Jacek Rysiewicz

Language aptitude as a predictor of success in the learning of English as a foreign language. A quantitative study.

Praca doktorska napisana pod kierunkiem
prof. dr hab. Krystyny Droździał-Szelest
w Instytucie Filologii Angielskiej
Uniwersytetu im. Adama Mickiewicza
w Poznaniu

Poznań 2004

Serdeczne podziękowania mojej promotorce
Pani Profesor Krystynie Drozdział–Szelest za pomoc przy pisaniu niniejszej pracy oraz za więcej niż anielską cierpliwość przy jej czytaniu.

In memory of John B. Carroll
(1913 – 2003)

CONTENTS

INTRODUCTION

0.1. Background of the study

The notion of foreign language aptitude as one of the factors influencing second language learning success has a long tradition of research and has occupied scholars from the field of both psychology and (applied) linguistics for at least eight decades now. Those who see language aptitude as a major determinant of successful second language learning point to the fact of the relative autonomy of an aptitude factor in relation to other unquestionable determinants of success in L2 learning such as motivation, attitude, anxiety or learning styles. Others point to the fact that language learning success is also affected by factors other than those resulting from individual differences between language learners. Sex, age, family background, learners' first language or teaching methods are just a few variables whose importance cannot be denied.

With such a plethora of factors, all potentially important in determining whether a learner will successfully master a second language, it is hardly surprising that all efforts at disentangling the web of interrelationships and interdependencies between such a multitude of ingredients with the hope of univocally identifying a limited set of features responsible for success (or pinpointing just one as *the* factor) must look like a daunting task to any applied linguist. Yet, any practising L2 teacher will readily provide dozens of examples from her/his experience of learners who learn a language with extreme ease and a more or less equal number of those who are 'hopelessly disappointing'. Between these two extremities there lies a whole spectrum of learners who are 'mundanely average'. And all this in a class of highly motivated pupils of the same age, with (roughly) similar social and intellectual background, enjoying their FL classes with a devoted teacher *au courant* with the latest in the teaching profession. And although such homogeneity is rare, a certain number of variables will inevitably be shared by all or most learners depending on the context of the teaching/learning situation thus reducing the interplay of possible factors responsible for success or its lack to a handful. One cannot help noticing that under those vaguely homogenous conditions some learners proceed faster and with more ease than others, and, in common parlance, we attribute 'quick' learners with 'a gift' , 'a flair' for languages.

It is precisely this type of practical knowledge of the problem described in the previous paragraph that has become a part of the present author's professional experience as an ELT teacher in various educational settings over a period of twenty years, and which has formed the main impetus for the current research.

0.2. Purpose of the study

The general purpose of the present study was to investigate the nature of the relationship between measures of second language proficiency (SLP), foreign language (FL) aptitude and intelligence (verbal and reasoning). Apart from considering the dynamic aspect of the three cognitive domains, the present study also asks questions about the internal structure of FL aptitude and of SLP. From this general purpose of the study two objectives followed:

O 1. The first objective was to investigate to what extent (if at all) SLP, FL aptitude and intelligence are constructs independent of each other

O 2. The second objective was to investigate which of the variables of FL aptitude and intelligence used in the study were the best predictors of L2 learning success as measured by the SLP criteria.

Each objective had its theoretical and practical implications, briefly characterised in the sections below where the research questions, formulated on the basis of the objectives, are discussed.

The analytical approach adopted in the study is quantitative in nature and the results of the tests used served as input data for product-oriented analyses. The test measures employed in the study include the following:

1. SLP measure – an adapted version of UCLES Key English Test (KET-A)
2. FL aptitude measures:
    a. *Test Predyspozycji Językowych* (TPJ) – language aptitude test
    b. *Progressive Language Aptitude Test* (PLAT) written by the present author
3. intelligence measures:
    a. *Standard Raven Progressive Matrices* (SRPM) – non-verbal reasoning test
    b. *Test Językowy LEKSYKON* (TJL) – verbal test of crystallised intelligence.

The test measures are described in more detail in *Chapter 2*. The major analytical tools employed were factor analysis and multiple regression, also discussed in *Chapter 2*. The participants of the data-acquisition stages of the study were 227, 13-year-old learners in seven lower secondary schools from and around Poznań (cf. *Chapter 2*).

