6-19-2019

UZBEKISTAN MOD FOREIGN LANGUAGE APTITUDE TEST: A CRITICAL EVALUATION

Ilya Zverev Head of the English Language Department
Ministry of Defence Partnership for Peace Training Centre

Ilya Zverev

Ministry of Defence Partnership for Peace Training Centre, Head of the English Language Department

UZBEKISTAN MOD FOREIGN LANGUAGE APTITUDE TEST: A CRITICAL EVALUATION

ABSTRACT

Globalization has entailed a growth in importance of the second/foreign language teaching and learning all over the world with the number of both voluntary and involuntary language learners increasing on daily basis. There is, however, a widely attested discrepancy in actual results achieved by those engaged in second/foreign language learning usually explained by means of invocation of a specialized talent that certain individuals have, whilst others lack. Such a talent is thought to be measurable and the results obtained are regarded as valid predictors of success for intensive foreign language programs. The present article deals with critical appraisal of one of such instruments in terms of both its theoretical and practical validity. A number of points to be addressed for the purpose of the instrument improvement are demonstrated via referral to both basic statistic techniques and scientific consensus in the field of language learning aptitude research.

Илья Зверев

Мудофаа вазирлиги Тинчлик учун ҳамкорлик ўқув маркази, Инглиз тили кафедраси мудири

МВ ХОРИЖИЙ ТИЛ ЛАЁҚАТ ТЕСТИ: ТАНҚИДИЙ БАҲОЛАШ

АННОТАЦИЯ

Глобаллашув жараёни бутун дунё бўйлаб иккинчи чет тилни ўқитиш ва ўрганишнинг аҳамиятини оширди ва шу билан бирга кўнгилли ва мажбуран тил ўрганувчилар сонини ошишига ҳам сабаб бўлмоқда. Бироқ иккинчи чет тилни ўрганишга киришганларнинг натижалари орасида катта фарқ бўлиб, бу маълум шахсларнинг тилга бўлган иқтидори юқорилиги, қолганларда эса етишмаслиги билан тушунтирилади. Бундай иқтидорни баҳолаш мумкин ва қўлга киритилган натижалар интенсив тил ўрганиш дастурида муваффақият олиб келади, деб ҳисобланади. Шу каби воситалардан бирининг ҳам назарий ва амалий томондан яроқлилиги мақолада муҳокама қилинади. Тил ўрганиш лаёқати борасида олиб борилган статистик ва илмий изланишлар ёрдамида юқорида келтирилган воситаларни такомиллаштириш мақсадида таклифлар ишлаб чиқилди.

INTRODUCTION

Uzbekistan Ministry of Defence (Uzbekistan MOD) is the only governmental agency that has been continually providing intensive foreign language training for its service members for more than a decade via respective courses offered by Uzbekistan Partnership for Peace Training Centre (Uzbekistan PfP).

Intensive foreign language training (IFLT) can be defined as a foreign language teaching and learning within a certain period of time during which a student with zero to none foreign language learning experience gradually moves through a series of foreign language proficiency levels in order to achieve the one required by his/her superiors dedicating to this endeavor up to 8 hours daily.

IFLT requires application of materials specifically designed with certain constraints in mind (language difficulty level, daily cognitive load, etc.) and, by its very nature, is not an option that can be pursued by anyone. A mandatory requirement for implementation of any successful IFLT is utilization of certain candidate selection instruments, which is due to one of the most widely observed phenomena of the foreign language classroom, namely the discrepancy in the speed and ease of progress exhibited by some learners and seemingly absent in others.

The discrepancy in language learning abilities is attested in both laypeople's and professionals' parlance, and the latter are often none the wiser than the former, having been perplexed by the juxtaposition of those learning a foreign language rather effortlessly and those failing at the task in identical circumstance with identical opportunities provided [Sparks and Ganschow; 1991; p. 3].

Candidate selection for IFLT, therefore, is a means of making sure that given similar chances and similar training, the majority of those undergoing it will be able to achieve the results desired by the party investing in such a training.

AIMS AND TASKS OF THE RESEARCH

The aim of the present article is to provide a critical assessment of IFLT candidate selection instrument employed by Uzbekistan MOD for the last 8 years (and counting), namely of *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 such an assessment, the following tasks have been set:

• to review existing approaches to conceptualization and operationalization of foreign language aptitude (FLA) construct with a particular focus on classical FLA theory remaining valid to this day;

• to provide an overview of existing FLA test batteries;

• to describe the structure of Uzbekistan MOD's Procedures;

• to assess the validity of The Procedures in terms of the data provided.

