

to overcome the limitation of a simple comparison of MTMM correlations, as CFA models are a powerful and direct means to test the relative contributions of traits and methods to test-takers' performance and to explain underlying relationships.⁵⁵The approach used in measuring the construct validity of the test is very different from this research. In this research, the researcher uses an exploratory factor analysis in term of measuring the construct validity.

The second is a journal entitled "The Construct Validity of a Test: A Triangulation of Approaches" by Mohammad Salehi at Sharif University of Technology. In his research he uses three different kind of approach in determining the construct validity of the test. The first approach is factor analysis, the researcher use a factor analysis in investigating whether the test items in the 'Reading Comprehension' sections of the UTEPT distinctly measure various sub-skills or not. The result shows that this test is good in measuring various sub-skills of reading. The second is multitrait-multimethod approach which use in finding the correlation between traits being tested and the methods of testing. The result of this approach shows that there are correlation between traits being tested and the methods of testing. The last approach is protocol analysis. The use of this protocol analysis is to distinguish the degree of correlation among sub-parts of the test. The approach use in his journal is different from the approach will be used in this research. As the researcher has said above, this research will be investigated using factor analysis.

⁵⁵ Lyle F Bachman, *Statistical Analyses for Language Assesment* (Cambridge: Cambridge University Press, 2004) , 279.

a Foreign Language (TOEFL), a semantic schema-based cloze multiple choice item test (S-Test) and a lexical knowledge test (LKT). The research focus on finding what factor structure is for the C-Tests, the Decontextualised C-Test, the Spelling Test and what factor structure is if the TOEFL, the lexical knowledge test and S-Test are included. The application of principal component analysis (PCA) to the responses of the participants on the three tests, i.e., C-Tests, Spelling Test and Decontextualised C-Test, revealed two components called *language proficiency* and *direction specificity* in this study. While the inclusion of the S-Test, TOEFL and the LKT in the PCA yielded the same two components, their rotation brought about the highest loadings of the included tests as well as the moderate loadings of the C-Tests on the first component, validating them as proficiency measures of language. However, they loaded the highest on the second component along with the Spelling Test and Decontextualised C-Test and thus confirmed their spelling and direction specificity. The difference between the study of C-test and this research is the focus of research. This research focuses on examining the construct validity of TOEFL-like test using factor analysis procedure.

The sixth is a study by Neil J. Anderson, Lyle Bachman, Kyle Perkins and Andrew Cohen entitled “An exploratory study into the construct validity of a reading comprehension test: triangulation of data sources”. Reading, while recent thinking in language testing has recognized the importance of gathering information on test taking processes as part of construct validation. And while there is a growing body of research on test-taking strategies in language testing, as well as research into the

relationship between item content and item performance, no research to date has attempted to examine the relationships among all three - test taking strategies, item content and item performance. This study thus serves as a methodological exploration in the use of information from both think-aloud protocols and more commonly used types of information on test content and test performance in the investigation of construct validity. The study is examining the construct validity of the reading comprehension test by investigating the relationship among test taking strategies, item content and item performance.⁵⁸ In investigating the construct validity of TOEFL-like test, the researcher will only use one data resource, test item. Types of data analysis were performed, providing both qualitative and quantitative data. Each data analysis task reflects part of the triangulation of data sources to examine the construct validity of the reading comprehension test. The first task involved reviewing each of the think-aloud protocols and coding them for the use of reading and test-taking strategies. The second data analysis task involved the content analysis of each test item on both forms of the test from two perspectives: that of the test designer and of Pearson and Johnson's (1978) taxonomy of relationships between texts and test items. Third, the data from the think-aloud protocols were then submitted to chi-squared analyses. The result of chi-square analyses indicate that these subjects are using these strategies differently, depending on the type of question

⁵⁸ Neil J. Anderson, Lyle Bachman, Kyle Perkins and Andrew Cohen, "An exploratory study into the construct validity of a reading comprehension test: triangulation of data sources". *Language Testing* , Vol.8 No.2, 1991, 41.

that is being asked. The results of the chi-square analysis also show that the subjects tended to use some strategies fairly consistently for the different types of items.

The seventh is a journal entitled “Testing the Convergent and Discriminant Validity of the Decisional Balance Scale of the Transtheoretical Model Using the Multi-Trait Multi-Method Approach” by Boliang Guo, Paul Aveyard, and Antony Fielding of University of Birmingham and Stephen Sutton of University of Cambridge. The authors extended research on the construct validity of the Decisional Balance Scale for smoking in adolescence by testing its convergent and discriminant validity. Hierarchical confirmatory factor analysis multi-trait multi-method approach (HCFA MTMM) was used with data from 2,334 UK adolescents, both smokers and non-smokers. They completed computerized and paper versions of the questionnaire on 3 occasions over 2 years. The results indicated a 3-factor solution; Social Pros, Coping Pros, and Cons fit the data best. The HCFA MTMM model fit the data well, with correlated methods and correlated trait factors. Subsequent testing confirmed discriminant validity between the factors and convergent validity of both methods of administering the questionnaire. There was, however, clear evidence of a method effect, which may have arisen due to different response formats or may be a function of the method of presentation. Taken with other data, there is strong evidence for construct validity of Decisional Balance for smoking in adolescence, but evidence of predictive validity is required. The difference between the research above and this research is that this research use factor analysis procedure in case of measuring the

construct validity of this TOEFL-like test while the research use MTMM in case of measuring the discriminant validity which is the part of construct validity.

The eight is a journal entitled “Does the text matter in a multiple choice test of comprehension? The case for the construct validity of TOEFL’s minitalks” by Roy Freedle and Irene Kostin. The current study addresses a specific construct validity issue regarding multiple choice language-comprehension tests by focusing on TOEFL’s minitalk passages: Is there evidence that examinees attend to the text passages in answering the test items? To address this problem, we analysed a large sample ($n = 337$) of minitalk items. The content and structure of the items and their associated text passages were represented by a set of predictor variables that included a wide variety of text and item characteristics identified from the experimental language-comprehension literature. Stepwise and hierarchical regression techniques showed that at least 33% of the item difficulty variance could be accounted for primarily by variables that reflected the content and structure of the whole passage and/or selected portions of the passage; item characteristics, however, accounted for very little of the variance. The pattern of these results was interpreted, with qualifications, as favoring the construct validity of TOEFL’s minitalks. Our methodology also allowed a detailed comparison between TOEFL reading and listening (minitalk) items. Several criticisms concerning multiple-choice language-comprehension tests were addressed. The difference between this research and the research which the researcher worked for is the method in examining the construct validity. As the researcher has said above, this research uses regression in case of

