DOCUMENT RESUME

ED 368 191 FL 021 949

AUTHOR Scholz, George E.
TITLE Exploring ESP and Language Testing.
PUB DATE Apr 93
NOTE 8p.; Paper presented at the Annual Meeting of the Southeast Asian Ministers of Education Organization Regional Language Center Seminar (28th, Singapore, April 19-21, 1993).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Construct Validity; Content Validity; *Curriculum Design; *English for Special Purposes; *Language Tests; *Second Languages; *Test Construction; Testing; *Testing Problems; Test Validity

ABSTRACT

A discussion of language testing in the context of a program in English for Special Purposes (ESP) focuses on the lack of "fit" between the two areas and makes some recommendations for improvement. It begins with overviews of recent trends in testing and recent issues in ESP. Overlap is seen in two areas: construct and content validity. It is argued that for learning, teaching, and testing to be in harmony, test specification and test items should be derived from the same need analysis and program consensus used for developing ESP instructional materials, but little research has been conducted on this issue, or indeed in ESP testing in general. It is proposed that the structure of the original need analysis may need to be reviewed and consensus-building be undertaken to increase the program's pedagogic coherence and operational cohesion. Program evaluation and improvement would subsequently benefit, and both instructional materials and tests could be made more appropriate. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

EXPLORING ESP AND LANGUAGE TESTING

BY GEORGE E. SCHOLZ

ED 368 191

U S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

George E. Scholz

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

FL021949

BEST COPY AVAILABLE

Over the last twenty-five years, both English for Specific Purposes (ESP) and language testing have made significant individual contributions to English language teaching (Bachman, 1991; Johns & Dudley-Evans, 1991). ESP has enriched the areas of needs assessment and discourse analysis, while language testing has enhanced our understanding of communicative competence by investigating models of language proficiency. However, the interaction between ESP and language testing has been rather limited. While dealing with issues of the components of language proficiency and task-based testing, the validation of communicative testing techniques or criterion-referenced testing in ESP has been slow. While ESP programs are empowered with needs analysis and materials development, the testing and evaluation of such programs are relatively weak, evoking general standardized language proficiency (concurrent validity) as proof of a successful program of instruction. Perhaps this paucity of correspondence between ESP and language testing is a result of their different focal points as specializations in the world of language learning. Alternatively, perhaps this general lack of "fit" between ESP and language testing may be due to both a lack of empirical rigor and professional consensus on the various validities implied by an applied ESP program model. This paper is a brief exploration of the issues and implications of language testing in an ESP context.[1] For the purposes of this paper, Language for Specific Purposes (LSP) can be substituted for English for Specific Purposes (ESP). LSP is more global in focus, and the issues raised here are germane to any language program.

Language testing in the last ten years has focused on three areas (Bachman 1991 672-690), with implied issues of validity:

1) Theoretical issues of the nature of language proficiency and their implications for practical evaluation (construct validity),
2) Methodological advances in terms of tools for test analysis (content validity), and,
3) Language test development as 'communicative testing' (predictive & concurrent validity).

The nature of language proficiency leaves much room for discussion. Testing researchers concur that language proficiency cannot be viewed as a unitary global ability (as proposed by John Oller in 1979). Language proficiency is "mulitcomponential, consisting of a number of interrelated specific abilities as well as a general ability or set of general strategies or procedures" (Bachman 1991:637). Language testing is attempting to answer the perennial questions of what language is and how it is learned. Construct validity addresses the theoretical justification of a test, which in many ways is similar to a classroom language teacher's approach, or teaching /learning beliefs.

Statistical tools and research methodologies assist in identifying learner, test, and task characteristics. Content validity is involved here, where the issue is whether the test indeed measures the skill(s) that the test is claiming to measure. Identifying the components, abilities, strategies and procedures of language proficiency (competence), however, is a more complex task, and an area currently under investigation. Testing experts constantly try to develop a proposed model of language proficiency (construct), in turn justifying enhanced test development (content). What language testing researchers do know is that "a language test score represents a complexity of multiple influences" (Bachman 1991: 676). The classroom is very similar, where

[1] This paper is a revision of "ESP and Language Testing" in the TESOL France News, Vol.XXII, No. 2, Summer, 1992.

the instructor's or program's approach (construct) determines the activities for learning materials (content).

Concurrent and predictive validity can be statistically measured. Concurrent validity examines the relationship of skills in one test to similar skills in another. For example, if a dictation correlates highly with a listening mulitple-choice test, then these two components can be related to a hypothesized 'listening' construct. If the dictation test, does not correlate significantly with a spelling test, then the two components (dictation to spelling) are not necessarily related. Consequently, one cannot claim that a dictation test captures spelling test or vice-versa. Predictive validity explores the future. For example, assuming all other variables are controlled, if on-the-job performance can be statistically linked to prior language learning history, then there is a good case for the predictive validity of the language tests administered.

