UZBEKISTAN MoD FOREIGN LANGUAGE APTITUDE TEST BATTERY PREDICTIVE POWER ANALYSIS

Ilya Zverev
Uzbekistan PfP
Training Center
UZBEKISTAN
zverev.elijah@gmail.com

ABSTRACT

Discrepancy in the speed and ease of progress exhibited by some learners and seemingly absent in others is one of the most widely observed phenomena of the foreign language classroom. It is usually explained via an invocation of foreign language aptitude, a set of cognitive abilities considered conducive to the tasks in question. Uzbekistan MoD Foreign Language Aptitude Test Battery is one of the instruments utilized for measurement of such abilities. Our goal is to establish whether the results yielded by each of the Subtests comprising the Battery in question can be used to predict the eventual success of the testee in his or her language studies. Here we demonstrate the lack of significant predictive power of three out of four Battery Subtests through both correlational and multiple-regression analyses conducted in IBM SPSS Statistics (version 26) and JASP (Version 0.12.2) based on the data from 39 male service members of Uzbekistan MoD enrolled in an intensive English language course. The results of the research suggest that exclusive reliance on the scores obtained by the testee in majority of Battery-comprising Subtests for the purposes of his eventual enrollment in an intensive language course might be detrimental.

Keywords: Uzbekistan Ministry of Defense, foreign language aptitude, foreign language aptitude test battery, predictive power.

INTRODUCTION

Intensive foreign language courses offered by "Partnership for Peace" Training Center of the Armed Forces of the Republic of Uzbekistan (Uzbekistan PfP) are conducted in the city of Tashkent and typically last for six months.

A trainee enrolled in any of the courses is freed of any service-related responsibilities at the place of his service. Nevertheless all the support from the Government that the trainee is entitled to in virtue of his rank and/or position does not cease. The only responsibility that such a person would have during the entire period of his studies is that of actually learning the foreign language.

Resources required for intensive foreign language training of a single individual (be it time, money or effort), consequently, demand that the selection instrument put to use for the purposes of candidate selection possess significant predictive power.

The primary instrument utilized by Uzbekistan Ministry of Defence in order to select candidates for intensive foreign language courses offered by "Partnership for Peace" Training Center of the Armed Forces of the Republic of Uzbekistan (Uzbekistan PfP) is "The Procedures of psychological assessment of candidates for intensive foreign language courses offered by 'Partnership for Peace' Training Center of the Armed Forces of the Republic of Uzbekistan".

For the purposes of brevity, throughout the study it will be referred to as "Uzbekistan Ministry of Defence Foreign Language Aptitude Test Battery – Uzbekistan MoD FLA TB).

The Battery comprises four Subtests: "Lexical Analogies", "Shape Selection", "Linguistic Decoding" and "Narration Summary". Every subtest carries a separate score subsequently utilized for the testee's allocation to a particular Professional Fitness Group (PFG) and eventually regarded as grounds for his enrollment in an intensive foreign language course.

The aim of the present study is to establish the existence (or lack thereof) of the predictive power of Uzbekistan MoD FLA TB constituents. The achievement of the aim set would require execution of the following tasks:

- to select (an) appropriate measurement instrument(s) that would enable us to quantify language learning success of an intensive foreign language course student;
- to conduct a correlational test using the results yielded by the application of the language learning success measurement instrument(s) and those demonstrated by an intensive foreign language course student in each of the four Uzbekistan MoD FLA TB Subtests;
- to compare the results of the correlational analyses conducted with those provided by Uzbekistan MoD FLA TB developers;
- to conduct sequential multiple correlation analyses in order to assess the predictive powers of each of the four Uzbekistan FLA TB Subtests in terms of the language learning success measurement instrument(s) selected.

LITERATURE REVIEW

Foreign language aptitude (FLA) is a psychological construct comprising a set of certain cognitive abilities conducive to foreign language learning. There are, consequently a number of approaches to both construct conceptualization and operationalization.

The Classical Model of Foreign Language Aptitude (FLAC) was established by an American educational psychologist John Bissel Carroll in the 1950s (Zverev, 2019, p. 141). According to this Model, there can be distinguished four basic abilities underlying one's aptitude for a foreign language: phonetic coding ability, grammatical sensitivity, inductive language learning ability and rote memory.

