Validity in the Assessment of Listening Skills in Language Learning

Autor: Rebecca Barker, Helen (Maestra. Especialidad en Lengua Extranjera (Inglés)).
Público: Profesores de lenguas extranjeras. **Materia:** Lenguas Extranjeras. **Idioma:** Inglés.

Title: Validity in the Assessment of Listening Skills in Language Learning.
Abstract
Listening is arguably the most complex language skill to assess due to the fact it cannot be observed directly. It is a receptive skill involving internal processes; hence, any test employed has to produce a reliable external quantification of such processes. The need for assessment in the classroom is unquestionable as it can provide invaluable information about the teaching and learning processes and outcomes, but only when it is properly developed and implemented. This article discusses construct and cognitive validity in the assessment process.
Keywords: Listening assessment, construct validity, cognitive validity

Título: La validez en la evaluación de las habilidades auditivas en el aprendizaje de los idiomas.
Resumen
La evaluación auditiva es, sin duda, la más compleja de todas las destrezas lingüísticas puesto que no se observa directamente. Es una habilidad receptiva, por lo tanto, de cualquier evaluación ha de obtenerse una cuantificación externa y fiable de los procesos internos implicados. La necesidad de evaluar en la aula es indudable, se facilita información esencial sobre el proceso de enseñanza-aprendizaje y sus resultados, pero solo cuando se desarrolla y se implementa de una manera adecuada. Este artículo habla de la validez de constructo y la validez cognitiva en el proceso de evaluación.
Palabras clave: Evaluación auditiva, validez de constructo, validez cognitiva.

Recibido 2018-05-02; Aceptado 2018-05-08; Publicado 2018-06-25; Código PD: 096006

VALIDITY

In 1961 Lado introduced the concept of 'validity' in language testing (D'Este, 2012). It can be defined as the extent to which a test accurately measures what it is intended to measure. This article examines two different types of validity in the field of listening assessment.

CONSTRUCT VALIDITY

The term 'construct validity' refers to the *"the extent to which a language test is representative of an underlying theory of language learning"* (Huang, 2013). According to Vandergrift and Goh (2009) *"Construct validity is important for assessment because it entails defining construct, operationalising the behaviours that need to be assessed, and then creating tasks (appropriate texts and response items) to elicit these behaviours."* This can be done by either defining the competence - the knowledge, skills and abilities - that learners should possess, or the tasks they should be able to perform.

The primary stage curriculum advocates a communicative approach to language teaching and learning. The main general stage objective for the area of foreign languages refers to the achievement of a basic communicative competence in the target language; thus, a listening construct based on communicative theory is required in order to assess listening.

Buck (2001) proposes frameworks for describing listening competence and listening tasks (see Buck, 2001:104 and 107 for the full descriptions).

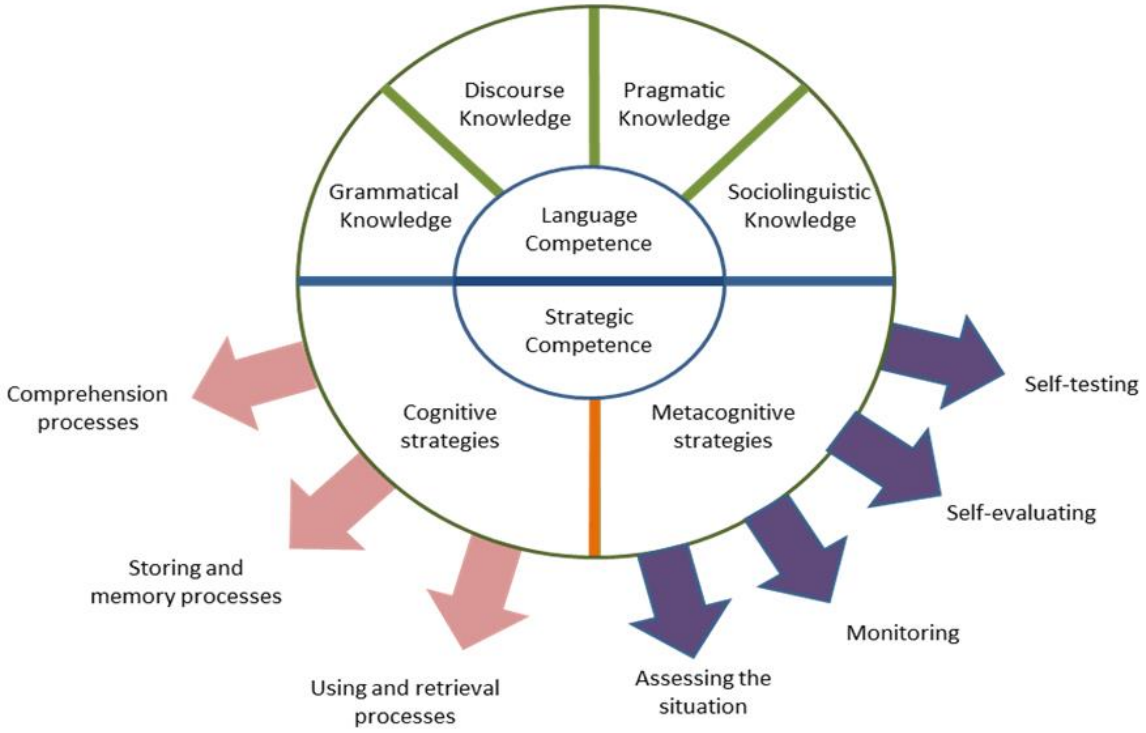

Figure 1. Author's own illustration of Buck's framework of listening competence (2001).

