitly during the

and unrelated linguistic chunks functioning independently of each other, integrative approach tackles language from a more holistic perspective. That is, language use is no longer viewed as putting the small decontextualized bits together, but as a dynamic process in which various skills and sub skills are employed in accomplishing real life tasks. Hence, any comprehensive account of one's language involves assessing his/her ability to use the language in the appropriate context with the integration of different language (Farhady et al., 2006). According to Heaton (1990),

This [integrative] approach involves the testing of language in context and is thus concerned primarily with meaning and the total communicative effect of discourse. Consequently, integrative tests [unlike discrete-point items] do not seek to deparate language skills into neat divisions in order to improve test reliability.

Instead, they are often designed to assess the learner's ability to use two or more skills simultaneously. Thus, integrative tests are concerned with a global view of language proficiency ... (p.16).

Among the tests most peculiar to this era, cloze, dictation, composition writing, oral interview and translation are a few to be mentioned. As is the case here, almost all these test types entail assessing language use in such a way that requires the examinee to draw upon more than one aspect of his language knowledge in successfully accomplishing the test task. Indeed, this re-orientation from discrete -point to integrative approach took place mostly under the influence of Chomsky's *language competence hypothesis*. In fact, transformational linguists on the one hand, and *rationalist psychologists* on the other brought to the forefront the notion that language and language learning is completely a systematic and rule governed enterprise influenced by learners' cognitive abilities. With such new notions, a

demand was made of language testing specialists to re-direct their theories and practices so that they better mirror the realities of language and language use knowledge.

● *Communicative Language Testing*

Influenced by the cognitive psychology and Hymes' model of communicative competence, which highlighted the ability to use language in its appropriate social context as an indispensable aspect of language proficiency, language testing in the late 1970s and the early 1980s witnessed one of its biggest changes of its direction. The new era, known as *communicative language testing* and labeled as "the Promised Land" of language testing history (Morrow, 1979; p. 144, qtd. in Fulcher, 2000), was indeed a reaction to the previous overemphasis of validity and reliability at the expense of *authenticity* and face validity. Heaton (1990, pp. 19-20) Characterizes communicative language tests follows:

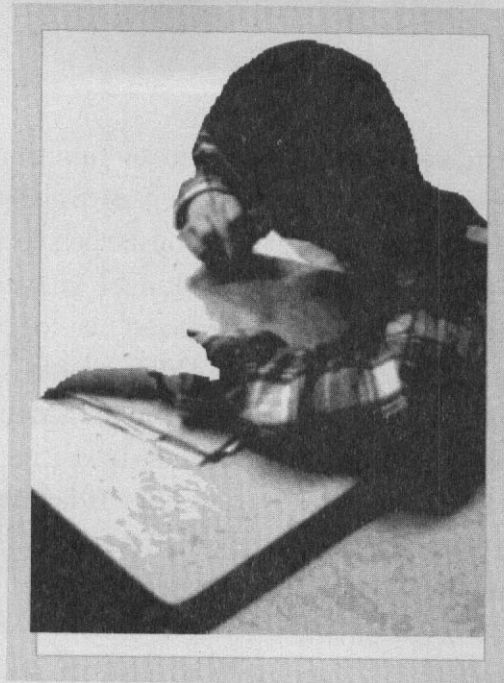
Communicative tests are concerned primarily (if not totally) with how language is used in communication. Consequently most aim to incorporate tasks which approximate as closely as possible to those facing the students in real life. Success is judged in terms of the effectiveness of the communication which takes place rather than formal linguistic accuracy. Language 'use' is often emphasized to the exclusion of language 'usage'. ...Unlike the separate testing of skills in the structuralist approach, moreover, it is felt in communicative testing that sometimes the assessment of language skills in isolation may have only a very limited relevance to real life.

Fulcher (2000) in his account of communicative language testing identifies communicative tests with three basic qualities of authenticity, performance orientation and real life outcomes. Each of these concepts are explained in more details in the following sections.

the *Discrete Point Model* of language testing came into being which was primarily concerned with testing language skills (listening, speaking, reading and writing) and components (vocabulary, grammar and phonology) as the backbones of any linguistic system's knowledge. This model, which has Lado's (1961) skills/components of language proficiency as its heart, is described by Heaton (1990, p.15) along the following lines:

This [structuralist] approach is characterized by the view that language learning is chiefly concerned with the systematic acquisition of a set of habits. It draws on the work of structural linguistics, in particular the importance of contrastive analysis and the need to identify and measure the learner's mastery of the separate elements of the target language: phonology, vocabulary and grammar. Such mastery is tested in using words and sentences completely divorced from any context on the grounds that a larger sample of language forms can be covered in the test in a comparatively short time. The skills of listening, speaking, reading and writing are also separated from one another as much as possible because it is considered essential to test one thing at a time.