Since its original creation in the middle of the XX century, FLAC has exerted profound influence upon numerous approaches at reconceptualization of FLA construct. Among such approaches and theories it is necessary to mention the Linguistic Coding Differences Hypothesis (LCDH) of Richard Sparks and Leonore Ganschow (Sparks & Ganschow, 1991, 2001), CANAL-F FLA model of Elena L. Grigorenko, Robert J. Sternberg and Madeline E. Ehrman (Grigorenko et al., 2000), Information processing perspective and Macro-SLA aptitude model of Peter Skehan (Skehan, 2002), Aptitude Complexes/Ability Differential framework of Peter Robinson ((Robinson, 2005, 2012), High Level Language Aptitude Battery model of University of Maryland Center for Advanced Study of Language (Doughty et al., 2010), etc. However, it is FLAC that is still considered to be the most widely-used FLA model with Modern Language Aptitude Test as its principal operationalization instrument.

At the heart of every abovementioned approach, lies the understanding that FLA is distinct from general intelligence and, therefore, can only be measured by an appropriate model operationalization instrument, an FLA test. One of such tests is Uzbekistan MoD FLA TB.

Uzbekistan MoD FLA TB is a set of four subtests employed for the purposes of candidate selection for intensive foreign language training courses offered by Uzbekistan Partnership for Peace Training Center (Uzbekistan PfP).

Uzbekistan MoD FLA TB Subtest 1 is "Lexical Analogies". During the Subtest, the test taker is presented with thirty pairs of lexical items, each of which expresses "a certain relationship between its constituents" (Zverev, 2019, p. 145). The test taker is to understand what the relationship in question is and to select among the five options provided an element that might produce the identical relationship when placed alongside the stimulus word. The correct answer is worth one point; with no penalty for an incorrect response imposed, the maximum number of points the candidate might obtain for this Subtest is thirty.

Subtest 2 is "Shape Selection". For the purposes of this Subtest, the test taker is presented with two Sets of geometrical shapes, each containing five items marked with Cyrillic letters (а – д). Below each Set, there are twelve sub-Sets, each comprising up to four geometrical shapes. The sub-Sets are marked with Arabic numerals (1 – 12). The test taker is to regard each of the sub-Sets, to mentally "merge" each of their constituents together and to match the resulting shape with one of the five comprising the Set immediately above the sub-Set (1 – a, 2 – б, etc.). The correct response is worth one point. No penalty is imposed for an incorrect response. The maximum number of points for Subtest 2 is twenty.

In order to deal successfully with Subtest 3 of Uzbekistan MoD FLA TB, "Linguistic Decoding", the test taker is to perform translation from an unknown (artificial language) into his native language of ten sentences using the word-bank provided. The time allocated for this Subtest is five minutes and the maximum number of points the test taker can obtain as the result is ten.

Uzbekistan MoD FLA TB Subtest 4 ("Narration Summary") is the last subtest of the battery. The invigilator is to read a short text in test takers' native language, from which they are to remember as many details as they can. The task for the test taker is to replicate the text in the space provided in his answer sheet. Excluding the time necessary for the text reading by the invigilator, test takers have three minutes to perform this task. The maximum number of points for this subtest is ten.

Interestingly enough, but only Uzbekistan MoD FLA TB Subtest 3 ("Linguistic Decoding") might be viewed as bearing some sort of relation to any of the major FLA conceptualization approaches. The remaining four subtests resemble general intelligence tests (for a detailed discussion of the issue, see Zverev (2019)).

Any correlational analysis requires two pieces of data between which a correlation is to be established. Being guided by the data provided by Uzbekistan MoD FLA TB test developers (Akhrorov & Rakhimmirzaev, 2011b, p. 7), we faced the necessity of selection of an appropriate instrument for the measurement of "learner's foreign language learning success". No data as to what quantification procedure the Battery developers employed to that end has been provided in any of the available sources.