There are conceptual parallels between teacher and tester: both are language teaching professionals concerned with the nature of language and language learning. Construct validity is related to the question of whether posited language abilities in a test are related to the language abilities targeted for a learner. Content validity deals with the question of whether the language test tasks are similar to the tasks for the targeted language use (cf. Bachman 199: 681, Oller 1979: 50). These issues in language testing are alive and well. ESP should be concerned with these issues of construct and content validity, as the same needs analysis which generated syllabus and resultant materials should also be used to assist ESP test developers in creating test specifications and test items in sync with the syllabus/materials design. Moreover, communicative language testing is concerned with what the learner needs to do with the language, as specified by the target language use. The same issues for ESP materials design are also concerns of language test writers: that the materials/test items be practical, cost-effective and have instructional value.

Let's now examine current issues in ESP. A robust ESP program should have appropriate materials which facilitate learning and appropriate tests which facilitate decision-making and support instructional delivery. Of course, this is in the ideal world. The actual issues in ESP center on three questions (Johns 1991: 304-305), which also revolve around issues of validity:

1) How specific should ESP courses and texts be (content validity)?
2) Should they focus upon one particular skill, e.g. reading, or should the four skills always be integrated (construct validity)?
3) Can an appropriate ESP methodology be developed (content validity)?

There is overlap between issued in ESP and language testing, and they revolve around construct and content validity. For example, if an ESP program decides (and the needs analysis indicates) that all four of the traditional language skills should be integrated, how does this impact on ESP test design? How can skill-integration be broken down into language abilities (construct validity)? How do integrated language tasks in the classroom translate into content-valid testing tasks with instructional value? Does a unique ESP methodology evoke specific language abilities and tasks? Has the ESP program's content been validated by subject matter specialists? J. C. Alderson (1988) provides some perceptive insights into the relationship of ESP, needs analysis and language test development.

ESP's concern with appropriacy of materials relevant to a learner's targeted need should focus on the construct and content validity of a program's testing instruments. Test specification and test items should be derived from the same needs analysis and program consensus used for developing ESP materials. This way, the learning, teaching and testing troika could be placed in harmony. Language testing not only should provide valid and reliable information for decision-making, but also have instructional value (Oller 1979: 50-52) or a positive "washback" (a term used by J.C. Alderson). As common-sensical as this may seem, little research has been done with respect to the instructional value of tests and their tasks. Even more amazing is the lack of agreement, spurious assertions, unverifiable claims, and scarcity of empirical data for many tests which determine many learners' futures! On the other hand and in all fairness, one only has to sit in on a faculty meeting to hear similar issues concerning the classroom and program learning environment.

Resolving these issues is no easy matter. The language program must go on, the teachers must teach, the students must take tests. However, I would like to propose that a closer structural look at needs analysis and a programmatic attempt at consensus building with consequent goal-setting may bring ESP and the language testing together again.

INSERT FIGURE A. ABOUT HERE.