0.2.1. Research Question 1

Following from the first objective (O1) as specified above the first *research question* (RQ) of the present study addresses the issue of the structure of human cognitive abilities:

RQ 1. **Is intelligence, as measured by standardised tests, an effective predictor of FL proficiency?**

The question was deemed of crucial importance for at least one reason. Namely, in order to meaningfully claim any validity for tests of second language learning potential (aptitude) in predicting a FL proficiency level it needs to be satisfactorily demonstrated that the abilities on which language learning is claimed to depend are in effect different from (or at least not identical to) general cognitive abilities associated with intelligence.

Three answers to RQ 1 are logically possible and have been pursued by various researchers at least since Spearman suggested that a general intelligence factor, the so called 'g-factor', is responsible for all human cognitive processes (Spearman 1927). The first of those answers would, somewhat extremely, claim that intelligence and language proficiency are in effect the same thing. This general cognitive version of the 'unitary hypothesis' assumes virtual equivalence between intelligence (understood as a reasoning ability) and language proficiency and was strongly supported by Oller (1981, 1983). The equivalence between non-verbal cognitive abilities (intelligence) and verbal (linguistic) skills underlying language proficiency would have important implications for the causal structure of human abilities. Namely, it could be assumed that the causal link goes from intelligence to language thus, attributing priority to intelligence over language specific/verbal abilities. In order to predict success in L2 learning it would be sufficient to obtain a measure of a subject's intelligence. However, earlier research on this issue produced conflicting evidence with some authors suggesting that intelligence, after all, was one of the poorest predictors of L2 achievement (von Wittich 1962) and others claiming just the opposite (Kaulfers 1931, 1939).

The second possibility, at the other extreme, would be that L2 proficiency and aptitude are totally independent of intelligence. This position known as 'completely divisible hypothesis' (cf. the contributions by Carrol, Oller and Volmer and Sang in Oller (ed.) 1983) assumes that there is no relationship between intelligence (reasoning) and language proficiency. If it could be demonstrated that there is a relationship (of whatever magnitude) between the two faculties the 'completely divisible hypothesis' would have to be rejected as a null hypothesis in the considerations of RQ1. Practically, any study that investigated the issue at hand has shown that this is the case. Consequently, it is expected that the data gathered in the present study will not support the null hypothesis.

The third answer that can be given to RQ1 is that while all the three domains enjoy relative independence from each other, there is some overlap between them. This 'compromise view' has been amply supported by research over the last few decades (cf. Gardner and Lambert 1965, Carroll 1983 or Boyle 1987) and will be pursued in this study as a viable hypothesis.

In order to address RQ1 it is hypothesised (after Boyle 1987) that intelligence and L2 proficiency/FL aptitude are neither identical nor completely independent but can be demonstrated to be relatively independent of each other with some overlapping areas common to either. It is also assumed that the hypothesis would be supported if factor analysis of intelligence, L2 proficiency and FL aptitude variables revealed that the variables of each of these three constructs emerged as separate factors. In other words, if it could be demonstrated that both the 'unitary' and the 'divisible' hypotheses are rejected, the 'compromise hypothesis' would receive support. On the other hand, if the intelligence, L2 proficiency and FL aptitude variables were shared by the same factor(s), then the hypothesis about their separateness would have to be refuted. To this end a statistical technique of factor analysis was employed in the study.

0.2.2. Research Questions 2a and 2b

Following from the second objective (O2) the second *research question* of the study has a more practical slant and addresses the problem of the extent to which language proficiency and/or its components can be best prognosticated from intelligence and aptitude measures. Depending on whether the integrated L2 proficiency or its sub-components are predicted this general **RQ 2** can be thus formulated as two separate but connected research questions: **RQ 2a** and **RQ 2b** below:

RQ 2a. What combination of six independent variables used (4 aptitude and 2 intelligence variables) allows for the most effective and efficient prediction of L2 success as measured by KET-A (Total) in the sample studied?

In order to address the question of choosing the best prediction model the following three competing hypotheses were envisaged:

1. Only aptitude variables best predict SLP
2. Only intelligence variables best predict SLP
3. A combination of aptitude and intelligence variables (verbal and reasoning) best predict SLP

First, the variables were used to predict the general SLP as measured by KET-A Total score (RQ 2a), and then to predict two components of SLP (cf. RQ 2b) identified in the process of the test's construct validation (cf. ARQ 2 [ancillary research question] below).