1. CLASSICAL VIEW OF FOREIGN LANGUAGE APTITUDE

One of the principal individual differences evoked to explain the success of some and failure of others in foreign language learning is the "aptitude for languages"

or "foreign language aptitude".

Foreign language aptitude (FLA) is a ***componential psychological construct***, i.e. a "postulated attribute of people" [Cronbach and Meehl; 1955; p. 178] not subject to immediate observation or measurement encompassing a number of cognitive abilities [Wen, Biedroń and Skehan; 2017; p. 2] underlying certain observed differences in a person's behavior in response to an external stimulus [Wesche, Edwards and Wells; 1982; p. 128].

Any construct must be ***conceptualized***, i.e. provided with a *precise definition* in terms of the behavioral performance it is used to explain. Nevertheless, when it comes to FLA, there is a great heterogeneity of approaches to construct conceptualization [Dörnyei; 2005; p. 33; Li; 2016; p. 3]. Those approaches, in their turn, directly influence ***operationalization*** of the construct, i.e. provide it with conditions and procedures under which the performance enabling an observer to draw conclusions pertaining to the construct can be elicited [Bachman; 2004; p. 15]. A large enough number of similar performances (behavioral samples) can be regarded as a proof for the existence of the postulated construct, which must also be demonstrated to be distinct from those already accepted as valid by the research community.

Heterogeneity in construct conceptualization approaches is reflected in a number of existing instruments purporting to elicit certain performance and eventually provide some sort of measurement of the construct. Foreign language aptitude (FLA) is no exception to this rule [Dörnyei; 2005; p. 33; Li; 2016; p. 3; Winke; 2018; p. 2].

FLA has always been viewed through the prism of its possible application in two interconnected areas: student screening and student placement [Lutz; 1967; p. 3]. In the 1920s and 1930s, the two most commonly utilized types of FLA tests were [Carroll; 1962; p. 91]:

• tests measuring one's native language proficiency level and encompassing spelling accuracy, vocabulary and grammar structure range tasks;

• tests based on samples of tasks language learners would be likely to encounter during their foreign language studies ("work-sample" tests).

Early FLA tests correlated highly with intelligence measurement instruments and required the subject to possess some amount of previously obtained linguistic knowledge (including that of grammar-related terminology and comprehension of basic word-formation processes in their native language). The tests were in line with the objectives foreign language training was pursuing back in the day: it was chiefly concerned with inculcating students with reading and translation skills in the foreign language and did not pay much attention to communication.

High correlation between early FLA tests and intelligence measures raised a question of the need for specialized instruments of FLA measurement and assessment in the first place. Indeed, what is the point of having two tests yielding highly correlating results if one test would suffice just as well?

The FLA-as-IQ era ended with the advent of World War II and United States Government's realization of a rather dire state of affairs in the area of foreign language training of its personnel employed in missions overseas. It was then that the

intensive language training method (or "Army" method) was developed. The method is still widely used in all educational establishments all over the world utilizing materials of American Language Course (ALC) and requiring students enrolled to dedicate no less than six (usually nine) months of their lives to foreign language studies (up to 8 hours daily).

It can be argued that it was in direct response to the requirements of intensive foreign language training programs (and the Army Method) that the first scientific investigation of FLA and development of reliable FLA assessment batteries was initiated. Where such programs are in place (for instance, in Uzbekistan Ministry of Defence), there must necessarily be valid instruments utilized for the purposes of candidate selection for intensive foreign language training.

The *Classical View of FLA* was espoused by the founding father of the field, an American educational psychologist John Bissell Carroll during his work at Harvard University in the early 1950s. The foundational premises of this view are as following [Carroll; 1971; Li; 2015; Wen, Biedroń and Skehan; 2017]:

1) intensive foreign language training is a demanding endeavor in terms of various resources (both physical and temporal) required for its successful implementation;

2) there has been found *no way* of training program duration reduction beyond a certain critical point that would still guarantee attainment of satisfactory results [Carroll; 1962; p. 87];

3) *everybody* is capable of learning a foreign language, yet *not everybody* is capable of achieving satisfactory results in terms of foreign language speaking and understanding within the framework of "*rigorous*, *intensive*, *expensive* training programs in foreign languages operated by military and governmental organizations", where language classroom setting might be more favorable to those with high rather than with low aptitude scores [Carroll; 1962; p. 89; Sparks, Ganschow and Patton; 1995; p. 640; Winke; 2018; p. 6];

4) there are some people who possess "a fairly specialized talent (or group of talents)" [Carroll; 1962; p. 89] allowing them to successfully undergo intensive foreign language training, yet their share in the general population seems to be small;