The listening competence framework, as illustrated in figure 1, comprises of a range of sub-components. However, Buck admits that this framework is limited in nature because it does not offer test developers guidance as to what the most important elements are, nor what ones should be included and to what extent in a given test. It is difficult to implement because *"determining the competencies that underlie performance on a set of test tasks is a complex and indirect process, and we have no way of knowing for certain which competencies are required by any particular task"* (Buck, 2001).

On the other hand, a task-based approach to a listening construct definition involves creating a suitable list of target-language use/ real-life performance tasks and designing a set of test tasks to replicate them. It is assumed that the skills used to perform the real-life and the replicated tasks are the same, hence overcoming the difficulty to determine them as in the competence-based approach. But is it possible to accurately replicate tasks in test situations? If tasks cannot be reproduced exactly, then the assumption that the same skills are being used can be called into question, leading us back to the same problem with validity we have with the first approach. To overcome this, Buck (2001) suggests a matrix approach to construct definition by taking into account both competence and task because *"we can get the best of both approaches, and hope that the strengths of each will compensate for the weaknesses of the other. To do this, it may be sufficient just to produce lists of relevant abilities and lists of the most important target-language use tasks, and then go ahead and develop test tasks that require both."*

COGNITIVE VALIDITY

This concept refers to *"the extent to which the tasks employed succeed in eliciting from candidates a set of processes which resemble those employed by a proficient listener in a real-world listening event"* (Field, 2013). It emphasises the need to gain a better understanding of the construct under consideration (in this case, listening) through the development of empirically-substantiated process models. Such models identify the phases through which a listener typically proceeds, thus establishing a framework *"for determining in a systematic way how the various processes which make up performance in a given skill are represented, explicitly or implicitly, in the types of response required of the test taker"* (Field, 2013).

The process listening model proposed by Field (2013) is composed of five different levels of processing or representation. The first three levels, as illustrated in figure 2, represent the lower-level listening processes, that is, those processes involved in encoding the message into language. Each level requires a distinct type of knowledge, which is used to transform the message. The higher-levels are involved in the addition of meaning. It is important to note that *"comprehension does not follow a strictly linear progression from the lower to the higher processing levels; rather, different*

levels may be operating concurrently, with breakdowns at one level compensated by 'positive information' at another, or with simultaneous breakdowns at higher- and lower-levels leading to miscomprehension altogether" (Harding, Alderson and Brunfaut, 2015).

The levels and what happens in each one are as follows:

- **Input decoding**
  The acoustic cues are converted into phonological forms of the language, for example phonemes and syllables.
- **Lexical search**
  It entails the identification of the words and their meanings that best match the sequences of sounds. To do this, listeners make use of the frequency with which words occur in natural speech and how words are associated through lexical networks, called spreading activation.
- **Parsing**
  This involves the integration of groups of words into larger units, normally clauses. It is facilitated by an understanding of a language's standard word order. When listeners perceive the completion of a unit, they construct a proposition, that is, they form an abstract idea about the unit. This idea replaces the linguistic form of the message.
- **Meaning construction**
  A proposition is essentially a literal understanding of the speaker's words. Due to the highly abbreviated nature of human communication, a listener must draw on various knowledge sources - pragmatic knowledge, external knowledge and discourse representation (the listener's knowledge of the interaction up to the point in question) - in order to make sense of it. It is an enrichment process in terms of meaning.
- **Discourse construction**
  This involves the integration of the new information into a representation of a larger listening event and testing its congruency with previous information.

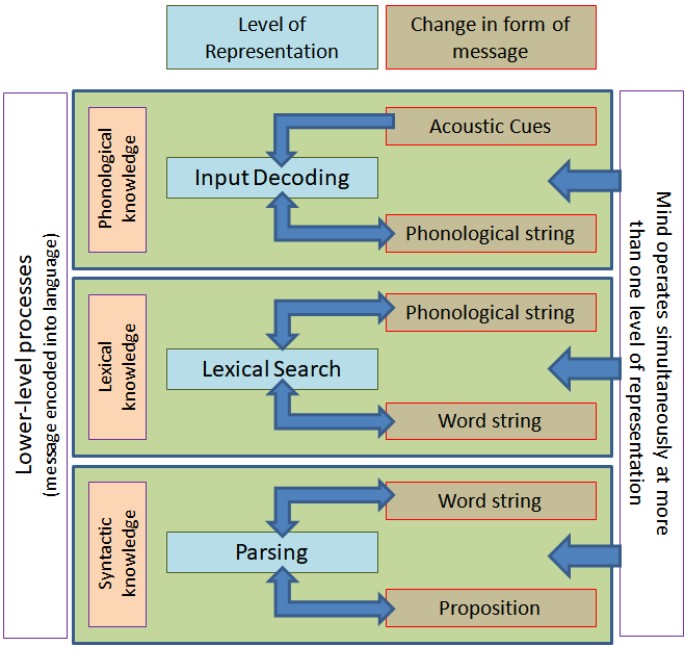

Figure 2 Author's own illustration of the first three levels of Field's process model (2013).