The answer to this question had both practical and theoretical implications. Practical, because the identification of the best possible predictors of second language proficiency would make the task of prognosis less of a fortune-telling and more of a predictable forecast. Ideally, with the age group in question we should be hoping for a test which could be administered within 45 to 60 minutes. Consequently, only a limited set of abilities can be included in a test aimed at predicting L2 learning success. Thus, the question of great practical importance was to identify the most efficient predictors.

The theoretical importance of the question lies in the implications it has for the validation of the FL aptitude construct itself. The assumption is that if the resultant best prediction model relies in its efficiency (exclusively, or to a large extent) on a given predictor then it may be assumed that the predicted performance (SLP) is related to the predictor(s) in a non-accidental, and theoretically valid, way. In other words, if the independent variables represent theoretically valid constructs (cf. ARQ 1 and 2) for the performance predicted then the fact of synergy between the predictors and the criterion may be interpreted as an indication of a tentatively causative link between the two.

In a similar vein, and given the composite nature of SLP, if it could be demonstrated that a different patterning of predictors is capable of discriminating between different components of SLP, then this fact would additionally lend support to the componential and learner-profile oriented treatment of aptitude as advocated by Skehan (1989, 1998) and others. This extension of RQ 2a is formulated as a separate RQ 2b below.

RQ 2b. Does the combination of the best predictors (as established in RQ 2) change when two different components of SLP (*Competence* and *Skills*) are considered as dependent variables?

If it could be demonstrated that the two components which emerged from factor analysis of SLP are differentially linked with the predictors i.e., that a given SLP component is best predicted by a given combination of the variables, then this fact, while lending additional support to the predictive potential of the model could also be indirect evidence for the componentiality of FL aptitude. Previous research on the structure of FL aptitude (Skehan 1986c, 1989, 1998; Wesche 1981) made it clear that, in practical terms, FL aptitude is best treated as componential in nature rather than as a monolithic, undifferentiated construct. Such a view of FL aptitude can lead to important insights concerning, among others, optimal instructional methods best suited to a particular learner profile (Skehan 1989). Skehan goes even further and claims having reached an integrated understanding of language aptitude, language teaching and language learning which he relates to an information-processing model of language. He then remarks that:

> *Aptitude tests are not simply predictive instruments which have no explanatory power. They have a rationale in relation to language learning processes, and an even wider justification through general psychological processes* (1989: 205).

Accordingly, higher scores on tests linked with, let's say, the skills component of SLP could be an indication to a teacher that the learner is better suited to more communicative-oriented instruction emphasising active language skills rather than knowledge of grammar and lexis. In other words, the results of aptitude tests if used informatively rather than purely predictively, may be an invaluable indication to teachers and learning process administrators of learners' strengths and weaknesses. Accordingly, instruction might be adapted to individual students' aptitude profiles.

0.2.3. Ancillary Research Question 1

However, before the above mentioned research questions could be properly addressed, the validity of FL aptitude measures (TPJ and PLAT) and the validity of the SLP measure (KET-A) had to be established. Hence, two *ancillary research questions* (ARQ) were found necessary as subservient to the research questions proper:

ARQ 1. What is the structure of second language proficiency (SLP) as revealed by the scores from KET-A (Total) test?

As it was not clear whether the Cambridge KET English test and teachers' grades of the subjects' oral performance would be an adequate measure of the subjects' L2 proficiency defined here rather traditionally as a four-skills (listening, reading, writing, speaking), multi-dimensional (morphology, syntax, lexis, phonology) construct, it had to be established first to what extent the results obtained represented second language proficiency thus defined. The construct of SLP as defined above was mostly measure-imposed[1] to the extent that the discrete dimensions of SLP as operationalised in the original KET (whose blueprint KET-A followed) included the following sub-tests: listening comprehension, reading comprehension, speaking[2], writing, vocabulary and grammar. This model of SLP which can be dated back to Lado (1961, 1964) and whose dominance in much of testing practice has been unchallenged since then, implies a Chomskian distinction into **knowledge (**competence**)** and **skills (**performance**)**. Thus understood, the model of SLP does not include pragmatic components postulated by Bachman in his framework of communicative language ability (CLA) and focuses only on, what he calls, organisational competence (Bachman 1990).

