Critical Studies of the Notion of "Language Ability" in Bachman and Palmer's Test Model

Byron Gong, Associate Professor, PhD

English Department, Soochow University, Taipei, Taiwan

Email: byronygong@yahoo.com

Abstract

This paper presents a critical discussion about the relationship between "language ability" in Bachman and Palmer's AUA (Assessment Use Argument, 2010) test model and the common concept of language ability. The purpose is to point out that although the definition of "language ability" in the influential AUA model is thought-provoking, its application would seem to be almost impossible in practice, especially regarding standardized tests. It is important to realize that the common concept of language ability (four language skills) is not the main concern in the AUA model. This paper claims that the test validation should be more focused on the observable test performance, i.e. language proficiency, rather than the psycholinguistic "competence or ability". As limited by space, the empirical analysis of the test scores (N=78, using Paired-Sample T Test and Pearson test) is just presented briefly to consolidate the author's claim. Generally, it is important for test designers to realize the limitations of the AUA model in practice. Rather than the unobservable competence, it is test performance to which we need our interpretations about language ability, i.e. language skills, to generalize.

Keywords: AUA test model, language ability, test performance, competence

Introduction

This paper provides a critical study on a fundamental theoretical issue in language testing, i.e. the relationship between "*language ability*" in Bachman and Palmer's AUA (Assessment Use Argument, *hereafter referred to as AUA*) test model and the commonly used linguistic concept of *test performance* as a reflection of testees' foreign language proficiency. This paper aims to discuss the limitations that the concept of *language ability* adopted by Bachman and Palmer's AUA test model might impose on the development of language testing at grassroots level in practice. It is important for test designers and stakeholders to realize that different definitions of *language ability* would lead to different interpretations of test results or performance, which in turn would lead to different interpretations of test validation. If test designers are mainly interested in test candidates' language "ability", which refers to the "hidden" linguistic competence from the perspective of Chomsky's psycholinguistics, cultural interaction and cognitive psychology, rather than interested in test candidates' test performance in terms of language skills (listening, speaking, reading, and writing), test designers may work out a number of different models concerning test takers' language competence, typically, such as the theoretical test models of CLA and AUA (Bachman, 1990; Bachman and Palmer, 1996, 2010); and those of Lado's (1961) for language and communication; and the post-Hymes model of Canale and Swain (1980). However, to a large degree, test models such as CLA and AUA are considered an essentially psycholinguistic view of language competence and are considerably lacking important interactional and sociolinguistic performance. In this sense, Bachman's model is extremely difficult for language testing practitioners to use it as an operational framework in practice (O'Sullivan and Weir, 2011, p.15).

It is certainly beyond the scope of this paper to conduct a comprehensive discussion of every aspect of the AUA model. In reality, it is worth noting that test performance based on language skills, i.e. listening, speaking, reading, writing, rather than the "language ability or competence" defined as a process of psychological or mental ability that is proposed in AUA, should be the main observable data from which test interpretation and validation can be

generated, which is especially true regarding large-scale standardized testing systems.

Language Ability and Language Performance

As for the relationship between the two terms of *language ability* and *test performance*, the prevailing concept in the academic circle of language testing is that test designers should be mainly interested in test candidates' "language ability" (Bachman and Palmer, 1990, 2010; Carr, 2011) rather than their actual test performance, because test performance is not equal to language ability, and performance may be affected by many other factors. The influential test models, such as the models of Communicative Language Ability (CLA) proposed by Bachman (1990), Bachman and Palmer's Task Language Use (TLU) domain (1996), and their latest Assessment Use Argument (AUA) in 2010, are generally based on a hypothetical framework of multi-components so that the issue of test validation can be established with evidence of test candidates' *language ability*. These multi-componential test models have provided a broader view of the development of language testing. At first glance, such view of assessing testees' language ability or competence sounds quite acceptable, and the goal of many foreign language tests is indeed to assess test candidates' language ability through their test performance.