Since all the participants of the study were enrolled in an intensive English language course, language learning success measurement instruments we have decided to employ were American Language Course Book Quiz (ALC BQ) and American Language Course Placement Test. Both of them are standardized assessment tools developed and validated by USA Defence

Language Institute English Language Center. They constitute an integral part of the intensive English language course at Uzbekistan PfP.

ALC BQ is a multiple-choice test delivered to language learners upon their completion of studies of an American Language Course book (usually every five days). An ALC BQ requires that the test taker respond correctly to fifty questions (25 audio- and 25 reading-based) within forty-five minutes. Every correct response is worth one point. No penalty is imposed for an incorrect response. ALC BQ, consequently, is an interim progress test. At the end of the course of studies, every student will have passed twenty-four such tests (one per every week), the mean score of which we have utilized as one of the language learning success measurement instruments.

ALCPT consists of one hundred multiple-choice questions. Sixty-six of the questions are audio-based and thirty-four are reading-based. The test taker is given up to sixty minutes to respond to all the test items. No penalty for incorrect responses is imposed. Every correct response is worth one point. Typically, ALCPT is taken at the end of the course of studies and, therefore, might be viewed as an achievement test. We have decided to use ALCPT as an additional course success measurement instrument in order to account for possible discrepancies arising from the student's standard performance throughout the course and his performance upon its completion.

METHODOLOGY
The study included 39 native Uzbek language speakers enrolled in an intensive English language course at Uzbekistan PfP. All of the participants were male service members of Uzbekistan MoD.

The principal selection criterion for participation in the study was availability of the data on the participant's score in each of the four Subtests of Uzbekistan MoD FLA TB alongside with information on his performance in the intensive language course in question reflected through the measurement instrument(s) application.

As shown in Table 1, the majority of participants were commissioned officers with non-commissioned officers accounting for around 15% of the total number.

Table 1: *Study participants' distribution in terms of their ranks*

Rank	*f*	Rel. *f*	*cf*	Percentile
Lieutenant Colonel	1	0,026	39	100,00
Major	9	0,231	38	97,44
Captain	10	0,256	29	74,36
Senior Lieutenant	6	0,154	19	48,72
Lieutenant	7	0,179	13	33,33
Sergeant	1	0,026	6	15,38
Private	5	0,128	5	12,82

All the participants of the study were tested by means of Uzbekistan MoD FLA TB in groups of ten people maximum in the specifically designated quiet rooms on the territory of their military installations by trained Uzbekistan PfP officers.

Before the commencement of the test proper, all the participants were asked to fill in the answer sheets providing information on the date when the test was being conducted as well as the full name and military rank of the test taker. The time required for this stage is not included into the total time of the test.

The Battery Subtests were administered in a single sitting in the order established by the developers: "Lexical Analogies" (8 minutes), "Shape Selection" (9 minutes), "Linguistic Decoding" (5 minutes) and "Narration Summary" (3 minutes) (Akhrorov & Rakhimmirzaev, 2011a).

No audio or video equipment was used. The test takers were allowed to utilize either pencils or pens. Should there be a need for correction of the response provided, the test takers were asked to place a signature next to the point of the correction made in order to ascertain that no outside interference occurred.

We were not granted permission to work with the participants at the time of the testing; nor did we have access to their answer sheets upon their completion. The only information we were provided was the participants' scores in each of the four subtests.

During the entire period of the participants' studies at Uzbekistan PfP, we gathered data on their performance in each of the twenty-four ALC BQs and on the final ALCPT. For ALC BQs we calculated both the sum total of the scores and the mean score. The ALCPT score utilized was that reported upon the completion of this test at the very end of the course. Thus, for every participant of the study we had nine pieces of data: rank, scores in the four Uzbekistan MoD FLA TB Subtests, total Uzbekistan MoD FLA TB score, ALC BQ sum total score, ALC BQ mean score and ALCPT score. All the information on the participants' identity has been subsequently removed from the study.

RESULTS
Correlational Analysis
The number of participants in the study (thirty-nine) was not high enough to exclude the chance of our violation of the assumption of normality, one of the central for calculation of parametric Pearson's correlation coefficient statistic quantifying the correlation between two variables. Consequently, for the purposes of our study we decided to utilize Spearman's rho correlation coefficient, "a standardized measure of the strength of relationship between two variables that does not rely on the assumptions of a parametric test" [Andy Field; 2009; p. 794]. The results of the analysis conducted are provided in Table 2.