The results of earlier studies on the structure of SLP as discussed by Sasaki (1996:3-4) gave in effect four different views on the nature of the organisation of second language abilities. According to those views SLP can be explained by the postulation of:

1. several specific but uncorrelated factors (Lado 1961)
2. several specific but correlated factors (Bachman and Palmer 1983)
3. one general, higher-order factor with several correlated factors associated with it (Oller 1981, 1983)
4. one general, higher-order factor with several independent and specific factors (Bachman and Palmer 1983; Bachman 1982).

The four models have evolved from two extreme views on the trait structure of second language proficiency believed to underlie L2 test performance, customarily referred to in literature as: the completely divisible competence hypothesis and the unitary competence hypothesis (Sasaki 1993, 1996; Bachman and Palmer 1983). The first model originated with structuralist ideas, whereby SLP was viewed as a function of the four language skills and the components of language knowledge. The strong version of the completely divisible model would posit one independent factor for each cell in the component-by-skill matrix (Lado 1961). The strong version of the unitary hypothesis maintained that only one general factor can explain the entire amount of common variance in L2 tests.

Thus, the importance of demonstrating construct validity of the SLP measures, apart from obvious practical needs, would also have theoretical implications pointing towards a particular model of second language proficiency. This question was also necessary in the light of the fact that due to the limited testing time available, the original KET had to be adapted (hence: KET-A[dapted]) to suit the practical needs of the study. This mainly involved sacrificing the test's volume for its testing format.

0.2.4. Ancillary Research Question 2

ARQ 2. What is the nature of the abilities measured by the 'aptitude' tests used in the study?

This question was necessary for the following reason. Until recently the interest in the field of second language aptitude/ability testing has been virtually neglected by researchers in this country (cf. *section 1.6.* for a brief survey of studies done in Poland). The little research that did take place relied on aptitude tests constructed abroad or on purpose-made and largely *ad hoc* measures based on the then-dominant theory of FL aptitude operationalised in Carroll and Sapon (1959), and was limited in scope and impact to academic circles of professionals only.

It was not until a few years ago that signs of popular interest in testing individuals' learning potential for diagnostic and/or prognostic purposes became noticeable. The interest was sparked off by the rising importance of language education at school level (in the light of post-1989 socio-political changes taking place in the country) and associated with it a shift of the onset age for L2 instruction to Grade 4 at a primary school. Also the rise of language-profile secondary schools at both lower and upper level created a practical need for selection-on-admittance tools which would screen candidates for desirable L2 learning abilities before entry to those highly demanding programmes. And although such prognostication of language learning success is not a widespread practice in Poland, there have appeared a few tests claiming to be capable of doing just that.

For the purpose of the present study two aptitude tests were used. The first test (TPJ) was chosen on the basis of the test's possible impact due to its circulation and widespread use. The test was written by Kuliniak (2002) in co-operation with *Instytut Badania Kompetencji* (Institute for Competence Studies) in Wałbrzych in the south of Poland, an independent testing service institution which developed from, and as a result of, a project piloting a new secondary school leaving examination, the so called *"Nowa Matura"*. The test's author did not seem to have drawn on past research in the field of FL aptitude in the process of test construction but rather on his intuition and implicit ideas as to what abilities he deemed important for L2 learning (Kuliniak 2002:3). Because of this, and because of the test's possible impact on the process of diagnosis and prognosis in lower secondary schools it was considered of paramount importance to validate the test's theoretical construct to ascertain whether it measures what it claims to be measuring before it could be used in a further analysis.

The emphasis on construct validity of testing tools used for prediction purposes is particularly important in view of the problems with predictive validity often and persistently claimed to be a sufficient indicator of a test's validity. In this connection Bachman aptly observes:

> *One potential problem with examining predictive utility alone is that it can largely ignore the question of what abilities are being measured. We might find, for example, that scores on a test of signed number problems in mathematics are excellent predictors of performance of language courses. Regardless of such a finding, however, we would not seriously consider using this test as a predictor, because of the obvious mismatch between the ability the test appears to measure and the performance we are trying to predict. That is, in this case we would probably reject predictive utility as evidence supporting the validity of this test use* (1997: 250-1) [3].