As the above quote clearly indicates, decontextualization of test items, i.e., divorcing samples of language from its surrounding linguistic and non-linguistic context, is a common practice in the tests of this period. Besides, an effort used to be made to separate language skills and components into neat division for testing purposes. As a result, a good test of listening was considered to be the one which only and only drew upon the listening skill and did not have intervention from other language skills and components. In practice, however, developing such pure tests was doomed to failure since the real use of language entails the integration of a



number of skills and components. Indeed, people hardly ever come across a situation in their language use in which only a single skill or component suffices for the successful completion of real-life tasks.

Generally speaking, within the framework of the discrete-point model, there is an overemphasis of reliability and validity and a craze for objectivity in language tests. This, in turn, accounts for the huge popularity of multiple-choice and true/false items in the tests of this period, e.g., TOEFL. Although the skills/components model of language proficiency on which the structuralist approach is based was welcomed in the 1950s, it had its own critics too. Among the most frequently cited shortcomings of the skills/components model were 1) lack of an explicit indication of how language skills and components are related and 2) failure to recognize the context of language use (Bachman, 1995). Realization of these inadequacies was contemporaneous to the introduction of Hymes *Communicative Competence* model of language proficiency which marks a major turning point in the language teaching history.

● *Integrative Approach*

Instead of viewing language as a set of discrete

language use when designing their tests” (Alderson & Clapham, 1992;p.149).

Lacking any particular linguistic or psycholinguistic foundations, Grammar-Translation Method (GTM) of language teaching experienced its heydays during the nineteenth and early decades of the twentieth century. In deed, GTM was the heritage of teaching Latin in Europe which was later on exploited for the instruction of ‘modern’ European languages, e.g., French, English, German. Also known as the Prussian Method in the United States, GTM is based on the following pre mises (Richards and Rogers, 2002):

1. The goal of foreign language study is learning to read its literature and benefiting from the mental discipline and intellectual developments associated with it;
2. Reading and writing are the principal components of any language;
3. Translation is one of the most effective techniques for learning a language;
4. Sentence is and should be the basic unit of teaching and language practice;
5. Accuracy, particularly in translation and study of grammar, is very important;
6. Grammar should be taught deductively and in an organized and systematic way;
7. The student’s native language must be the medium of instruction.

The testing approach commonly referred to as *the Essay Translation Approach* shares many of these assumptions about the nature of language and language use. Belonging to the so-called ‘pre-scientific’ era of language testing, a test based on the essay translation approach typically consists of essay writing, translation and grammatical analysis with a heavy literary and cultural bias (Heaton, 1990). Test takers are usually asked to write native-like and sometimes scholarly essays

with rigid literary qualities. Failure to demonstrate such an expertise is usually interpreted as an indication of limited language ability. The translation technique used also demands testees to translate decontextualized and literary laden sentences from their first language to the second and vice versa. In addition, grammatical analysis, usually in form of verb congregation and explicit stating of the rules, is another indispensable aspect of this outdated testing tradition. On the contrary, when it comes to language testers, no particular skill is required, only their subjective judgments on the quality of learners’ responses without any objective criteria for correctness.

● *Discrete-Point Approach*

The introduction of the Audiolingual Method, along with the increased opportunities for communication, brought about the decline of GTM by the mid-twentieth century. As Richards and Rogers (2002,p.53) put it, by the mid 1950s “[the] combination of structural linguistic theory, contrastive analysis, aural-oral procedures and behaviorist psychology led to the Audiolingual Method.” The structural linguistics’ view of language proficiency was that of a system composed of structurally related components, e.g., phonemes, morphemes, words, phrases, clauses and sentences. In fact, “learning a language, it was assumed, entails mastering the elements or building blocks of the language and learning the rules by which these elements are combined, from phonemes to morphemes to words to phrase to sentence” with the speech being the primary medium of language use (Richards, 2002; p.55). Equally influential on Audiolingualism was the behavioral psychology whose view of learning was that of habit formation with particular attention given to *stimulus*, *response* and *reinforcement* as the prerequisites of any human learning including the verbal behavior.