5) the fairly specialized talent or FLA is a psychological trait *independent* other variables in both affective and cognitive domains influencing foreign language learning process [Carroll; 1962; p. 90]. Intelligence tests in particular have been demonstrated to be *rather weak and unreliable* for selection of candidates for intensive language training [Wesche, Edwards and Wells; 1982; p. 128] essentially providing a general cut-off point, yet being unable to deal with wide fluctuations at aptitude levels in those who crossed it [Carroll; 1962; p. 90];

6) FLA is a relatively stable, immutable and impervious to any attempts at mechanical development construct (it is thought to be impossible to teach a person to be better at his/her FLA);

7) the most direct manifestation of FLA is the rate at which language learners eventually achieve the required mastery level within certain time constraints, other conditions (motivation, instruction quality, learning opportunities, etc.) being optimal;

8) FLA is activated within contexts of *formal* instruction with direct manipulation of linguistic input and does not play a central role in natural language learning setting with no strict time limits imposed [Winke; 2018; p. 1];

9) FLA provides the learner with "initial stage of readiness" [Carroll; 1981; p. 81] for foreign language learning;

10) FLA is conceptualized as comprising four basic components (abilities): ***"Phonetic coding ability"*** is defined as the ability to encode novel auditory material (sounds or stings of sounds) in such a way as to make a delayed retrieval of it possible, i.e. the material can be "recognized, identified and remembered over time" [Sparks and Ganschow; 1991; p. 6]. The key element of the definition is the "delayed retrieval": a person capable of immediate reproduction of the material heard might not be equally successful in this task should any sort of disruption be introduced in between the two instances. Carroll claims that this ability is important for language learners since in the process they must not only learn "the identities of the new phonemes…, but must also recognize and remember the phonetic sequences represented by the morphemes, words, and intonation contours" [Carroll; 1971; p. 4]; ***"Grammatical sensitivity"*** is defined as the ability to perceive the differences in function various lexical items play within the boundaries of an utterance without overt instruction; ***"Inductive language learning ability"*** is defined as the ability to induce grammatical rules and forms, as well as meanings of lexical items from unknown language material without any assistance from an instructor. Despite Carroll's quite strong claims that language learning aptitude is not related to the general intelligence [Carroll; 1962; p. 90; Sparks and Ganschow; 1991; p. 4], this component admittedly "is most closely associated with general intelligence" [Carroll; 1971; p. 6]. The ability is operationalized thorough examination of linguistic material and drawing inferences as to the rules governing syntactical and grammatical relations between its various items. In other words, it is the ability to "notice and identify patterns of correspondence and relationships involving either meaning or grammatical form" [Carroll; 1971; p. 6]; ***"Rote memory for foreign language materials"*** is defined as the ability to memorize significant enough amounts of novel lexical or grammatical material in a foreign language.

The four-component model of FLA served as the basis for the Modern Language Aptitude Test (MLAT) first developed by Carroll and his colleague Stanley Sapon in 1959 and consisting of five subsections ("Learning numbers", "Phonetic script", "Spelling clues", "Words in sentences" and "Paired associates"). None of the MLAT subtests, however, is a pure manifestation of the four components outlined by Carroll himself, but rather "a hybrid mixture of the different underlying components" [Dörnyei and Skehan; 2008; p. 593]. Inductive language learning ability is hardly represented in the test at all [Wesche, Edwards and Wells; 1982; p. 130]. The MLAT has been in constant use for sixty years. It behooves us to point out the fact of the MLAT results' having been proven largely unsuccessful in prediction of foreign language achievement at advanced stages. A possible explanation, provided by Winke [Winke; 2018; p. 3] is that advanced foreign language learning stages to a larger degree draw upon cognitive and affective variables other than foreign language aptitude proper (motivation, effort,

learning strategies and study habits, attitude, personality, interest in the culture, etc.)

2. CURRENT FLA RESEARCH AND ASSESSMENT INSTRUMENTS

Both the MLAT and Carroll's ideas on FLA would dominate the field of FLA research for decades to come and they shape the way construct has been conceptualized and operationalized to this day. There have been, however, a number of developments that are well worthy of our attention.

First and foremost, as was pointed out by Skehan [Skehan; 2002; p. 69], in more than fifty years that passed since the original publication of Carroll's research result, there existed relative paucity in both theoretical and empirical research pertaining to FLA construct. That was due to three principal factors:

• conceptualization of the construct in terms of immutability and resistance to outside influence entailed the decrease in the interest towards construct investigation;

• impossibility of FLA alteration made it very difficult to align it with egalitarianism in education;

• the FLA was viewed as a product of an outdated language teaching methodologies of the 1950s and 60s that had been largely replaced by novel, communication-based ones.