Field (2013) argues that *"it is not possible for the conditions under which a language test is administered to approximate closely to those of an actual listening or speaking event. The goal is to establish whether the tasks proposed by the test designer elicit mental processes adequately representative of those which a language user would employ in a target real-world context. The processes in question might relate to the way in which the user assembles or interprets input; or they might reflect the types of response required of the user by facets of the task."* This argument contrasts somewhat with the need to replicate real-life tasks as per Buck (2001).

With regards to the operationalisation of this model, Harding et al (2015) believe there is scope for "*operationalising*

elements of this model through discrete assessment tasks...[though it] is likely to be much more straightforward at the lower-levels of processing where the influence of information beyond the text - such as context - and knowledge sources...is much less pronounced." These authors propose a series of potential tools for diagnostic listening assessment classified into the different levels of representation, for example:

Level of Representation	**Example of discrete listening assessment task**
Input decoding	Discriminating phonemes using different speakers.
Lexical search	Counting the number of words in an utterance.
Parsing	Use of true/ false statements to focus on particular elements of spoken syntax in literal utterances.
Meaning representation	Aural tests of metaphor comprehension.
Discourse construction	Distinguishing between relevant and irrelevant information in a long text when asked to listen for a specific purpose.

Table 1. Potential tools for diagnostic listening assessment. Taken from Harding et al (2015).

According to Huang (2013) the literature on cognitive validity also provides pedagogical guidance for developing tests. She states that test developers must *"recognise that a test should not be all about assessing the linguistic knowledge (the product), but also about applying the skill (the process) [and they need to] keep the purpose of the test or task in terms of its relevant cognitive processes clearly in mind [as well as considering] whether the range of processes elicited by a task/test or a series of tasks/ tests is comprehensive enough to be representative of behaviours in a real-world setting."*

THE WAY FORWARD?

Achieving validity in listening assessment is challenging given the covert and complex nature of this language skill. An approach to listening assessment based on construct validity involves defining listening within a communicative theory framework whilst cognitive validity relies heavily on speech processing i.e. from how it is perceived to it being enriched by the listener.

Both concepts attempt to define the same phenomenon but from different stances. This in itself proposes a series of questions, such as, is the communicative approach the most adequate theory of language learning to use as a basis for defining listening? Since the 1980s authors like Swan (1985a and 1985b) have voiced concerns about this approach whilst others, for example Rodgers (2001), talk about a post-communicative period in which language teaching and learning should take on an eclectic approach. A change in theory would involve a change in the listening construct definition.

Surely, at some point, there must be some overlap between these two concepts and, perhaps, their integration may path the way to a more complete definition of listening. Only further research into this field of study will determine whether this is true. As it stands, neither one nor the other ensures intended measures are actually assessed; rather, they allow test developers to become closer to achieving this by creating awareness of what is involved when someone listens and producing frameworks and guidance on listening assessment. Until that time it would seem wise to take into consideration both types of validity paying particular interest to their limitations and the specific circumstances of each given test.

Bibliografía

- Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.
- D'Este, C. (2012). New Views of Validity in Language Testing. *EL.LE,* 1/1, 61-76.
- Field, J. (2008). *Listening in the language classroom*. Cambridge: Cambridge University Press.
- Field, J. (2013). Cognitive Validity. In Geranpayeh, A and Taylor, L (Eds.), *Examining Listening* (pp. 77-151).Cambridge, UK: Cambridge University Press.
- Harding, L., Alderson, C. and Brunfaut, T. (2015). Diagnostic Assessment of Reading and Listening in a Second or Foreign Language: Elaborating on Diagnostic Principles. *Language Testing*, 32/3, 317-336.
- Huang, L. (2013). Cognitive Validity. *Key Concepts and Theories in TEAL*. Spring 2013, 12-16.
- Rogers, T.S. (2001). Language Teaching Methodology. *CAL Digest*. September 2001, 1-8.
- Swan. M. (1985a). A Critical Look at the Communicative Approach (1). *ELT Journal*, 39/1, 2-12.
- Swan. M. (1985b). A Critical Look at the Communicative Approach (2). *ELT Journal*, 39/2, 76-87.
- Vandergrift, L. and Goh, C. (2009). Teaching and Testing Listening Comprehension. In M. Long and C. Doughty (Eds.), *The Handbook of Language Teaching* (pp.395-411). Chichester: Wiley-Blackwell.