We have established the existence of a significant positive correlation between Uzbekistan MoD FLA TB Subtest 3 score and ALC BQ mean score, $r_s = .330$, p (one-tailed) < .05. Another positive significant correlation found was between Uzbekistan MoD FLA BT Subtest 3 score and ALCPT score, $r_s = .406$, p (one-tailed) < .05.

No other statistically significant correlations between the results obtained for any of the three remaining Subtests and either ALC BQ mean score or ALCPT score have been found. The results of the analysis conducted become ever more interesting when compared with those provided by the Battery developers (Akhrorov & Rakhimmirzaev, 2011b).

Table 2: *Correlation coefficient comparison (Spearman's Rho)*

	Sample Size	Spearman's Rho	Test of Significance	p
Akhrorov and Rakhimmirzaev				
Lexical Analogies vs Unknown Measure	86	.767	no information	< .05
Shape Selection vs Unknown Measure	86	.771	no information	< .05
Linguistic Decoding vs Unknown Measure	86	.808	no information	< .05
Narration Summary vs Unknown Measure	86	.742	no information	< .05
Zverev				
Lexical Analogies vs ALC BQ Mean	39	.176	one-tailed	.142
Lexical Analogies vs ALCPT	39	.218	one-tailed	.091
Shape Selection vs ALC BQ Mean	39	.061	one-tailed	.355
Shape Selection vs ALCPT	39	.057	one-tailed	.366
Linguistic Decoding vs ALC BQ Mean	39	.330	one-tailed	.020
Linguistic Decoding vs ALCPT	39	.406	one-tailed	.005
Narration Summary vs ALC BQ Mean	39	.251	one-tailed	.061
Narration Summary vs ALCPT	39	- .231	one-tailed	.922

Hierarchical Multiple Regression Analyses

In order to establish the predictive power of each of the four Uzbekistan MoD FLA TB Subtest scores in relation to both ALC BQ mean and ALCPT scores, two multiple regression analyses were conducted.

Based on the results of the assessment of each of the Subtests from the point of view of their compatibility with the existing FLA conceptualization models [Zverev; 2019], as well as those of the correlational analysis conducted within the framework of this study, for each of the outcome variables the sequence of predictor variable entry into the model was as following:

- Shape Selection score (r_s = .061, p (one-tailed) > .05);
- Lexical Analogies score (r_s = .218, p (one-tailed) > .05);
- Narration Summary score (r_s = .251, p (one-tailed) > .05)
- Linguistic Decoding score (r_s = .330, p (one-tailed) < .05).

The sequential regression analysis conducted to examine the effects other variables would exert on the model before the introduction of the Linguistic Decoding Score resulted in the unstandardized regression coefficients shown in Table 3.

The most predictive power among the three independent variables was that of Narration Summary score (ΔR^2 = .077) and Linguistic Decoding score (ΔR^2 = .112).

After the addition of all the predictor variables, the resulting model accounted for 25% of the variance in the outcome variable with Linguistic Decoding score accounting for a little over 11 % of the total variance. However, it was the only statistical variable in the regression model with all the four predictor variables with Confidence Interval Lower and Upper Boundaries (CI LB and UB) not crossing zero.

Table 3: *Multiple Regression Summary for ALC BQ Mean Score as the Outcome Variable*

Model		R	R^2	ΔR^2	B	B SE	B CI LB	UB	β
1	Constant	.187	.035	.035	40.313	2.030	36.20	44.43	
	Shape Selection score				.188	.162	-.14	.52	.187
2	Constant	.213	.045	.010	39.336	2.582	34.10	44.57	
	Shape Selection score				.181	.164	-.15	.51	.180
	Lexical Analogies score				.053	.085	-.12	.22	.101
3	Constant	.350	.123	.077	36.818	2.891	30.95	42.69	
	Shape Selection score				.089	.168	-.25	.43	.089
	Lexical Analogies score				.087	.085	-.08	.26	.169
	Narration Summary score				.446	.254	-.07	.96	.299
4	Constant	.484	.235	.112	35.149	2.840	29.38	40.92	
	Shape Selection score				.076	.159	-.25	.40	.076
	Lexical Analogies score				.007	.088	-.17	.19	.014
	Narration Summary score				.494	.241	.00	.98	.331
	Linguistic Decoding score				.480	.215	.04	.92	.373

Based on the results of the regression analysis conducted with ALC BQ mean score as the outcome variable, we can claim that each of the four independent variables has a rather low predictive power and taken together they explain only 23.5% of the total variance in it.