The lack of interest in FLA research meant the lack of development in FLA knowledge. The state of affairs remained largely unchanged until the advent of the new century: a stable psychological trait that is impossible to change was measured before the beginning of a language course and the results obtained would be later correlated with the performance measurement at the end of the rouse. Skehan [Skehan; 2015; p. 2] called this type of a foreign language training course design aimed at general success/attrition rate prediction "macro design".

The XXI century counterpart of Carroll's "The Prediction of Success in Intensive Foreign Language Training", which laid the foundation of FLA conceptualization and operationalization in the XX century, is the anthology edited by Peter Robinson and published in 2002. According to Wen [Wen, Biedroń and Skehan; 2017; p. 6], it marked the "turning point in the reconceptualization of the construct of FLA". Among the innovative perspectives on FLA, a detailed discussion of which lies beyond the immediate scope of the present article, one might mention the following: The learning difficulties perspective and the Linguistic Coding Differences Hypothesis (LCDH) model of Richard Sparks and Leonore Ganschow [Sparks and Ganschow; 1991, 2001]; The successful intelligence perspective and the CANAL-F model of Elena L. Grigorenko, Robert J. Sternberg and Madeline E. Ehrman [Grigorenko, Sternberg and Ehrman; 2000]; The information processing perspective and the Macro-SLA aptitude model by Skehan [Skehan; 2002]; Aptitude Complexes/Ability Differential' framework of Peter Robinson [Robinson; 2005],[Robinson; 2012]; High Level Language Aptitude Battery (Hi-LAB) model by researchers at the University of Maryland Center for Advanced Study of Language [Doughty *et al.*; 2010; Linck *et al.*; 2013]; Brain networks as FLA; Working memory as FLA.

All in all, successive models of language learning aptitude have expanded upon

Carroll's original idea to such an extent that at the moment FLA has become an "umbrella term"[Dörnyei; 2005; p. 33] encompassing a great number of rather heterogeneous behaviors and factors (working memory, phonological short term memory, noticing ability, etc.). The original concept of FLA, however, "has stood the test of time and remained a valid concept" in spite of momentous changes that took place in the foreign language classroom, where instruction moved from audiolingualism (dominant during Carroll's days) to communicative language teaching [Wen, Biedroń and Skehan; 2017; p. 4].

There are several types of FLA test batteries (besides the MLAT proper) most commonly used in the world [Winke; 2018; p. 2]:

The Pimsleur Language Aptitude Battery (PLAB). Originally designed in the late 1950s and early 1960s by Paul Pimsleur, the test be acquired from the Language Learning and Testing Foundation. There are two elements that make PLAB different from MLAT: the target audience (PLAB was developed for young adults at the ages between 12 and 18) and motivation measurements that PLAB includes and MLAT does not;

The Cognitive Ability for Novelty in Acquisition of Language – Foreign (CANAL-F). This FLA battery was created by Elena Grigorenko, Robert Sternberg and Madeline Ehrman in the late 1990s in order to measure how well an individual is able to cope with novelty and ambiguity when engaged in foreign language learning [Grigorenko, Sternberg and Ehrman; 2000]. CANAL-F, just like MLAT, consists of five separate sections: learning meanings of neologisms from context; understanding the meaning of passages; continuous paired-associate learning; sentential inference; learning language rules;

The LLAMA Language Aptitude Tests. Developed by Paul Meara in the mid 2000s, this relatively new test battery comprises four subtests (LLAMA B, LLAMA E, LLAMA D and LLAMA F), all of which are computer-delivered and independent of the test taker's native language. Underlying the LLAMA LAT are mostly Carroll's views on FLA, however LLAMA has little in common with its paper-based counterpart in terms of both design and task structure [Meara; 2005];

Included under this rubric are remaining FLA tests that are not directly available to the public. The foremost among them are **Defence Language Aptitude Battery (DLAB)** and **High Level Language Aptitude Battery (Hi-LAB)**. The DLAB is currently used by the United States military to select candidates for foreign language training sponsored by the US government. Hi-LAB test battery that is being developed by the Center for Advanced Study of Language at the University of Maryland (https:// www.casl.umd.edu/hilab) is intended to be used for the purpose of identification of individuals capable of achieving advanced levels of proficiency in a foreign language [Doughty *et al.*; 2010; p. 10].

At the heart of each of the tests described above lies a certain vision of FLA construct. Tests are means of construct operationalization, yet it is conceptualization of the construct that affects the way it is operationalized.