The other of the two aptitude measures was a test constructed by the present author. In the process of its construction the author tried to draw on two chosen aspects of recent research on FL aptitude as evidenced in Skehan (1980, 1982, 1998) and Meara et al. (2001), and, most notably, Grigorenko and Sternberg (1998) and Grigorienko et al. (2000). Skehan's re-examination of the nature of memory as "a capacity to memorise more auditorily complex material [and] the capacity to impose organisation and structure on the material to be memorised" (1989:200) rather than as a simple stimulus-response link, inspired the present author to include an array of tests of different facets of *memory for text* alongside traditional, Carroll and Sapon's *paired-associate memory for verbal material*. Another source of inspiration for the present-author-written test was the rekindled interest in the measurement of learning **potential** rather than **developed** abilities observed in cognitive sciences in the last two decades or so, and applied to FL ability testing by Grigorenko, Sternberg and Ehrman in a seminal article published in *Modern Language Journal* in 2000. The shift from product to process-oriented measurement holds the promise of reconciling quantifiable and objective measurement with, sometimes excessively subjective and often meaninglessly impressionistic, qualitative approaches to behavioural research. Although the theoretical foundations for, what seems to be a completely new paradigm in ability testing, can be traced back to the classics of psychometrics i.e., Binet, Thorndike and others "credit for introducing the concept of dynamic testing to modern psychology is usually given to Lev Vygotsky"[4] (Grigorenko and Sternberg 1998:77). The ideas of the dynamic testing paradigm as applied to FL aptitude testing are discussed in more detail in Chapter 1 and *section 2.4.2.* As in the case of TPJ, so PLAT's construct needed to be validated first (cf. *section 3.6.2.*) before it could be used in subsequent analyses.

0.3. Limitations of the study

There are several limitations that make the results of the present study less readily generalisable than otherwise could have been hoped for. The first limitation impinging on the force of the arguments of the present thesis is connected with the nature of the data acquired for analysis. According to Carroll (1974) and repeated later in Carroll (1993) ideal conditions for an aptitude test should involve two applications of an achievement test: once before training, at time *A*, and once after training, at time *B*; and one application of an aptitude test: at time *A*. If between time *A* and time *B* significant change in achievement is observed then an ability is a measurement of aptitude for future learning success if there is no significant correlation of aptitude and achievement at time *A*, but there is a significant correlation between aptitude measured at time *A* with achievement measured at time *B* (Carroll 1974). These conditions are rarely obtained and in practice it is sufficient to retain the condition of considerable gain in learning and assume that "an ability is an aptitude if it helps in predicting degree of learning beyond a prediction from degree of prior learning" (Carroll 1993:17). In the present study no proficiency/achievement measure was obtained at time *A* and it is assumed that there was a change in achievement due to learning which took place between the application of aptitude measures and an achievement measure (cf. *section 2.6.*). However, considering the fact that the majority of the subjects had had ca. 2 years of L2 training prior to the research, some of the variance contained in the true score might not be reliable. In other words, the variance might be due to prior learning and training experience rather than to abilities ascribed to aptitude for languages.

The second limitation has to do with the less than perfect psychometric properties of the PLAT and KET-A measures used in the study. Despite the fact that prior to research use PLAT had been piloted on a smaller sample (cf. *section 2.2.*), some characteristics of the test's descriptive statistics could be objected to (cf. *section 3.1.*). Also, the fact that no standardisation was attempted can raise doubts about the extent to which the test's results are generalisable. With respect to KET-A, the problem of the commensurability of the three versions of the test used in the study appeared. Consequently, it might be argued that since different versions of the criterion test were used with different sub-groups of subjects the proficiency levels measured are not compatible with one another. The way the problem was partially dealt with is discussed in *section 2.5.*

The third qualification applies to the appropriacy of analytical tools used in the study with the view to seeking causal relationships between predictors and the criterion variables. With respect to the issue of causal relationships between variables in a multiple regression analysis Rietveld and Hout caution: "in many situations a multiple regression analysis can only provide part of the evidence required to substantiate the claim of causality"(1993:71). To achieve the level of explanation in the interpretation of gathered data more powerful models of statistical analysis such as structural equation modelling or confirmatory factor analysis enable a researcher to avoid some of the problems arising from the limitations of the analytical tool employed in the present study.