However, it is important for test designers and stakeholders to notice that the claimed "language ability" advocated in Bachman's models does not completely refer to the common concept of language ability, nor the "four skills". Furthermore, according to Bachman and Palmer (1990, 2010), test researchers should not be mainly interested in the testees' performance of the four skills. Instead, they should pay more attention to understand the constructs or traits through their performance in a language test. In Bachman and Palmer's words, the familiar language skills (listening, reading, writing, and speaking) "are not the constructs to be assessed" in a test, and they are to be considered as "language use activities" according to Bachman's AUA model (2010, p.220). As for this claim, it might come as a shock to many of us test designers and ELT teachers who hold that test results through test takers' performance in a test are the most important point of reference. This is especially true when we talk about large-scale language tests, not in a classroom case of portfolio assessment of three or five pupils' homework. In other words, although the AUA model proposed by Bachman and Palmer (2010) provided a new perspective of language testing development, many concepts of the AUA model are rather debatable and even confusing, which may not be most applicable to those working as designers of large-scale standardized testing systems or classroom teachers at grassroots level.

Distinction between Language Ability and Language Performance

We can see the distinction between competence and performance from the early work of the linguist Noam Chomsky (1965). Competence is the knowledge that all speakers have of their *native language* which allows them to produce well-formed structures. It is tacit knowledge that speakers normally have no access to and cannot describe. Performance is "the actual use of language in concrete situations" (Chomsky, 1965:4), which may contain structures that, for various reasons, are not actually well-formed.

Meanwhile, the term performance refers to the application of one's competence or knowledge of the rules of language to actual communication (Davis et al., 1999, p.143). According to the traditional linguistic notion, one's knowledge of language is "hidden", and is referred to as **linguistic competence, or ability**. According to Chomsky's transformation-generative grammar, performance means a person's actual use of language. But in foreign language learning, a learner's performance may indicate his or her competence (Richards and Schmidt, 2002, p.392). For the purpose of this paper, **test performance** refers to test consequence, which is a sample of a test candidate's performance on the required test tasks, and such a sample should be one that examiners can assess.

The relationship between language competence and language performance based on Chomsky's linguistic concept is similar to the linguistic relationship between langue and parole proposed by Saussure. Although Saussure's distinction between "langue" and "parole" is similar to Chomsky's distinction between *competence* and *performance*, whereas for Saussure the repository of "langue" is the Speech Community, while for Chomsky the repository of "competence" is the "ideal speaker/hearer". Therefore, Chomsky's concept is basically psycholinguistic (Richards and Schmidt, 2002).

Main Weakness of the Notions of CLA, TLU and AUA

Generally speaking, the term "language ability" in Bachman and Palmer's AUA test models may place ordinary test designers at grassroots level in a difficult predicament, seemingly caused by unnecessary divergence of terminology. The definitions of *language ability* in the notions of Bachman and Palmer's AUA test model in 2010, Target Language Use (TLU) in 1996, and Bachman's Communicative Language Ability (CLA) in 1990 overlap in meaning considerably.

The three elements of "language competence, strategic competence, and psychophysiological mechanism" are the basis of Bachman's conceptual framework, and the CLA test model is a quite complicated theoretical framework, from which the detailed notions of TLU and AUA have been developed. But it is important to be aware that the CLA notion is based on its a priori assumption that communicative language ability being considered not only as a mental ability, but also as responses to settings of specific social contexts. According to some critics, the model of CLA proposed by Bachman (1990) is quite problematic as his view "represents an essentially psycholinguistic view of performance and is largely missing important interactional and sociolinguistic elements" (O'Sullivan and Weir, 2011:14-15). Furthermore, in criticizing Bachman's CLA model, O'Sullivan and Weir (2011:14) pointed out:

> a critical weakness of the model is that it proved to be extremely difficult if not impossible to operationalise, not least because of its daunting breath and depth, but also due to its lack of clear prioritization as to what might constitute criterial parameters for language testing purposes… As such, Bachman's model (1990) has contributed less than might have been hoped to empirical test validation.

