Validation of the Gap-fill as an Overall Measure of English Language Proficiency among Iraqi EFL Learners

Ahmed Sattar

English Department, University of Thi Qar / college of Arts

ahmedsattar@utq.edu.iq

https://orcid.org/0009-0002-5684-3501

Abstract

The purpose of this study was to validate the Gap-fill among Iraqi EFL university students. To this end, two types of multiple-choice Gap-fill were administered to a sample of EFL students along with a number of other tests as criteria. Correlational analyses revealed moderate to strong correlations between the Gap-fill and the criterion measures. Principal components analysis resulted in a one-factor solution with the Gap-fill highly loading on this single factor. These were interpreted as validity evidence for the. Gap-fill Implication of the study for classroom testing are discussed.

INTRODUCTION

Preparation of tests can be an immense challenge for EFL teachers, since many of them do not receive professional language testing instruction. Therefore, many teachers have a technique for "cut and paste", i.e., copying items from existing tests or classroom lessons and workbooks when they require a test (Coniam, 2009). That may appear to be a shortcut for teachers who struggle to reach targets for testing in the context of certain commitments. This strategy is considered to be legitimate if the content and complexity of the selected tests are sufficient for the curriculum (Coniam, 2009).

An alternative strategy is to develop tests based on the reduced redundancy principle (RRP, Spolsky, 1968). RRP stipulates that languages contain redundancy, i.e., languages contain many elements that are not essential for communication. Redundancy exists in languages to protect them against noise. This is the reason why we can understand a text with some parts washed away in the rain. Or we can understand our friend in a very loud and noisy party. Languages have many repetitive or redundant parts that if we miss one element there other elements that compensate for the missing parts. The RRP has been employed to develop language tests. In RRP tests, noise is introduced into authentic written or spoken language and examinees are required to process the language with the noise. The logic is that if examinees are proficient in the language, they should be able to understand the language when noise is imposed.

Perhaps the most well-known, most widely-used RRP test is the Gap-fill. In Gap-fill, noise is introduced into language by deleting every n^{th} word (n is usually between 5 and 7) in an authentic text. Examinees have to read the text and restore the missing words. The purpose of this study is to validate a multiple-choice Gap-fill among Iraqi English as a Foreign Language university students. This is the first study on the validation of the Gap-fill among Iraqi students.

METHOD

Participants and Setting

In order to collect the required data, 157 Iraqi undergraduate EFL (English as a Foreign Language) students at al Baghdad University were recruited. The participants were all non-native speakers of English whose first language was Arabic. The participants included 95 females and 62 males. They ranged between 19 and 25 years in age (M=22.40, SD=3.67).

Instrumentation

In order to collect the data for the purpose of the current study, the *Iranian National University Entrance Examination* (INUEE) for admitting candidates to English programs in state universities was used. The 70-item multiple -choice test was administered to Iraqi EFL university students. Version 1398 (2019) of the INUEE was used for this study. The test contained four parts including vocabulary,

grammar, Gap, and reading comprehension. Previous research has showed that grammar and vocabulary are important components of language proficiency (Aryadoust & Baghaei, 2016). Therefore, due to the difficulty of testing listening comprehension and speaking, vocabulary and grammar were selected as criterion measures along with reading comprehension.

RESULTS

Correlational analysis

Table 1 depicts the correlations between the cloze test and the criterion measures and also the correlations between the criterion tests.

Table 1: Matrix of correlations between the tests

	Gap Com.	Gap Wrt.	Gap Spk.	Gram.	Vocab.	Reading	Proficiency
Gap Com.	.83	.92	.90	.70	.60	.68	.78
Gap Wrt.		.77	.67	.66	.63	.70	.78
Gap Spk.			.63	.62	.46	.53	.63
Gram				.70	.55	.56	.81
Vocab					.71	.65	.84
Reading						.83	.90

Note: All correlations are significant at the 0.01 level (2-tailed); Gap Wrt.= Gap-fill based on written discourse;

Gap Spk= Gap-fill based on spoken discourse; Gap Com.=Combined scores of written and spoken Gap-fill. <mark>Cronbach's alpha reliabilities on the diagonal.</mark>

As Table 1 shows, performance on the combined Gap-fill is strongly and significantly correlated with the grammar test ($r=.70$, $p<.01$, $n=156$), the vocabulary test ($r=.61$, $p<.01$, $n=156$), and the reading comprehension test ($r=.68$, $p<.01$, $n=156$). The Gap-fill is highly correlated with the scores of vocabulary, grammar, and reading comprehension <mark>combined</mark> which is named proficiency in Table 2 ($r=.78$, $p<.01$, $n=156$). This combined score is a more accurate measure of general language proficiency.