The fourth limitation follows from an insufficiently broad sampling of the abilities included in PLAT. Specifically, it is acknowledged that the non-inclusion of a test of phonemic coding along the lines of Carroll and Sapon's *Part II, Phonetic Script* of MLAT and re-examined by Reynolds (1999) constitutes a constraint on the generalisability of the study's findings. However, as it was stated in *section 0.2.*, one of the reasons for non-inclusion of phonemic coding in PLAT was the consideration of the test's practicality as well as the desire to focus on chosen aspects of FL aptitude i.e., memory and inductive language reasoning. This, however, will limit the significance of the FL aptitude factor model.

Finally, in investigating the extent to which intelligence and the components of aptitude studied influence the level of L2 proficiency other possible sources known to contribute to differential levels of L2 proficiency such as motivation, anxiety, cognitive styles, learning strategies and an array of others are not taken into consideration. Consequently, any claims of causal effects on the criterion performance are by necessity limited, on the one hand, to the model of learning potential abilities postulated for PLAT and, on the other, to the model of general cognitive abilities as measured by the two intelligence tests used (Raven and TJL).

0.4. Significance of the study

There are several reasons that make the present study important for the field of applied linguistics. The results of the research will have practical as well as theoretical implications.

First and foremost, the present study is the only piece of actual research in Poland in which an attempt has been made to address the issue of FL aptitude in a complex way. Earlier studies on FL aptitude were largely fragmentary in their focus and limited in their conclusions (cf. *section 1.6.*). They were based mostly on secondary school learner-samples (Figarski 1984; Lewowicki 1975 and Niżegorodcew 1975, 1979, 1980) or dealt with university candidates and/or students (Ozga and Tabakowska 1973, 1975 and Jankojć 2002). None investigated learners younger than 15 years at their early stages of L2 instruction. Furthermore, a few of those studies focused exclusively on FL aptitude (Niżegorodcew 1975, 1979, 1980; Ozga and Tabakowska 1973, 1975) and considered other factors such as learning styles, motivation and years of L2 instruction (Figarski 1984; Lewowicki 1975; Jankojć 2002), thus never going beyond the obvious in their implications for FL aptitude question. None of the studies mentioned investigated the relationship between intelligence (reasoning) and FL aptitude/L2 proficiency to the degree it is done in the present study.

Secondly, it is the first study to date trying to isolate native language specific predictors of FL achievement. The bulk of the prediction studies were done on native speakers of English, and there have been quite a few studies done for German, French, Japanese, Chinese, Finnish, Spanish and Russian. No single study has been identified which would do the same for Polish.

Thirdly, the present study is original in that it tries to construct-validate a new test of language aptitude (PLAT) partially based on the idea of dynamic testing (Grigorenko 2000), a proposal likely to become a new paradigm in ability testing in the years to come. All the previous studies relied exclusively on traditional FL aptitude tests as modelled by, mostly, Carroll and Sapon's MLAT (Modern Language Aptitude Test) and its followers.

The fourth reason that makes the present study significant has to do with the research method employed in data analysis. It is commonly agreed among researchers dealing with predicting educational success that the nature of the sample and the type of instruction offered may produce results which are largely artefactual i.e., occurring due to methodology-specific procedures rather than naturally present in the population. Thus, researchers agree that conditions favourable to prediction include, among others, the following two conditions: 1) heterogeneous sample – this provides a spectrum of possible abilities as reflected in a wide range of test scores; 2) intensive L2 instruction – this allows for a sizeable increment of the measured ability (L2 proficiency) to reveal itself to a researcher distinctly enough to be attributed to a predictor or a set of predictors rather than to instruction-external sources. Non-favourable conditions, on the other hand, are characterised by: 1) homogenous sample – where explicit or implicit pre-selection of subjects narrows the potential range of observable performance (low score variation); and 2) non-intensive language courses. While most research on FL aptitude done abroad and in Poland was carried out in contexts largely conducive to prediction, this study constitutes the first attempt to address the issue of 'language talent' in contexts in which most L2 instruction takes place in Poland i.e., in heterogeneous groups of learners going through a non-intensive L2 course. Such conditions are typical for virtually all lower secondary schools in this country.