3. UZBEKISTAN MOD IFLT TEST BATTERY

The test battery employed by Uzbekistan MOD for the purposes of IFLT candidates' selection has incorporated some of Carroll's ideas on FLA. The development of the battery reflects the acceptance of the fact of the existence of a specialized talent/group of talents for foreign language learning found in some, but not all service members of Uzbekistan MOD (point 4 in the list provided above). This acceptance, however, should not be interpreted in terms of the test taker's general inability to learn a foreign language, but rather in terms of his/her being capable of demonstrating satisfactory results within the constraint imposed by the entire IFLT system (points 1, 3 and 7). FLA seems to have been conceptualized as a componential construct, whose constituents (though not identified directly) are measured via four distinct subtests with the overall score taken to reflect the test taker's "readiness" (or, in terms of the ['The validation results of psychologic assessment of candidates for Uzbekistan PfP intensive foreign language training courses'; 2011; p. 15], subsequently referred to as "The Validation Results", his or her "***professional fitness***") for IFLT (points 9 and 10).

"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", as the IFLT candidate selection instrument is officially named, were developed by the Professional and Psychological Assessment Center of Uzbekistan MOD based on "classical psychological tests employed for assessment of complexes of specialized skills of service members" ['The validation results of psychologic assessment of candidates for Uzbekistan PfP intensive foreign language training courses'; 2011; p. 3].

The Procedures comprise four subtests designated by means of Arabic numerals (NB: examples provided for each of the Subtests are SIMILAR to those included in the actual test and NOT directly taken from it).

Subtest 1 ("Lexical analogies"). The subtest proper consists of 30 tasks of identical structure. The test taker is provided with a pair of lexical items in his/her native/primary language (Uzbek or Russian). The pair is structured in such a way as to express a certain relationship between its constituents. For instance:

SEDATIVE : DROWSINESS

(A) epidemic : contagious

(B) vaccine : virus

(C) laxative : drug

(D) anesthetic : numbness

(E) therapy : psychosis

In the elements above, the relationship is that of CAUSALITY, i.e. a sedative causes drowsiness. The test taker is to identify the relationship and select the option that expresses the same one (the correct answer for the task provided is D). The total amount of time allocated for this subtest is 8 minutes.

Subtest 2 ("Shape selection"). This subtest consists of 20 tasks of varying degree of complexity. The test taker is to perform mental manipulation with graphic

elements presented in random order and to select one of the five shapes those elements might constitute. The total amount of time allocated for this subtest is 9 minutes. An example of the task is provided in Figure 1. The test taker is presented with a combination of two shapes (#1), which he/she should mentally combine and select the best answer (a, b or c).

Subtest 3 ("Linguistic decoding"). The test taker is given a list of 21 words in an invented language with their exact translation into his/her native/first language. Below this word bank, there is a list of 10 sentences in the test taker's native/primary language. The task is to translate those 10 sentences into the artificial language utilizing the material provided in the word bank. The total amount of time allocated for this task is 5 minutes. The Subtest is supposed to assess the candidate's "linguistic maturation level", i.e. his/her ability to discover integral facets of language phenomena. High linguistic maturation level is viewed as connected with both native and foreign language proficiency achievement potential ['The validation results of psychologic assessment of candidates for Uzbekistan PfP intensive foreign language training courses'; 2011; p. 4].

Subtest 4 ("Narration summary"). The test taker listens to a short story comprising a number of interconnected "meaningful units" read by a test invigilator. The task is to remember as much information as possible and then to write it down within three minutes since the moment the oral presentation of the material is over. The total amount of time allocated for this task is 3 minutes. The Subtest is designed to assess "meaningful memory" of the test taker, that is his ability to memorize novel material by means of the so-called mnemonic props allowing him/her to overcome limitations of his/her short-term memory ['The validation results of psychologic assessment of candidates for Uzbekistan PfP intensive foreign language training courses'; 2011; p. 5].
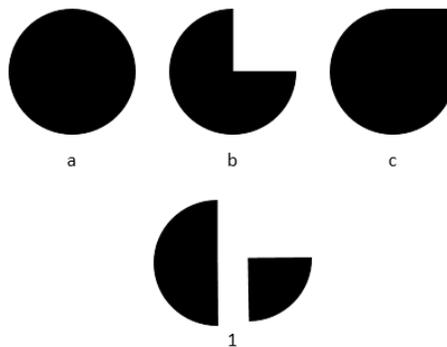
4. UZBEKISTAN MOD IFLT TEST BATTERY CRITICAL APPRAISAL

It must be emphasized that the discussion that follows is by no means directed against the developers of the original battery and is exclusively based on the content of the materials that we had at our disposal (both "The Validation Results" and "The Procedures"). The aim pursued is initiation of a meaningful dialogue on the issues raised and subsequent development of an instrument capable of complementing rather than completely replacing the existing one.