In fact, the CLA model was considered as basically psychological, viewing language ability as a mental ability. McNamara (2003) criticized the Bachman CLA test model as being essentially psychological, and its assumed competence is increasingly questioned in studies of language use in context, and it is being found to be problematic in practical language testing contexts too.

Meanwhile, another feature of the CLA and AUA test models is that they should be used to assess a person's language ability through doing a test that is designed according to some specific requirements. Unfortunately, there lacks a clear clarification of what task can be considered within a specific setting. Although Bachman and Palmer (2010) provided a Framework of Language Task Characteristics, which is based on Bachman's proposal in 1990, the daunting depth of this framework leaves readers more questions. Bachman and Palmer (2010:142) mentioned that the AUA model is suitable for both low-stakes and high-stakes assessments, but there still remains a big question if the four skills are not their real concern. For example, we are not clear about the coverage of TLU regarding large-scale standardized testing systems, such as TOEFL, IELTS, TOEIC, GEPT in Taiwan, or CET in Mainland China. It seems that the more specific settings TLU tasks should be, the more difficult that TLU tasks would be for large-scale standardized proficiency tests.

On the other hand, Bachman and Palmer claim (2010:116), in most cases, they "are not interested so much in how test takers perform on the assessment itself, but rather in the information this performance provides us about test takers' language ability that applies

beyond their performance on the test itself" (p.116). It seems that they want to interpret and generalize test takers' language ability by means of "beyond the assessment itself to a particular TLU domain and its associated TLU tasks" (p.116). However, this sounds quite difficult to implement in practice because good performance tasks that can be considered within a specific TLU (Target Language Use) domain are usually not only time consuming, but also likely to be of small samples. Still, it is difficult for small samples to be representative of the TLU tasks within a particular domain to be reliable and valid for test users, not to mention the possible problems of subjectivity. The rational is that the more a test is designed according to a specific situation, the less generalizability of the language ability and validity would be.

As for the AUA model, the most problematic aspect seems to be what has been mentioned by O'Sullivan and Weir (2011) and McNamara (2003), i.e. an essentially psycholinguistic view of language ability and performance; largely missing important interactional and sociolinguistic elements. Furthermore, it seems that the AUA model is also extremely difficult to operationalise in practice, not least because of its daunting breath and depth, but also due to its lack of a clear list of priorities, especially for ordinary test designers and classroom teachers, as to what might constitute key parameters for language testing purposes at grassroots level.

The Divergent Definition of "language ability" in the AUA Test Model

The changed definitions of certain terminology in Bachman and Palmer's AUA test model have received critical attention. According to the book review in the journal of *Language Testing* (2011), we can see some comments on divergence of terminology as a perceived problem in the book *Language Assessment in Practice* by Lyle Bachman and Adrian Palmer (2010). Kane (2011) has pointed out the problem of divergence of terminology in this book, indicating that development of differences in terminology may have 'a failure to communicate.' (p.581)

The implied criticism in Kane's comments on the differences in terminology used in the book *Language Assessment in Practice* (Bachman and Palmer, 2010) is one of the problems to which researchers and developers in the circle of language testing should pay meticulous attention. Otherwise, it is possible for ordinary test designers to misunderstand the AUA test model in their test development, because the same terminology carries different meanings. Consequently, it is not difficult to imagine there could be "a failure to communicate" between researchers and ordinary test designers (or teachers) at grassroots level when the two sides adopt the same terminology to refer to different concepts. It would be an even sad situation if the latter adopted a certain test model while they were not aware of what their own traditional and familiar terms have been used to mean a concept that is miles away. As a result, all the related work in the test development, from test specifications to the validation of a whole testing system, may be significantly affected. In addition, it is important for test designers and stakeholders to realize that test performance would lead to different interpretations of test validation when there are different definitions of *language ability*. Therefore, from these viewpoints, divergence of terminology in AUA test model and its impact on testing development in practice appear to be the main research problem for the purpose of this paper.