Fifthly, the present study tries to implement advanced statistical techniques in data analysis for research on FL aptitude. Specifically, a multivariate technique of exploratory factor analysis is used to construct-validate the testing tools employed in the study, and the data analysis technique of multiple regression is used in deciding on the best prediction model for the criterion measure used. The use of these techniques has not been attested in any research on the aspect of applied linguistics dealt with in the study in Poland. The application of the two powerful techniques constitutes a definite instrumental advantage over the simple correlational techniques made use of in all the previous research cited above, and as such contributes to the field of applied linguistics in Poland.

Finally, the present study hopes to raise the quality and standards of FL aptitude testing in Poland beyond the current state of *ad hockery* and improvisation. The need for a more rigorous approach in constructing second language aptitude tests becomes apparent when one considers the fact that the only validation the published tests have had so far is the public interest sparked off by the press coverage the tests have received[5].

NOTES

1. Not surprisingly, widely-recognised, standardised SLP examinations have a 'normative' washback on the quality (hopefully) of teaching as well as on the instruction methods prevalent in public schools in Poland. In the hands of a reflective and qualified teacher the practice of 'teaching to the exam' (or a coursebook) does not have to be a vice.
2. This skill was not objectively tested in the present study for reasons of convenience. Instead, teachers' grades, both semester and year final were obtained. The grades were based on the subjects' overall performance throughout the year. Despite the impressionistic nature of such assessment, an effort was made to evaluate each subject on at least two occasions during regular class hours. This was done in an unobtrusive and stress-free way so that the subject did not know she/he was being assessed. Typically, a teacher would involve her/his pupils in a speaking activity structured around a task in the course of a regular lesson and would focus on one or two learners for the purpose of assessment. An analytic marking scheme was adapted from the KET Examiner Testing Manual to suit the level of each group of pupils. Marks were awarded on the following four scales: grammar, vocabulary, discourse management and intelligibility. Despite the efforts to objectivise the marking it was not possible to entirely avoid normative assessment. However, this is not a major problem since the interest of the research is in how each individual compares to her/his own achievement on various measures rather than in how individuals compare with each other.
3. It will be seen in *section 2.4.1.* and in *section 3.6.2.* that this quote applies pertinently to TPJ which, while correlating moderately highly with the KET-A criterion test ($r= 0.60$; $N=203$), does not seem to tap relevant abilities implicated in L2 learning. Again, great caution must be advised in relying too much on the predictive validity of a test if this has not been backed up by the test's construct validation.
4. The concept of dynamic testing has been known to cognitive scientists under various guises. Most common synonymous concepts are: ZPD (zone of proximal development), interactive, process or assisted testing, or tests of learning potential.
5. cf. www.gazeta.pl/edukacja for *Test Predyspozycji Językowych* (TPJ); and *Wiedza i Życie 1*, 1993, 26-27 for a test of L2 aptitude by Z. Dzięgielewska and co-authors.

CHAPTER 1

ISSUES AND PROBLEMS IN APTITUDE RESEARCH

This chapter presents a selective overview of issues and problems in FL aptitude research. *Section 1.1.* gives a comprehensive background to the history of research on FL aptitude. *Section 1.2.* examines the definitions of the field’s key concepts: aptitude, ability and achievement. In *section 1.3.* the issues of specificity, modifiability and componentiality of aptitudes are discussed. The discussion of the nature of second language proficiency (unitary or componential) in *section 1.4.* is linked with the debate on the structure of FL aptitude (cf. *section 1.3.*). The literature on the relationship between intelligence versus aptitude and second language proficiency is discussed in *section 1.5.* Finally, *section 1.6.* surveys research done in Poland on individual differences as determinants of L2 learning success.

1.1. Foreign language aptitude research

The scholars from the early period of language aptitude research (from around 1920 to 1939) had no clear understanding of what their object of study was and how it should be defined and so their studies were mainly psychometric in nature. The history of research on language aptitude as evidenced in the writings of such researchers as Henmon et al. (1929), Symmonds (1930), Kaulfers (1931) and Spoerl (1939) shows that, at the time, the interest in language aptitude focused on constructing convenient tools (tests) with the help of which it would be possible to prognosticate future learning outcomes. The ultimate aim was predominantly practical – to be able to determine, with a fair degree of accuracy, a learner's future success in learning a foreign language or, for that matter, any other subject or ability. The practical need of screening learners for desirable 'amounts' of aptitude before accepting them on a second language course was what shaped the thinking of the researchers of the period. In other words, the efforts of the researchers were aimed at assessing the usefulness of a tool by examining its practicality.