The first issue that cannot help but attract one's attention has to do with statistical data provided in "The Validation Results" and concerns the results utilized for test validation in the first place.

Thus, for every subtest there was conducted a correlational analysis (Spearman's rho - "a standardized measure of the strength of relationship between two variables that does not rely on the assumptions of a parametric test" [Field; 2009; p. 794]) of the results the test taker obtained and his/her foreign language learning achievement. The values for Spearman's rho for every subtest are provided in *Table 1* below (reproduced with minor changes from 'The validation results of psychologic assessment of candidates for Uzbekistan PfP intensive foreign language training courses' [2011; p. 7]).

For the test overall, or rather for the sum total of points gained by the test taker correlated with his/her foreign language achievement Spearman's rho reported was 0.783161 ($p. < 0.05$).

Each of the correlation coefficient values provided for the four subtests and for the whole test exceed correlational values of FLA with foreign language achievement reported for similar tests by leading researchers in the field by sometimes 50 points (the range provided for such values by Dörnyei & Skehan [Dörnyei and Skehan; 2008] based on the overview of research results in the field is between .20 and .60, "with a median value a little above .40"). Such a great discrepancy in the values cannot but raise some doubts as to the validity of statistical analysis conducted.

The score obtained by each test taker was standardized via application of the "Three-sigma rule", whose basic tenets are as following: in ***normally distributed data***, 68% of scores fall within 1 standard deviation unit of the mean, 95% - within 2 standard deviation units of the mean and 99.7% - within 3 standard deviation units of the mean. The key element of the Rule in question, presented in italicized bold type, is normality of data distribution (whether or not the distribution is perfectly symmetrical, i.e. characterized by values of skewness and kurtosis equal to 0).

Table №1

Subtest results and foreign language proficiency achievement correlation coefficients ($p < 0.05$)

Subtest	Sample size	Spearman's rho
"Lexical analogies"	86	0.766996
"Shape selection"	86	0.771079
"Linguistic decoding"	86	0.808089
"Narration summary"	86	0.741684

Normally-distributed data entails application of Pearson's correlation coefficient (r – "a standardized measure of the strength of relationship between two variables" [Field; 2009; p. 791]), whilst non-normally-distributed data necessitates application of a non-parametric statistic, that is of Spearman's rho. The fact of the latter's having been chosen for correlational analysis allows use to make a logical conclusion of non-normality of data distribution. Unfortunately, it is impossible to actually confirm

or refute this corollary, since no information on actual values of skewness or kurtosis was provided in the materials we had at our disposal (neither were there given any values of Kolmogorov – Smirnov or Shapiro – Wilk tests of normality of distribution).

The "Three-sigma rule", however, cannot be applied in any but normally distributed data set: in any other circumstances, the Rule's validity is reduced to zero. Actual utilization of the Rule by test developers can, therefore, be taken as an indication of normal distribution of raw scores.

What we have, consequently, is a logical paradox: Spearman's rho is only used for non-normally distributed data, whilst the "Three-sigma rule" is only applicable in normally distributed data. Their simultaneous application is impossible, yet it seems to have taken place, which violates the basic tenets of statistical analyses and bounds certain of the values obtained to be erroneous.

All of the remaining issues have more to do with the actual content of the test, rather than with statistical analyses of its results.

First, none of the four subtests of The Procedures is attested and included in any of commonly recognized FLA assessment instruments (briefly discussed in section 2 of the present paper).

"Lexical analogies" section (employed in the GRE test (used for the purposes of primary selection of candidates for graduate level courses in the United States of America)) usually serves a dual purpose: on the one hand it is designed to assess the test taker's vocabulary range; on the other hand, it is used in order to assess his/her ability to actually establish logical relationship between the items provided. Should both of those qualities of the analogy-based test have been actually preserved in The Procedures, it would not be possible to raise any objections whatsoever, since one of the FLA conceptualization approaches espoused, in particular, by Richard Sparks and Leonore Ganschow actually links one's native language literacy skills (including vocabulary range) with his/her FLA [Sparks and Ganschow; 1991; p. 4; Sparks *et al.*; 2011; p. 254; Wen, Biedroń and Skehan; 2017; p. 6].