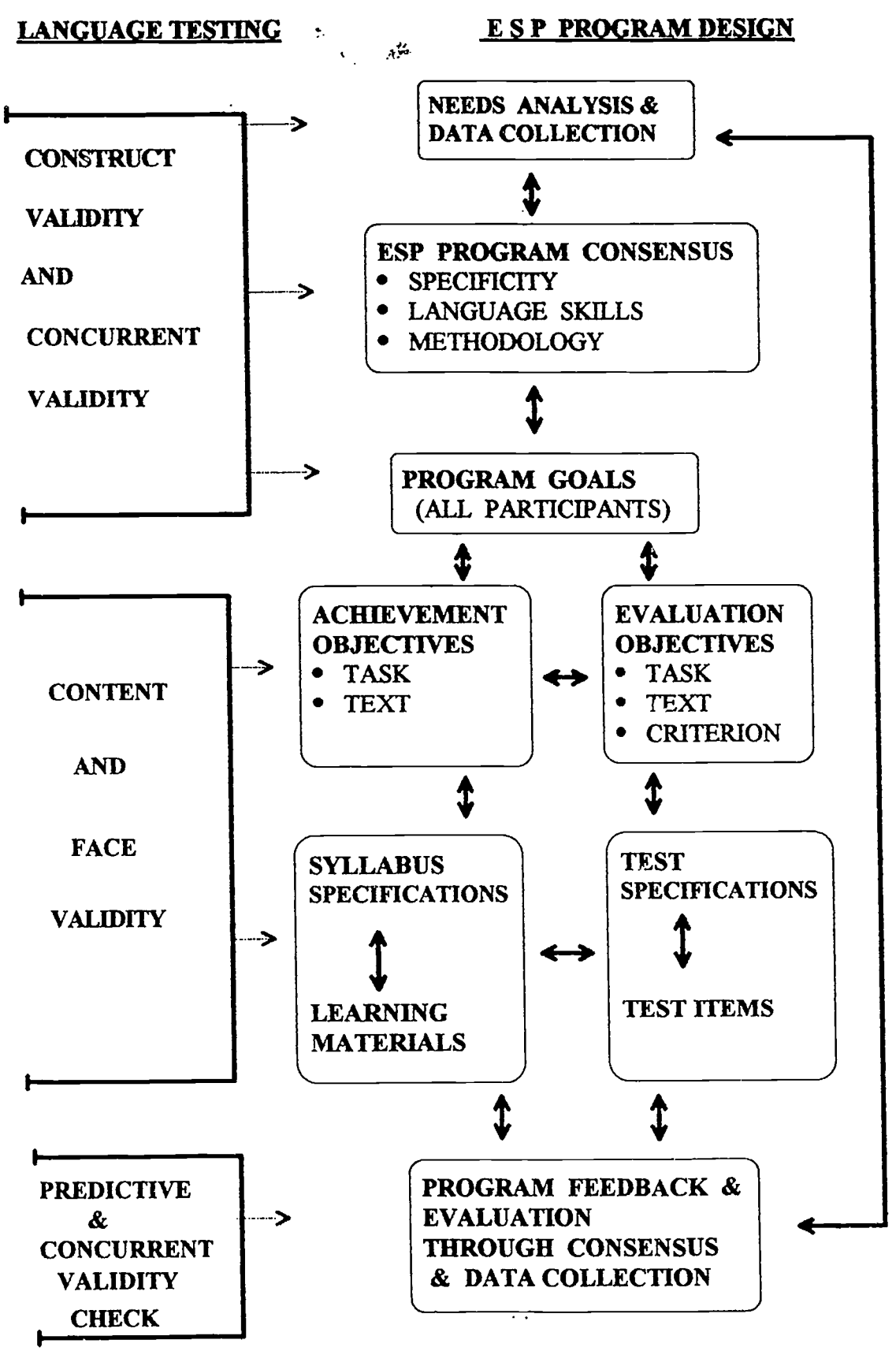
These questions of validity from both the testing and ESP perspective, however answered, will most likely generate discussion in any language program. The structure of the original needs analysis may need to be reviewed for areas of agreement and divergence among client, instructors, subject matter specialists, and the learners. All participants in an ESP language program should reach a consensus in attempting to resolve these three issues. Consensus-building in the part of client, administration, faculty and learners should increase a program's pedagogic coherence and operational cohesion. Once consensus (with the needs analysis as a starting point) is achieved, with coherence reflected in the goals and objectives, the building of operational cohesion (derived from the needs analysis) in test and syllabus specification can begin. In turn, test items and learning materials are interactive with their respective specifications. Program feedback and evaluation again gains strength through consensus and discussion, with a feedback loop to all the elements in a language program design. Truly effective and efficient language programs require revision through consensus and data collection.

These are questions which will generate much discussion among ESP practitioners and testing specialists alike. Unfortunately, there are no clear answers. The proposed ESP program model attempts to show where testing and related issues of validity may be incorporated into an ESP program.

Are there and ESP programs which have appropriate (valid) tests but inappropriate materials? Are there any ESP programs which have appropriate (valid) materials but inappropriate tests? My personal experience suggests that there are more of the latter. Perhaps this is because the appropriacy of materials is less constrained by the strength of the needs analysis, while a test is more subject to the rigors of empirical verification. Alternatively, given the decision-making power embedded in tests, the evaluation and analysis of language proficiency or achievement is

avoided by an ESP program by using standardized norm-referenced tests from an outside source to establish concurrent validity. Thus the entire issues of construct and content validity could be side-stepped.

This short article has attempted to focus language testing in an ESP program model. While most models are linear in theory, an ESP program model should be considered highly interactive in reality. In other words, while the proposed model offers a sequential view, the actual operationalization of the model involves interwoven interactions, indicating that the model should be considered dynamic. The human element cannot (and should not) be itemized by modular components. The collection of data for a language program, its transformation into useful information and consequent application is dependent on the consensus of the professionals involved in that program. How they agree to use the information reflects the learning effectiveness and operational efficiency of their program..


LANGUAGE TESTING
E S P PROGRAM DESIGN
NEEDS ANALYSIS & DATA COLLECTION
CONSTRUCT VALIDITY AND CONCURRENT VALIDITY
ESP PROGRAM CONSENSUS
• SPECIFICITY
• LANGUAGE SKILLS
• METHODOLOGY
PROGRAM GOALS (ALL PARTICIPANTS)
CONTENT AND FACE VALIDITY
ACHIEVEMENT OBJECTIVES
• TASK
• TEXT
EVALUATION OBJECTIVES
• TASK
• TEXT
• CRITERION
SYLLABUS SPECIFICATIONS
LEARNING MATERIALS
TEST SPECIFICATIONS
TEST ITEMS
PREDICTIVE & CONCURRENT VALIDITY CHECK
PROGRAM FEEDBACK & EVALUATION THROUGH CONSENSUS & DATA COLLECTION

Bibliography

Alderson, J. Charles. (1988). New procedures for validating proficiency tests of ESP? Theory and practice. Language Testing, 5 (2), 220-232.

Bachman, Lyle F. (1991). What does language testing have to offer? TESOL Quarterly, 25 (4), 671-704.

Johns, Ann M. & Tony Dudley-Evans. (1991). English for specific purposes: International in scope, specific in purpose. TESOL Quarterly, 25 (2), 2997-314.

Oller, John. (1979). Language tests at School. London: Longman.