In their book *Language Assessment in Practice*, Bachman and Palmer (2010) provided a detailed explanation of their definition of language ability for their AUA test model, which is essentially based on Bachman's earlier proposal in 1990 for his model of Communicative Language Ability (CLA). Bachman and Palmer (2010:33) explained their concept of language ability by stating that language ability contains two components: language competence and strategic competence. As for language ability, they stated:

> In this chapter we describe **language ability** as a capacity that enables language users to create and

interpret discourse. We define language ability as consisting of two components: language knowledge and strategic competence. Other attributes of language users or test takers that we also need to consider are personal attributes, topical knowledge, affective schemata, and cognitive strategies. (Bachman and Palmer, p.33)

As for "language knowledge", Bachman and Palmer (2010) explained: "Language knowledge can be thought of as a domain of information in memory that is available to the language user for creating and interpreting discourse in language use" (p.44). Next, as far as "strategic competence" is concerned, Bachman and Palmer (2010:49) stated as follows:

Strategic competence can be thought of as higher-order metacognitive strategies that provide a management function in language use, as well as in other cognitive activities. We view strategic competence as a set of metacognitive strategies….We would thus hypothesize that the metacognitive strategies we discuss here are involved not only in language use, but in virtually all cognitive activity.

From the above quotations, we can see that Bachman and Palmer's definition of "language ability" is basically psycholinguistic, focused on mental cognitive process. (*Note*: in the book, Bachman and Palmer explained that they prefer to use "strategies" instead of "processes" so as to cover both "controlled and automatic" cognitive activities, pp.57-58.) Regarding their explanation of "language knowledge", we can tell that it is actually a very true reflection of Chomsky's concept of "language competence". Meanwhile, since they stated that "We view strategic competence as a set of metacognitive strategies", this again illustrates psycholinguistic cognitive conception. Thus, Bachman and Palmer's definition of language ability is just within the scope of psycholinguistic framework, similar to Chomsky's "competence".

Accordingly, we feel interested in their concept of the commonly accepted concept of language ability in terms of language skills, and we found them holding a very different view of language skills. According to Bachman and Palmer's (2010:55), language ability has traditionally been considered, by language teachers and language testers alike, to consist of four skills: listening, reading, speaking, and writing.

It is not the author's intention of this paper to make a judgment of Bachman and Palmer's criticism of the limited traditional concept of language ability, and their divergence of terminology of language ability in AUA test model. However, we can see the traditional concept of language ability as the "four skills" has virtually been rejected in the AUA model. Bachman and Palmer (2010) have stated that they do not consider the commonly used language skills as language ability, because language skills, such as listening, speaking, reading, and writing, are considered as only activities of using a language. Specifically, Bachman and Palmer (2010:33-34) claimed "we argue that the notion of 'language skills' is problematic, and that what are commonly referred to as 'listening,' 'speaking,' 'reading,' and 'writing' can more usefully be conceptualized as language use activities". Thus, the common concept of language ability as the "four skills" has been turned down by Bachman and Palmer. In other words, it seems that we are actually speaking "different languages" in our discussion of the relationship between language ability and performance due to the divergence of terminology, at least.

As for the definition of "language ability" in AUA, Bachman and Palmer (2010) hold that it is necessary to define language ability "in sufficiently precise terms to distinguish it from other individual attributes that can affect assessment performance. We also need to define language ability in a way that is appropriate *for each particular assessment situation*" (p.43). Bachman and Palmer (2010:43) further explained that the term "language ability" used for their AUA model is also called a "construct" for a particular testing situation as follows: "When we define an ability this way, for purposes of measurement, we are defining what we call a **construct**. For our purposes, we can consider a construct to be the specific

definition of an ability that provides the basis for a given assessment or assessment task and for interpreting scores derived from this task."