Another sequential regression analysis was conducted with ALPT score as the outcome variable. The unstandardized regression coefficients for this analysis are presented in Table 4. The analysis showed that the most significant predictor of ALCPT score was Linguistic Decoding score ($\Delta R^2 = .091$) followed by Lexical Analogies score ($\Delta R^2 = .068$) and Narration Summary score ($\Delta R^2 = .020$). The lowest predictive power, just as was the case with ALC BQ mean score as the outcome variable, was demonstrated by Shape Selection score that taken alone accounted for a meagre .007% of the total variance.

It should be noted, however, that there was not a single variable in any of the models constructed whose CI lower and upper boundaries did not cross zero. Overall, the four predictors do not account even for a fifth of the variation in the outcome variable (Model 4 R^2 = 0.186).

Table 4: *Multiple Regression Summary for ALCPT Score as the Outcome Variable*

Model		R	R^2	ΔR^2	B	B SE	B CI LB	B CI UB	β
1	Constant	.082	.007	.007	69.614	5.552	58.37	80.86	
	Shape Selection score				.222	.443	-.68	1.12	.082
2	Constant	.274	.075	.068	62.810	6.851	48.91	76.70	
	Shape Selection score				.169	.435	-.71	1.05	.063
	Lexical Analogies score				.366	.225	-.09	.82	.262
3	Constant	.308	.095	.020	66.249	7.914	50.18	82.33	
	Shape Selection score				.294	.459	-.64	1.23	.109
	Lexical Analogies score				.319	.232	-.15	.79	.228
	Narration Summary score				-.609	.695	-2.02	.80	-.151
4	Constant	.431	.186	.091	62.201	7.896	46.16	78.25	
	Shape Selection score				.261	.442	-.64	1.16	.096
	Lexical Analogies score				.124	.245	-.37	.62	.089
	Narration Summary score				-.493	.671	-1.86	.87	-.122
	Linguistic Decoding score				.164	.598	-.05	2.38	.336

DISCUSSION

As can be seen from Table 2, the values of Spearman's Rho correlation coefficient reported by Uzbekistan MoD FLA TB developers are significantly different from those that we have obtained based on the data obtained from the participants of the research.

Such a difference in the correlation coefficient values can be construed in terms of sufficient grounds for claiming that the test battery that we have utilized for the purposes of our research does not yield results as good as its developers claimed.

A possible explanation for the discrepancy discovered might be the difference in sample sizes that served as data sources for the two studies in question (86 vs 39 people). Nevertheless, we have already pointed out the fact that none of the Subtests comprising Uzbekistan MoD FLA TB can be regarded as having any sort of connection with any major FLA conceptualization models (most importantly, FLAC).

In the previous research (Zverev, 2019, p. 147) we have provided qualitative analyses of the Subtests and demonstrated that they cannot be viewed as FLA tests. The quantitative results obtained in the present study seem to confirm that claim of ours. In particular, "Shape Selection" Subtest (quite "suspicious" in terms of bearing any direct relation to FLA) has been shown to yield the results having the lowest impact upon one's success in intensive foreign language course.

Outcome variable notwithstanding, only one of the four subtests can be viewed as providing results possessing sufficient predictive power in terms of both language learning success instruments selected for the research. This Subtest ("Linguistic Decoding"), as expected, can be viewed as operationalizing inductive language learning ability of FLAC, however tenuous such operationalization might be.

CONCLUSIONS

The principal aim of the study conducted was establishment of the predictive powers of the four Subtests comprising Uzbekistan MoD FLA TB: Lexical Analogies, Shape Selection, Linguistic Decoding and Narration Summary.

