One of the issues that has been the focus of extensive thought, debate and research in language testing is the validity of tests. A test, or part of a test, is said to have validity if it can be demonstrated that it measures just the ability that it is supposed to measure (Hughes, 1996). The language skills or sub-skills that tests attempt to measure are not directly observable, but must be elicited by an observable tool (test). The decision on an individual's ability depends on the information gathered or the competence of this tool. However, independent factors, which are not supposed to be involved in the ability measured, may obstruct or contribute to the information required by this tool. Nevo (1989) states that a review of the literature on language testing suggests that test makers' assumptions concerning what they are testing often do not coincide with the actual processes undergone by the respondent when taking the test. However, in exploring validity of tests, the test takers' perspective has been somewhat ignored. This study, therefore, may elicit vital information in this regard.

McDonough (1995) addresses two questions on test validity and suggests they can be answered by looking at test-taking strategies:

1. How well does the test actually measure what it is supposed to measure?
2. Second, how good a theory of reading (or whatever) is the tester's construct? In other words, how well does the tester's construct match developments in theorizing about the skill involved outside testing?

Recently several authorities have suggested that one way of answering these questions is to look at the strategies of test takers using the think-aloud method (McDonough, 1995: 107).

An example that provides a response to the first question, from a test-taker's perspective, has been noticed by the present writer when he sat a real IELT. In the listening comprehension section in the that test, in some of the test items each of the four multiple-choice options was about three quarters of a line in length. Since the time provided for the response was very short, answering such items, in fact, needed speed-reading competence.
as well as listening comprehension. In other words, some examinees had comprehended
the given listening material very well, but they were unable to give the right responses
because the written items, which supposedly tested listening comprehension, actually I
measured a second skill at the same time. This criticism is in line with Cronbach's (1980)
assertion that "the job of validation is not to support an interpretation, but to find out what
might be wrong with it" (cited by Bachman, 1990:257).

With regard to McDonough's second question, experts (e.g. Dolly & Williams, 1986;
Allan, 1992; Yi'an, 1998) argue that when students are taking a test - any test - they are
really being tested on two things: (a) how much they know about a subject, and (b) how
adept the individual is at taking tests. If this is true, then the test may fail to discriminate
between the test takers' ability in the tested area and may also fall short of measuring the
extent of the test takers' knowledge in the given task. For example, in some tests, which
involve speed as a factor test takers may vary in test performance due to variation in
time management strategy, rather than varying degrees of linguistic knowledge. Complex
test instructions or formats could also be problematic for some students. Furthermore, in
test construction, various clues are sometimes involved in tests, which may lead some test
takers to select the correct answer even though they do not actually know the knowledge
tested. For example, in a filling in the gap problem, they may analyse the context of the
gap and use some methods (test-taking strategies) to fill it, which may be quite
independent of the intended knowledge needed to fill it. Such test-taking strategies bypass
the knowledge of the subject matter which the test items aim to measure and subsequently
affect the validity of examination, scores. Research (Brozo et al., 1984; Carter, 1986) has
shown that teachers and college instructors are largely unaware of the extent of flawed
construction appearing in teacher-made tests; possibly providing an unfair advantage to
the student who is good at test-taking strategies. Also, others (Slakter, et al., 1970; Dolly
and Williams 1986) assert that some individuals have better test-taking strategies than
others, which may result in test score variations and may weaken the validity of the test.

Although the problem of possible influence of successful test-taking strategies on a test's
validity is applicable to all types of tests, the multiple-choice question is particularly
susceptible to such criticism. Multiple-choice questions do not provide evidence of the
cognitive processes by which alternatives are selected or rejected. A test taker may pick
an incorrect answer based on correct reasoning, perhaps by thinking of reasons not
anticipated by the test maker. On the other hand, a correct answer may be picked by applying certain strategies rather than language knowledge. Some research suggests that knowledge of test taking strategies tends to earn students higher scores for multiple-choice tests than scores on examinations of other types. Kesselman and Peterson (1981) claims that the multiple-choice test "is the type of exam that rewards 'test-wise' students with the most extra points for their 'test-wiseness'." (Kesselman and Peterson, 1981: 36). Cohen (1984) mentioned findings that support this claim. He reported that students were given questions of 4 multiple-choice options, without the passage to which they were supposedly related. The random possibly for selecting a correct answer in such a situation is 25%. However, the subjects scored nearly twice that. This suggests they used some independent strategies rather than just random ticks. With regard to the case of the current study, test items have appeared to require particular knowledge of lexical meaning to fill in a gap. However, subjects may have used quite independent strategies to make a choice for the gap based on elimination strategies, grammatical knowledge, collocations, morphological elements, positioning sense or a whole range of other test-taking strategies involving other kind of knowledge than the one targeted.

All in all, test's validity is subject of a good deal of concern; yet, one angle that has been largely neglected in this issue is the test-takers' perspective. Without exploring the processes evoked when they tackle multiple-choice questions, or any type of test format, test theorists are hindered in their attempt to verify and increase validation. Thus, content validation needs to be extended beyond the test maker's 'expectation', analysis of the items, and even beyond a panel of expert's judgment of the validity of items, to include getting students to do the items, tapping their process and analysing them to see what sort of knowledge really is being used to answer the items and hence how valid the items are.

References