Most importantly, Bachman and Palmer (2010) remind us that their concept of "language ability" differs from what are commonly referred to as the four skills 'listening,' 'speaking,' 'reading,' and 'writing', which are actually not language ability, but activities of language use. In other words, it is worth noticing that this distinction has again demonstrated Bachman and Palmer's proposal that test developers should be more interested in test candidates' real "language ability" rather than the performance of their four skills. Bachman and Palmer (2010:115) further explained that "in language assessment, the construct in which we are most interested are aspects of language ability". More specifically, when writing about language skills as the construct, Bachman and Palmer (2010) stated that they have taken the position that the "familiar 'language skills' (listening, reading, speaking, and writing) are not the constructs to be assessed. Rather, we consider them to be language use activities." (pp. 220-221)

In addition, "language ability" for the purpose of AUA test model is to be defined precisely so that other factors that can affect assessment performance should be excluded or clarified. Based on AUA model, it is necessary to define language ability in a way that is appropriate for each particular assessment situation. As the AUA model emphasizes *each particular assessment situation*, there must be adequate samples for each particular assessment situation, or specific TLU tasks for different purposes and conditions. Consequently, "language ability" needs to be designed for specific purposes to reflect each particular assessment situation, such as email writing and note writing, or chatting over a phone and talking with a professor, or reading a fiction and reading a newspaper article. In fact, all this distinction may not be considered necessary to us ordinary English teachers and test designers because designing a test for each particular case is impractical; and high foreign language proficiency covers such general language ability.

Moreover, according to the AUA test model, Bachman and Palmer (2010) provided a new perspective of language testing development, but as being mentioned above, many concepts of the AUA model are rather debatable and even confusing. They pointed out that test designers should be more interested in test candidates' "language ability" than their test performance (Bachman and Palmer, 2010, p.60, p.115). At first glance, this claims seems to be quite acceptable. However, we should be fully aware that the claimed "language ability" used in Bachman and Palmer's CLA and AUA models differs considerably from the traditional linguistic concept of language ability in terms of foreign language proficiency, which has been widely used by ordinary test designers and teachers. The familiar "language skills" (listening, reading, writing, and speaking) are not the constructs to be assessed according to Bachman's models (Bachman and Palmer, 2010:220). In fact, the concept of "language ability" used in the CLA and AUA test models is much broader than the common concept of language ability. Therefore, this phenomenon is a confusing picture of terminology. To some degree, these test models may send confusing signals to ordinary test designers in test development and validation, because the conception of "language ability" in the AUA model may not carry the same meaning of the common terminology that is widely used by test designers or classroom teachers at grassroots level. In short, it appears that Bachman and Palmer used the same terminology to refer to different concepts of "language ability" for their models of CLA and AUA. Based on the author's practical experience in language testing over the past two decades, it is believed that such a development or divergence of terminology in their test models of CLA and AUA would provide little functional use in practical testing development, although some scholars in the testing circle may have avid interest in the debate of divergence of terminology.

Another issue that is worth considering is the difference between the interests of psycholinguists and applied linguists. We need to notice that what "competence" refers to

according to Chomsky's concept is the ideal speaker/hearer of the native speaker, that is an idealized but not a real person who would have a complete knowledge of the whole language, i.e. mother tongue (Richards and Schmidt, 2002). Meanwhile, the language competence or ability of the four skills (listening, reading, writing, and speaking) used by most of our test designers in their development of a foreign language test refers to the foreign language proficiency in terms of a testee's skill in using this foreign language through a test performance. The point is that these two terms of competence are two different concepts, not to mention the fact that our test takers are all EFL learners, and most of them have already passed the biological "critical period" so that their lateralization has already occurred by this time. Thus, as for psycholinguists, they would be more interested in one's mental language ability and cognitive process. But as EFL teachers and learners, performance of using a foreign language functionally would be their focus of teaching and learning. Therefore, when we come to Bachman and Palmer's term of language ability used in the test models of CLA, TLU, and AUA, we should be fully aware of what the difference could be between their definition of "language ability" and the commonly used concept of *language ability* used by EFL teachers, students, as well as the test designers of foreign language proficiency tests.