Original test developers claimed to have demonstrated high correlation between each of the four Subtest scores and the test taker's eventual foreign language learning success.

The correlation coefficients can be interpreted as reflecting actual predicting powers of a measurement instrument: arguably, high correlation between a particular entrance test score and a particular exit test score can be viewed as the former's ability to predict the results of the latter. Consequently, in order to achieve the aim set, the first step would be to conduct a correlation analysis following where possible the procedure utilized by the test developers.
Since no information was provided as to what "language learning success" quantification instrument had been employed by the Battery developers, we have decided to use two USA DLI ECL-validated instruments constituting an integral part of the intensive English language course at Uzbekistan PfP: American Language Course Book Quiz (ALC BQ) mean score and American Language Course Placement Test (ALCPT) score.

Based on the data obtained from thirty-nine students enrolled in an intensive English language course at Uzbekistan PfP, a correlational analysis was conducted. Its results were different from those reported by the test developers: the only statistically significant correlation with quite low correlation coefficient values was demonstrated between Linguistic Decoding and ALC BQ / ALCPT scores.

In order to establish the predictive powers of each of the Subtest scores taking into account the results of correlational analysis as well as of a critical analyses of the Subtests conducted previously (results published in [Zverev; 2019]), we have carried out two sequential multiple regression analyses utilizing the data we had at our disposal. The results of the regression analyses demonstrated rather weak predictive powers of each of the four Subtests scores for either ALC BQ or ALCPT.

In both analyses, the models constructed showed that the strongest predictive power was that of Linguistic Decoding score, which accounted for around 11% of the total variation in ALC BQ outcome variable and approximately 9% of it in ALCPT variable.
Based on the results of the study, we suggest that constituents of Uzbekistan MoD FLA TB have rather weak power to be applied for prediction of one's successful performance within intensive English language course conducted by Uzbekistan PfP.

REFERENCES

Akhrorov, S., & Rakhimmirzaev, S. (2011a). *The Procedures of psychological assessment of candidate for intensive foreign language course offered by 'Partnership for Peace' Training Center of the Armed Forces of the Republic of Uzbekistan* (p. 20).
Akhrorov, S., & Rakhimmirzaev, S. (2011b). *The validation results of psychologic assessment of candidates for Uzbekistan PfP intensive foreign language training courses* (p. 16).
Doughty, C. J., Campbell, S. G., Mislevy, M. A., Bunting, M. F., Bowles, A. R., & Koeth, J. T. (2010). Predicting Near-Native Ability : The Factor Structure and Reliability of Hi-LAB. *Selected Proceedings of the 2008 Second Language Research Forum*, *December*

2014, 10–31. https://doi.org/10.1071/WR04129

Grigorenko, E. L., Sternberg, R. J., & Ehrman, M. E. (2000). A Theory-Based Approach to the Measurement of Foreign Language Learning Ability: The Canal-F Theory and Test. *The Modern Language Journal*, *84*(3), 390–405. https://doi.org/10.1111/0026-7902.00076

Robinson, P. (2005). Aptitude and Second Language Acquisition. *Annual Review of Applied Linguistics*, *25*, 46–73. https://doi.org/https://doi.org/10.1017/S0267190505000036

Robinson, P. (2012). Individual Differences, Aptitude Complexes, SLA Processes, and Aptitude Test Development. In M. Pawlak (Ed.), *New Perspectives on Individual Differences in Language Learning and Teaching* (pp. 57–75). Springer International Publishing AG. https://doi.org/10.1007/978-3-642-20850-8_4

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual Differences and Instructed Language Learning* (pp. 69–93). John Benjamins Publishing Company.

Sparks, R. L., & Ganschow, L. (1991). Foreign language learning difficulties: affective or native language aptitude differences? *Modern Language Journal*, *75*(1), 3–16.

Sparks, R. L., & Ganschow, L. (2001). Aptitude for learning a foreign language. *Annual Review of Applied Linguistics*, *21*, 90–111. https://doi.org/10.1017/S026719050100006X

Zverev, I. (2019). Uzbekistan MoD Foreign Language Aptitude Test: A Critical Evaluation. *Filologiya Masalalari*, *29*(2), 138–151.