However, the first aspect (vocabulary range) is hardly tested via The Procedures: the absolute majority of words in Russian (no data for the Uzbek language is available) are rather frequent and cannot be used to infer high Russian (native) language proficiency in the test taker. The only facet remaining, therefore, is the test that has more to do with a person's IQ level rather than with his FLA proper. Subtest 1, therefore, can be regarded as a pure IQ test that cuts off those below a certain threshold but cannot contribute to a better distinction among those who are above it (as per Carroll's ideas).

"Shape selection", a test of spacial imagination skills, has zero to none relationship to FLA proper and to assessment of a person's *language* abilities. The inclusion of this test into the battery designed for the purposes of candidate selection for intensive foreign language training is somewhat perplexing. The only possible explanation is that Subtest 2, just like Subtest 1, taps into the test taker's *general intelligence* abilities that, as has already been mentioned above, can only serve the purpose of cutting off those below a certain value. Subtest 2, therefore, is also an IQ rather than FLA test.

The third subtest of the battery, "Linguistic decoding", at first does seem as an appropriate for FLA, due to its being proposed as a measure of the test taker's ability to draw inferences and conclusions from the material presented without over partici-pation on the part of the instructor. However, a closer investigation of the test reveals a very simple procedure that anyone can follow in order to deal with the task presented: nothing more required than simple search of appropriate lexical items and copying them into the space provided in the answer sheet. Decoding, therefore, turns into the race with time rather than into a meaningful search for possible approaches to the task. The test that does not seem unique to Uzbekistan MOD FLA battery actually is unique in the sense of its having been completely stripped off its essence and difficulty. It is, consequently, quite difficult to state what the subtest actually measures (beyond, of course, the test taker's ability to find the only viable solution in place and to exercise sufficient speed in subsequent dealing with the tasks provided).

The fourth and final subtest, "Narration summary", can be regarded as the test that assesses the test taker's working memory (at the moment taken by some research-es as equal to FLA, (see Wen & Skehan, 2011). However, none of the standardized FLA measures has a part so dependent on its delivery (speed may vary from speaker to speaker, as may dialects in case of the Uzbek language, both of which can potentially contribute to the invalidation of the results). Moreover, it actually remains unclear why it was this subtest that was given the heaviest weight among all others in the final decision made as to the candidate's "vocational fitness".

CONCLUSION

Within the framework of the present article, we have provided an overview of approaches to both conceptualization and operationalization of FLA as a psycho-logical construct. We have demonstrated the basic tenets of the classical view of FLA alongside with a succinct overview of the existing instruments of FLA assessment.

We have described the current FLA test battery employed by Uzbekistan MOD for the purposes of candidate selection for IFLT and have demonstrated that it is com-prised of four subtests, three of which are mostly focused on IQ measurement. Results of IQ tests should not be used for the purposes the test battery is put, since FLA is a construct related, yet distinct from general intelligence of the test taker.

Statistical analyses provided by the test developers do not contain sufficient data required to draw a definitive conclusion as to their soundness and robustness. Key element missing is the information on normality/non-normality of data distribution, which underlies the whole premise of the research conducted.

Overall, taking into account the growing need for service members proficient in foreign languages, development of a new FLA test might be considered for the pur-poses of not replacing, but rather complementing the existing one in order to select the best candidates for intensive foreign language training.

The test will have to be built upon the results of existing research, scientific con-sensus as to the structure of FLA construct and best practices employed by developers of similar tests all over the world.