So far we have discussed the definition of "language ability" in the AUA model. Based on our discussion above, we can come to a tentative conclusion that "language ability" used by Bachman and Palmer in their notion of AUA has provided a new perspective for language testing. But the definition of "language ability" proposed in the AUA test model differs considerably from our commonly accepted concept of language ability of the four skills. In their words, listening, reading, writing, and speaking are not the main constructs to be assessed according to Bachman's models. Therefore, the definition of "language ability" in AUA (or in its early version of CLA) might be a confusing picture of terminology. To some degree, these test models may send confusing signals to ordinary test designers in test development.

What should test designers consider more: "language ability" or test performance?

In testing practice, language *ability* and *performance* are two sides of the same coin. There is no denying that testing is used as a tool for evaluating the learner's language competence, knowledge, and performance so as to collect information for making decisions, which is the basis for how test validation can be proved. It should not be overemphasized that what test designers should be mainly interested in is a testee's "hidden" language competence. Instead, more attention should be paid to the candidate's actual test performance, i.e. foreign language proficiency. It is the test performance that reflects both directly and indirectly of the test candidate's language competence.

In their discussion about these concepts, Mihalicek and Wilson (2012:6) have provided a simple explanation: "You can think of linguistic competence as a person's unseen potential to speak a language, while linguistic performance is the observable realization of that potential: our performance is what we do with our linguistic competence." Since linguistic competence cannot be observed directly, it is understandable that test designers, stakeholders, and applied linguists have to use test takers' linguistic performance as a basis for making judgement and conclusions. Note that psycho-linguists are generally interested in how linguistic competence works rather than how to assess the way and function of such competence works, which may differ considerably from what many language test developers or designers are interested in. At this point, let us take a look at Bachman's view on assessing constructs, or *traits* in his words. Bachman (1990:309) stated:

> Language tests, however, like all mental measures, are *indirect* indicators of the abilities in which we are interested in. That is, we are not generally interested so much in how an individual performs on a particular test task in a specified setting on a given day, as in what that performance indicates about that individual's ability. The identification of trait with performance, however, does not permit us to make

inference beyond the testing context, and thus severely limits both the interpretation and use of test results and the type of evidence that can be brought forth in support of that interpretation or use.

In Bachman's view, what test designers should be interested in is test candidates' language competence rather than their test performance. However, this is a doubtful assertion as this view is much based on the concept of psycho-linguistic cognitive theory, and little empirical evidence is provided in support of this claim, not to say empirical evidence from large-scale proficiency tests. In this sense, Bachman's basic test model of Communicative Language Ability (CLA) was criticized as "it represents an essentially psycholinguistic view of performance and is largely missing important interactional and sociolinguistic elements" (O'Sullivan and Weir, 2011:14-15).

Moreover, let us consider this issue of competence and performance from the view of classical test theory so as to see why the separation of real language ability from other factors or attributes in the models of CLA and AUA is unrealistic or unfeasible in practice. Hypothetically, it is believed that a person's language performance is a kind of reflection of one's language competence plus some other factors. In other words, a test candidate's performance contains two parts: the candidate's language competence, and other interfering factors which affect the true score of the candidate. Still, in terms of classical test theory, a candidate's test score is not his true score of language ability. Instead, one's observed test score consists of true score plus some measurement error. Therefore, some people suggested that what test developers should be really interested in is how a test can demonstrate a test candidate's true competence, i.e. the "hidden" ability, rather than one's language performance. However, the concept of "true score" in classical test theory does not necessarily mean that language performance can be separated from "competence". In fact, it is through language performance that test reliability and validity can be actually achieved so as to prove the validation of tests. In fact, the concept of "hidden" *ability* poses a challenging question for classroom teachers, test users and stake-holders about how they can focus on language *ability* in practice. Hughes (2003:40) has correctly pointed out, "It is this score, which for obvious reasons we can never know for certain, which is referred to as the candidate's *true score*." Similarly, paper holds that such a theoretical concept of "hidden" ability is far from the reality, because we can never really observe a person's true score in testing practice, but only his/her performance, i.e. an ***observed score*** (Hughes, 2003:40; Davis et al., 1999: 215). Therefore, efforts should be made in reducing mitigating factors in candidates' testing performance rather than focusing on test competence psycho-linguistically.