Under the heavy influence of Audiolingualism,

students' achievement. Finally, testing can also be used as a tool for clarifying instructional objectives and, in some cases, for evaluating the relevance of these objectives and the instructional materials and activities based on them to the language use needs of students following the program of instruction.

All in all, the impacts and status of language tests, along with the decisions made on their basis, has turned testing/assessment¹ into potent tools whose influences go far beyond the limits of classrooms. In fact, many of the important decisions our society makes for individuals' academic and professional future is directly or indirectly influenced by how well language testing specialists and practitioners fulfill their commitments (Hamp-Lyons, 2000; Wall, 2000; Bachman, 2000; Bachman & Palmer, 2000). Therefore, it is imperative that our testing practices, at whatever level and scope, provide a truly representative sample of individuals' language abilities so that informed decisions could be made about test takers. This, in turn, requires that our testing traditions be enlightened with the latest research findings regarding a) the new language testing trends and assumptions, and b) the practical utility of different testing procedures. In fact, this seems to be the prerequisite for the development of any fair and reliable language test.

This state-of-art paper presents an overview of major language testing developments, both theoretical and practical, within the last century. In fact, we try to recapitulate the most influential testing paradigms along with their underlying theoretical assumptions and their subsequent practical realizations. This can provide a useful base of knowledge for those interested in language testing issues, including language teachers. Furthermore, knowledge of what has been, and is, going on in language testing domain seems to be an indispensable aspect of any apt learning and teaching practice.

Major Language Testing Paradigms

During its history, language learning and teaching has witnessed quite a few changes of direction culminating in the reconceptualization of previously taken-for-granted theoretical assumptions. These changes are in turn the result of modification in the underlying assumptions about the nature of language, i.e., linguistic theory, and the language learning process, i.e., psycholinguistic theory. When it comes to language testing, similar paradigm shifts can also be traced which are mostly stimulated by the changes in the corresponding linguistic and psycholinguistic theories of the time. In the other words, what goes on theoretically and practically in the domain of language learning and teaching has an immediate consequence for language testing practices.

Influenced by language learning/teaching assumptions and practices, language testing developments in the twentieth century can be described within five more or less distinct paradigms: *essay-translation approach*, *discrete-point approach*, *integrative approach*, *communicative language testing* and *alternative assessment procedures*. Each of these testing traditions has its own peculiar account of what the nature of language is and what it means to know a language. What follows is a brief overview of these four major language testing trends, their theoretical bases and their practical realizations.

● *Essay-Translation Approach*

Underlying any sound language test should be a sound model defining what it means to know a language, i.e., a coherent model of language proficiency and language use (Spolsky, 1989). As a matter of fact, "since language tests inevitably embody a view of language and indirectly a notion of language learning, it is important that test developers take account of generally accepted views of the nature of language, language proficiency and

ترجمه‌ای با تأکید بر اهمیت دانش ساختاری و توانایی ترجمه در فرایند ارزیابی؛ ۲. دوره‌ی تجزیه‌ای با تأکید بر ارزیابی مستقل مهارت‌های متفاوت زبانی؛ ۳. دوره‌ی ترکیبی با تأکید بر ارزیابی مهارت‌های متفاوت زبانی در ارتباط با یکدیگر؛ ۴. دوره‌ی ارتباطی با تأکید ویژه بر ارزیابی زبان در فرایند برقراری ارتباط؛ ۵. دوره‌ی شیوه‌های جایگزین با نگاهی نو به روش‌های جدید ارزیابی.

کلید واژه‌ها: دوره‌ی ساختاری - ترجمه‌ای، دوره‌ی تجزیه‌ای، دوره‌ی ترکیبی، دوره‌ی ارتباطی، دوره‌ی شیوه‌های جایگزین.

Abstract

The impacts and status of language testing, along with the decisions made on its basis, has turned testing/assessment into a potent tool whose influences go far beyond the limits of classrooms. In fact, many of the important decisions our educational system makes for individuals' academic and professional life is directly or indirectly influenced by how well language testing specialists and practitioners fulfill their commitments (Hamp-Lyons, 2000; Wall, 2000; Bachman, 2000; Bachman & Palmer, 2000). For this reason, it seems imperative that our testing practices, at whatever level and scope, provide a truly representative and unbiased sample of individuals' language abilities so that informed decisions could be made on their basis. This, in turn, requires that our testing traditions be enlightened with the latest research findings regarding a) the new language testing trends and assumptions, and b) the practical utility of different testing procedures.