REFERENCES

1. Bachman, L. F. *Statistical Analysis for Language Assessment*. Cambridge University Press, 2004, 365 p.
2. Carroll, J. B. 'The Prediction of Success in Intensive Foreign Language Training.', *Training research and education*, 64, 1962, pp. 87–136.
3. Carroll, J. B. 'Implications of Aptitude Test Research and Psycholinguistic Theory for Foreign Language Teaching'. Princeton, New Jersey: Educational Testing Service, 1971, p. 14.
4. Carroll, J. B. 'Twenty-five years of research on foreign language aptitude', in Diller, K. (ed.) *Individual differences and universals in language learning aptitude*. Rowley, MA: Newbury House., 1981, pp. 83–118.
5. Cronbach, J. and Meehl, P. E. 'Construct Validity in Psychological Tests', *Psychological Bulletin*, 52, 1955, pp. 281–302.
6. Dörnyei, Z. *The psychology of the language learner: Individual differences in second language acquisition*, *Second Language Acquisition Research Theoretical and Methodological Issues*, 2005. doi: 10.4324/9781410613349.
7. Dörnyei, Z. and Skehan, P. 'Individual Differences in Second Language Learning', in *The Handbook of Second Language Acquisition*, 2008, pp. 589–630. doi: 10.1002/9780470756492.ch18.
8. Doughty, C. J. *et al.* 'Predicting Near-Native Ability : The Factor Structure and Reliability of Hi-LAB', *Selected proceedings of the 2008 Second Language research Forum*, (December 2014), 2010, pp. 10–31. doi: 10.1071/WR04129.
9. Field, A. *Discovering Statistics using SPSS*. London: SAGE Publications Ltd, 2009.
10. Grigorenko, E. L., Sternberg, R. J. and Ehrman, M. E. 'A Theory-Based Approach to the Measurement of Foreign Language Learning Ability: The Canal-F Theory and Test', *The Modern Language Journal*, 84(3), 2000, pp. 390–405. doi: 10.1111/0026-7902.00076.
11. Li, S. 'The Associations between Language Aptitude and Second Language Grammar Acquisition: A Meta-Analytic Review of Five Decades of Research', *Applied Linguistics*, 36(3), 2015, pp. 385–408. doi: 10.1093/applin/amu054.
12. Li, S. 'The Construct Validity of Language Aptitude', *Studies in Second Language Acquisition*, 38(4), 2016, pp. 801–842. doi: 10.1017/S027226311500042X.
13. Linck, J. A. *et al.* 'Hi-LAB: A new measure of aptitude for high-level language proficiency', *Language Learning*, 63(3), 2013, pp. 530–566. doi: 10.1111/lang.12011.
14. Lutz, M. *The Development of Foreign Language Aptitude Test: A review of the literature*. Princeton, New Jersey, 1967.
15. Meara, P. 'LLAMA Language Aptitude Tests The Manual', 2005, pp. 1–22.
16. Robinson, P. 'Aptitude and Second Language Acquisition', *Annual Review of Applied Linguistics*, 25, 2005, pp. 46–73. doi: https://doi.org/10.1017/S0267190505000036.
17. Robinson, P. 'Individual differences, aptitude complexes, SLA processes, and aptitude test development', in Pawlak, M. (ed.) *New Perspectives on Individual Differences in Language Learning and Teaching*. Berlin: Springer International Publishing AG, 2012, pp. 57–75.
18. Skehan, P. 'Theorising and updating aptitude', in Robinson, P. (ed.) *Individual differences and universals in language learning aptitude*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2002, pp. 69–93.
19. Skehan, P. 'Foreign Language Aptitude and Its Relationship with Grammar: A Critical Overview', *Applied Linguistics*, 36(3), 2015, pp. 367–384. doi: 10.1093/applin/amu072.
20. Sparks, R., Ganschow, L. and Patton, J. 'Prediction of performance in first year foreign language courses: Connections between native and foreign language learning', *Journal of Educational Psychology*, 87(4), 1995, pp. 638–655.
21. Sparks, R. L. *et al.* 'Subcomponents of Second-Language Aptitude and Second-Language Proficiency', *Modern Language Journal*, 95(2), 2011, pp. 253–273. doi: 10.1111/j.1540-4781.2011.01176.x.

22. Sparks, R. L. and Ganschow, L. 'Foreign language learning difficulties: affective or native language aptitude differences?', *Modern Language Journal*, 75(1), 1991, pp. 3–16.
23. Sparks, R. L. and Ganschow, L. 'Aptitude for learning a foreign language', *Annual Review of Applied Linguistics*, 21, 2001, pp. 90–111. doi: 10.1017/S026719050100006X.
24. 'The validation results of psychologic assessment of candidates for Uzbekistan PfP intensive foreign language training courses'. Ташкент, 2011, p. 16.
25. Wen, Z., Biedroń, A. and Skehan, P. 'Foreign language aptitude theory: Yesterday, today and tomorrow', *Language Teaching*, 50(1), 2017, pp. 1–31. doi: 10.1017/S0261444816000276.
26. Wen, Z. and Skehan, P. 'A new perspective on foreign language aptitude research: building and supporting a case for "Working memory as language aptitude"', *Ilha do desterro: a journal of language and literature = revista de língua e literatura*, (60), 2011, pp. 15–44. doi: 10.5007/2175-8026.2011n60p015.
27. Wesche, M., Edwards, H. and Wells, W. 'Foreign language aptitude and intelligence', *Applied Psycholinguistics*, 3(2), 1982, pp. 127–140. doi: 10.1017/S0142716400006664.
28. Winke, P. M. 'Aptitude Testing', in DelliCarpini, M. (ed.) *The TESOL Encyclopedia of English Language Teaching*. First Edit. John Wiley & Sons, Inc., 2018, pp. 1–7. doi: DOI: 10.1002/9781118784235.eelt0376.