In summary, as for the relationship between language ability and test performance, the so-called language ability viewed from psycho-linguistic perspective is similar to the linguistic competence in Chomsky's view of transformational-generative grammar, but the latter is mainly concerned with the LAD (language acquisition device) of the L1, which differs greatly from foreign language proficiency. In this sense, Bachman and Palmer's test models of CLA or AUA adopted a psycho-linguistic view of their definition of language ability, which would be difficult to be carried out in testing practice. Therefore, this paper suggests that test performance is to be the focus, i.e. the four skills of listening, speaking, reading, and writing, because test performance is the reflection of language ability.

Empirical Studies of Linguistic Competence and Language Performance

If we accept Bachman and Palmer's concept of language competence (2010:33), we may assume that testees' linguistic knowledge might be assessed so as to measure their competence in L2. In addition, testees' performance of L2 proficiency test should be closely related to their language competence. Therefore, for the purpose of this paper, an empirical analysis of test scores of 78 sophomore English majors at a Taiwanese university was

conducted. All subjects took a test in English Linguistics at the end of the semester, which is their required course. Within a week, the same group of students took an English proficiency test (in Listening and Reading). By using SPSS (V20) to conduct a Paired-Sample T Test, and Pearson test, we obtained the following information: 1) students' test scores of English linguistics differ greatly from their scores of language proficiency, as T value is -4.033, p= .000 < .05. Meanwhile, the Pearson test of such two kinds of test scores shows that there exists no correlation between the scores of English Linguistics and English proficiency performance (r = -.029, p = .807>.001). Although this is a tentative empirical study, we can still obtain some implication that one's linguistic competence or knowledge of L2 may not be a valid reflection of one's real ability to perform L2 in reality. To some extent, this finding consolidates the author's claim that the test validation should be more focused on the observable test performance, i.e. language proficiency, rather than the psycholinguistic "competence or ability".

Discussion

It is important for test designers to consider test usefulness, or validation, from the viewpoints of the end users and stakeholders so as to determine what can be considered as an appropriate approach to test development. Test designers should focus on developing assessments whose use can be justified to stakeholders.

We should be aware of the difference between the interests of psycholinguists and applied linguists. It is important to notice that what "competence" refers to according to Chomsky's concept is the ideal speaker/hearer of the native speaker, an idealized but not a real person, who would have a complete knowledge of the whole language, i.e. mother tongue (Richards and Schmidt, 2002). Actually, no such kind of person really exists as no one can be perfectly knowledgeable and proficient in using his native language. Seeking such kind of "language ability" could be the matter of researchers' theoretical avid interest, but not the interest of test practitioners in foreign language test development and implementation. What we test stakeholders are interested is a different concept of language ability, i.e. the foreign (not mother tongue) language proficiency or ability of the four skills (listening, reading, writing, and speaking) in terms of a test taker's skill in using this foreign language. Therefore, when we come to the AUA framework, we should be fully aware of Bachman and Palmer's term of "language ability" used in their test models.