Overall, influenced by the learning and teaching assumptions, language testing throughout its history has witnessed quite a few paradigm shifts culminating in the rejection of many previously embraced testing traditions. In line with such swaying of pendulum in the underlying testing theories, one would logically expect to encounter reconsiderations of the old testing practices and their replacement with the ones reflecting newer trends. In brief, language testing history can be described within five more or less distinct periods of *essay-translation*, *discrete-point*, *integrative*, *communicative* and *alternative assessment*. This state-of-art paper tries to tap on each of these periods briefly and provide an overview of their implications for the classroom practices of language teachers.

Key Words: Alternative Assessment, Communicative Language Testing, Discrete-Point Testing
Essay-Translation Approach, Integrative Tests

Introduction

Among various functions attributed to language tests, Eckstein and Noah (1993) enumerate the following: encouraging higher levels of competence and knowledge, checking patronage and corruption, allocating sparse places in higher education, and measuring and improving the effectiveness of teachers and schools (qtd. in Wall, 2000). A somewhat similar comment is made by Bachman and Palmer (2000, p.8) who believe, *Language tests can be a valuable tool for providing information that is relevant to several*

concerns in language teaching. They can provide evidence of the results of learning and instruction, and hence feedback on the effectiveness of the teaching program itself. They can also provide information that is relevant to making decisions about individuals, such as determining what specific kinds of learning materials and activities should be provided to students, based on the diagnosis of their strengths and weaknesses, deciding whether individual students or an entire class are ready to move on to another unit of instruction, and assigning grades on the basis of

Language Testing at the Start of the Third Millennium: Theory and Practice Revised

چکیده

Professor Akbar Mirhassani, Full Professor in TEFL

Babak Dadvand, MA in TEFL

Email: babak.dadvand@gmail.com

Hadi Azimi, MA in TEFL

Email: azimi.ho@gmail.com

Tarbiat Modares University

در جوامع کنونی، محدودیت منابع و امکانات انسان را ناگزیر به گزینش کرده است. در واقع، برخی گزینش‌ها دارای چنان اهمیتی هستند که حتی می‌توانند، مسیر زندگی آینده‌ی افراد را تعیین کنند. در این بین، ارزیابی دانش زبانی نیز از جمله ارزیابی‌هایی است که پایه‌ی گزینشی بسیاری از تصمیمات در جامعه را تشکیل می‌دهد. در حقیقت، به دلیل نقش بسزایی که چنین ارزیابی در شکل‌دهی آینده‌ی تحصیلی و شغلی دانش‌آموزان و دانشجویان دارد، لازم است، تمامی کسانی که به نحوی با چنین گزینش‌ها، تصمیمات و ارزیابی‌هایی در ارتباط هستند، با آگاهی کامل از مسئولیت خطیر خویش، تمامی دقت و تلاش خود را در انجام آن به کار گیرند. این امر به نوبه‌ی خود مستلزم شناخت کامل دانش و توانایی زبانی، آشنایی با شیوه‌های روزآمد و علمی، و ارزیابی چنین توانایی است. با نگاهی اجمالی به تاریخچه‌ی آموزش زبان انگلیسی در یک صدسال گذشته، به راحتی می‌توان فراز و نشیب‌های فراوانی را که مقوله‌ی آموزش زبان پشت سر گذاشته است، مشاهده کرد. بخش عمده‌ای از این تحولات تحت تأثیر نظریه‌های نوین زبانی و یادگیری متجلی گشته‌اند. چنین تغییراتی در نظریه‌های زبان و یادگیری، به نوبه‌ی خود تأثیرات بسزایی را در نحوه‌ی نگرش ما به مبحث ارزیابی توانایی مذکور به همراه داشته‌اند. به طور خلاصه، تاریخ شیوه‌های ارزیابی مهارت زبانی را می‌توان به پنج دوره‌ی نسبتاً متمایز با ویژگی‌های خاص تقسیم کرد. این پنج دوره که به نحوی نشان‌دهنده‌ی تکامل نگرش به مقوله‌های زبان، یادگیری و ارزیابی هستند، عبارت‌اند از: ۱. دوره‌ی ساختاری-