Cronbach's alpha reliability showed that the Gap-fill is highly reliable with a Cronbach alpha reliability estimates of .83. For determining the validity of the Gap-fill criterion-related validity was used. Table 1 shows the correlations between the Gap-fill and the other tests of language ability which were used as criterion measures. The relatively high correlations found between the Gap-fill and the standard tests support the validity of the Gap-fill as an overall fill of general language proficiency. Furthermore, the scores on grammar, vocabulary, and reading comprehension were combined into a composite score of general language proficiency. The Gap-fill had the highest correlation with this combined measure ($r=0.78$, $p<0.01$) which is an indication of the validity of Gap as an overall language ability fill.

Principal Components Analysis

Principal components analysis (PCA) was also used to show the construct validity of the Gap-fill .The three passages (reading comprehension, vocabulary examination, and Gap) were submitted to PCA using SPSS version 21. PCA revealed the existence of one component with an own meaning above 1, which clarified 68% of the variance.

Table 2 shows the loadings of each variable on the single factor extracted from the data. All the loadings are very high (above .79) which supports the unidimensionality of the data. The fact that Gap-fill 2 (written discourse Gap) highly loads on a single factor on which all other measures load supports the validity of the Gap as a measure of overall language ability and reading comprehension.

Table 2: Component matrix and factor loadings

	Component
	1
Gramm	.822
Vocab	.798
Gap Spk.	.792
Gap Wrt.	.893
Read	.836

Note: Gap Wrt.= Gap-fill based on written discourse; Gap Spk= Gap-fill based on spoken discourse; Gap Com.=Combined score of written and spoken Gap-fill

DISCUSSION

This research sought to assess the usefulness of the Gap-fill as an indicator of general English language proficiency among Iraqi EFL learners. The findings showed that the Gap-fill scores and the criterion fill scores highly and significantly correlate. When vocabulary, grammar and reading were combined into an overall test of language competence, a high correlation was observed between the Gap-fill and this combined measure which supports the validity of Gap-fill as a fill of overall foreign language proficiency.

The criterion measures were also moderately correlated. Zare and Boori (2018) found a strong correlation between Gap-fill and reading comprehension (r=.81, p<.01, n=150) while Yazdinejad and Zeraatpishe (2019) found a correlation of 0.48 between reading comprehension and Gap-fill. They argued that this small correlation was due to the unreliability of measures and the correlation augmented to 0.81 when corrected for attenuation. In the present study a correlation of 0.68 was found between Gap and reading comprehension.

The correlation between the Gap-fill based on written passages and reading comprehension(r=.70) was higher than the correlation of the Gap-fill based on conversational passages and reading comprehension (r=.53). A z test for testing the difference between correlations in dependent samples was run to examine if the difference between these correlations was significant (Eid et al., 2011). Findings showed that the difference between the coefficients of correlation (.17) was statistically significant, z=3.52, p<.001. This finding suggests that if cloze is constructed from written discourse it is a better measure of reading comprehension than a Gap-fill constructed from conversations. Table 2 shows that the written discourse Gap-fill correlates almost as high as the combined Gap-fill with the criterion measures. This is evidence that the spoken discourse Gap-fill does not add to the predictive power of the Gap-fill although it improves reliability. In the context of C- fill - a fill form very similar to Gap - Baghaei et al. (2009) found that there is

not much difference between C- fill constructed from written discourse and spoken discourse passages in predicting reading and listening comprehension.

Principal components analysis showed that all the variables, i.e., the two Gap passages, grammar test, vocabulary fill, and the reading comprehension fill, loaded on a single factor with strong loadings which establishes the unidimensionality of the data. Sheybani and Zeraatpishe (2018), using exploratory factor analysis, showed that reading comprehension tests composed of multiple-choice Gap-fill and multiple-choice questions based on sustained texts are not unidimensional. They attributed this to the method factor created by Gap items. Baghaei and Ravand (2019) using bifactor multidimensional Rasch model demonstrated that Gap-fill items create an ancillary nuisance dimension in the data.

Principal components analysis resulted in a one-factor solution. It showed that all the variables measure one single trait that can be called overall language ability. Since reading, vocabulary and grammar are prerequisites for communicative success their loading along with the Gap-fills on a single factor is evidence of the validity of the Gap as a fill of general language proficiency. In other words, Gap-fill is a credible predictor of general language ability.

Alderson (2005) states that many teachers do not have any training in test development and their items are of poor quality. Coniam (2009) further argues that teachers usually resort to published tests and select their required test items from already published tests. They either do not have the expertise to develop their own

test items or do not have the time to do so or both. Besides, with low payments for teachers in most parts of the world many do not feel that they should spend time for test development.