Since the "language ability" used in Bachman and Palmer's model (2010) differs considerably from the familiar or traditional linguistic concept of "language skills" (listening, reading, writing, and speaking), how to combine the concept of the AUA model with practical testing in reality remains a tough question for grass-roots test designers and ordinary teachers to answer. If "language ability" refers to one's hidden language competence plus "other attributes", which are virtually not observable, the function of the above-mentioned AUA test model may remain unclear due to its lack of practicability. In other words, it seems impractical for test designers and users to adopt such test models in their practical development of a testing system and administration. Nevertheless, as Bachman and Palmer (2010, 132) themselves reminded that the AUA test model should not be treated as a prescriptive standard.

It is suggested that test designers and EFT teachers may choose the best tool in their own test development and practice. Meanwhile, we should be aware that linguists are generally interested in how one's native linguistic competence works as a mental process rather than how to assess one's proficiency level like that in a foreign language test. This is very different from what foreign language test developers or designers are interested in. In other words, what we need is not blind application of new test models or frameworks, nor is eclecticism, but judgement. As mentioned by Bachman and Palmer (2010:60), anyone interested in the AUA should be fully aware that the test developer should "identify and describe a specific TLU domain and select one or more TLU tasks as a basis for developing assessment tasks."

It is important to be aware that Bachman and Palmer's psycholinguistic concept of "real language ability" seems be a mission impossible for ordinary testing practitioners. The concept of "hidden" *language ability* poses a challenging question for classroom teachers, test users and stake-holders about how they can focus on such ability in practice when they adopt the AUA model. Instead, what we can obtain is testees' performance, i.e. observed scores. Therefore, efforts should be made in reducing mitigating factors in candidates' testing performance rather than focusing on test competence psycho-linguistically so as understand the mental cognitive process of linguistic decoding.

Conclusion

In the circle of language testing, the relationship between language ability and language performance is a fundamental issue for test designers and users to consider in their theoretical framework of test development and validation. Different perspectives of testing validation would lead to distinctive solutions for dealing with test usefulness in both theory and practice. It is suggested that the definition of "language ability" in the AUA model should be used with caution.

In conclusion, it is a very doubtful assertion that foreign language test designers should be mainly interested in "language competence" in Bachman's (or Chomsky's) terms rather than in language proficiency or performance. As designers of tests of English as a foreign language, we are actually more interested in test candidates' observable communicative language performance rather than their "hidden" language competence. Unlike psycholinguists, who may be more interested in how "linguistic competence" works, foreign language test developers should be more concerned with how to assess test candidates' performance effectively and validly so as to serve the purpose of the English language tests under our sociolinguistic conditions. Then, test performance is the valid reflection of one's language ability.

References

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: OUP.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: OUP.

Bachman, L. F., & Palmer, A. S. (2010). *Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World*. Oxford: OUP.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1-47.

Carr, n. T. (2011). *Designing and Analyzing Language Tests*. Oxford: OUP.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Mass.: The MIT Press.

Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: CUP.

Kane, M. (2011). Book review: Language assessment in practice: Developing language assessments and justifying their use in the real world. *Language Testing,* 28(4), 581–587.

Lado, R. (1961). *Language Testing*. London: Longman.

McNamara, T. (2003). Looking back, looking forward: rethinking Bachman. *Language Testing*, 20(4), 466–473.

Mihalicek, V., & Wilson, C. (Eds.). (2012). *Language Files* (11 ed.). Taipei: Bookman.

O'sullivan, B., & Weir, C. (2011). Test development and validation. In B. O'sullivan (Ed.), *Language testing: theories and practices* (pp. 13-32).

Richards, J., & Schmidt, R. (2002). Longman Dictionary of Language Teaching & Applied Linguistics (3rd Ed.). London, Pearson Education Ltd.

Saussure, F. d. ((1916) 1966). *Course in general linguistics*. New York: McGraw Hill.

Weir, C., & Shaw, A. (2006). Defining the constructs underpinning Main Suite Writing Tests: a socio-cognitive perspective. *Research Notes, 26*(November), 9–14.