Reduced redundancy tests in general and Gap-fills in particular are economical and easy to develop tests that all EFL teachers can use. Constructing Gap-fill is easy and does not need expertise. They can be constructed mechanically by deleting every nth word as long the text is at the right difficulty level for the students. The scoring is also simple and straightforward Gap-fills. can also be used as measures of reading comprehension. Therefore, EFL teachers can easily construct Gap-fills and us them in their reading courses. The problems of lack of time and expertise for EFL teachers can be mitigated by using Gap-fills or other reduced redundancy tests.

In this study multiple-choice Gap-fill was examined. Open-ended Gap-fills with different scoring procedures (acceptable word, exact word) should also be investigated in future and the findings be compared with the multiple-choice Gap in terms of validity and reliability. It is also suggested that further research should be concerned with making a comparative study about Gap-fill, C- fill and Gap elide test in Iraqi universities. That is, the other tests of the family of reduced redundancy tests should also be validated among Iraqi EFL learners. Validation at the gap level by examining the traits or subskills which contribute to successful Gap-fill performance and analysis of Gap with modern fill theory should also be given

attention (Baghaei & Doebler, 2019; Baghaei, Yanagida, & Heene, 2017; Baghaei & Kubinger, 2015; Widyaningsih, et al., 2021). Examining the relationship between Gap-fill and affective variables such as anxiety, motivation, and willingness to communicate (Baghaei, Hohensinn, & Kubinger, 2014; Baghaei, 2013) and the application of C-Test for measuring other abilities such as intelligence are also suggested (Baghaei & Tabatabaee, 2015).

References

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment.* London: Continuum.

Aryadoust, V., & Baghaei, P. (2016). Does EFL readers' lexical and grammatical knowledge predict their reading ability? Insights from a Perceptron Artificial Neural Network study. *Educational Assessment, 21,* 135-156.

Baghaei, P. (2013). Development and psychometric evaluation of a multidimensional scale of willingness to communicate in a foreign language. *European Journal of Psychology of Education, 28*, 1087-1103.

Baghaei, P. & Doebler, P. (2019). Introduction to the Rasch Poisson Counts Model: An R tutorial. *Psychological Reports*, *122* (5), 1967-1994.

Baghaei, P., Hohensinn, C., & Kubinger, K.D. (2014). The Persian adaptation of the foreign language reading anxiety scale: A psychometric analysis. *Psychological Reports, 114,* 315-325.

Baghaei, P. & Kubinger, K. D. (2015). Linear logistic test modeling with R. *Practical Assessment, Research & Evaluation, 20*, 1-11.

Baghaei, P., Monshi Toussi, M.T., & Boori, A. A. (2009). An Investigation into the validity of conversational C-Test as a measure of oral abilities. *Iranian EFL Journal, 4*, 94-109.

Baghaei, P., & Ravand, H. (2019). Method bias in cloze tests as reading comprehension measures. *SAGE Open*. doi: 10.1177/2158244019832706

Baghaei, P., & Tabatabaee, M. (2015). The C-Test: An integrative measure of crystallized intelligence. *Journal of Intelligence*, *3*, 46-58. Available: http://www.mdpi.com/2079-3200/3/2/46

Baghaei, P., Yanagida, T., & Heene, M. (2017). Development of a descriptive fit statistic for the Rasch model. *North American Journal of Psychology*, *19*, 155-168.

Coniam, D. (2009). Investigating the quality of teacher-produced tests for EFL students and the effect of training in test development principles on improving test quality. *System, 37*, 226–242.

Eid, M., Gollwitzer, M., & Schmitt, M. (2011). *Statistik und Forschungsmethoden Lehrbuch* [Statistics and research methods textbook]. Weinheim: Beltz.

Sheybani, E., & Zeraatpishe, M. (2018). On the dimensionality of reading comprehension tests composed of text comprehension items and cloze test items. *International Journal of Language Testing, 8*, 12-26.

Spolsky, B. (1968). *What does it mean to know a language? Or how do* you get someone to perform his competence? Paper presented at the Second Conference on Problems in Foreign Language Testing, University of Southern California.

Widyaningsih, S. W., Yusuf, I., Prasetyo, Z. K., & Istiyono, E. (2021). The development of the HOTS test of physics based on modern test theory: Question Modeling through e-learning of moodle LMS. *International Journal of Instruction, 14*(4), 51-68.

Yazdinejad, A., & Zeraatpishe, M. (2019). Investigating the validity of partial dictation as a test of overall language proficiency. *International Journal of Language Testing, 9*, 44-56.

Zare, S., & Boori, A. A. (2018). Psychometric evaluation of the speeded cloze-elide test as a general test of proficiency in English as a foreign language. *International Journal of Language Testing, 8*, 33-43.