The predominant methodology of the period, adopted to determine the skills thought to be of paramount importance to successful L2 learning, was that of a simple correlational study. It involved a comparison of results from an aptitude test with results from an L2 achievement (usually semester/year final grades; sometimes test scores). High values of correlation coefficients would suggest a relationship between a measured ability and L2 learning to be interpreted by an informed researcher. As noted by Spolsky:

by the end of the pre-war stage of aptitude testing all the items that might be included in the model of aptitude had been mentioned, so that the task was not to think of new ones but to show the contribution of each to the model (1995b:323).

The research of the period also clearly showed that it was widely accepted that aptitude, irrespective of the structure and tools it was measured by, was only one factor, in the range of factors, which accounted for the total variance in language learning success (cf. Spolsky 1995b:323). In this connection Symonds remarks that:

achievement (...) is conditioned by other factors than ability, such as studiousness, school attitude, outside activities, interest of the learner, organisation of the material, skill of the teacher and the like (1930:555).

In the later period (50's and 60's), as more refined instruments of data analysis were becoming available, researchers would typically start off with a set of test items aimed at measuring certain abilities which they thought important for successful foreign language learning. Then they would try to establish, using statistical method of factor analysis, which of the test items in the whole test battery best tapped the postulated construct. (cf. Carroll 1958 for an example of factor analytic studies).

In both periods of research the decision as to what skills to include in a test was the result of a researcher's intuitive conviction rather than an informed judgement. Still, the 'first research then theory' approach enabled researchers to accumulate a considerable body of data which together with developments in applied linguistics and linguistics itself led to a postulation of hypotheses, models and theories of L2 learning. Those, in turn, fed back onto research on the nature of language aptitude, its structure and role in L2 learning. Researchers' increased understanding of the object of their study resulted in theoretical constructs such as the Carroll Model of School Learning (Carroll 1973, 1983), the Good Language Learner Model (Naiman, Frohlich, Todesco and Stern 1978, cited in Skehan 1989), the Linguistic Coding Differences Hypothesis (Sparks and Ganschow, 1991b) or the Cognitive Ability for Novelty in Acquisition of Language – Foreign, the so called CANAL-F Theory (Grigorenko, Sternberg and Ehrman 2000).

A full appraisal of just a fraction of the available aptitude studies or even a comprehensive presentation of available strains of evidence is beyond the scope of this dissertation and the reader is referred to Skehan (1989 and 1998), Spolsky (1995) or Carroll (1981) for an overview of the issues. In the section that follows I limit myself to a selective presentation of current issues in L2 learning aptitude research which are directly relevant to the study described in subsequent chapters.

1.2. Aptitude, ability and achievement – definitions.

This section discusses some of the ambiguities surrounding the key concepts used in the study. It starts with a survey of definitions of the concepts, followed by a discussion of the use of the terms in some current writings in the field.

1.2.1. The New Oxford Dictionary of English

The definition given by *The New Oxford Dictionary of English* (350,000 entries) illustrates well a non-specialised, 'layman' understanding of the concept. Aptitude is defined as: "a natural ability and/or tendency to do something" (Pearsall 1998:82). The word 'natural' suggests that there are certain characteristics inherent in an individual prior to the actual performance of a task requiring those abilities. In other words, the definition implies that aptitude is a certain innate ability which is, in effect, inborn and available to an individual without prior learning or training. An aptitude test defined as: "a test designed to determine a person's ability in a particular skill or field of knowledge" (Pearsell 1998:82) puts an equation mark between aptitude and ability and totally disregards the temporal priority suggested earlier. This definition, it seems, is contradictory in itself and as such not satisfactory when significant theoretical constructs are made to depend on finer delineations between the three concepts. Specifically, it is the nature of the relationship posited for the aptitude-ability distinction that defines the scope of the problem associated with the two notions.

1.2.2. The Penguin Dictionary of Psychology

A different relationship between the two terms is envisaged by *The Penguin Dictionary of Psychology* (Reber 1985) which defines ability as: "an individual's potential **to** perform", and aptitude as: "an individual's potential **for** performance, or the possibility of the individual being trained up to a specified level of ability" (ibid: 1). In other words, abilities are understood as available **now** to an individual for task performance with no further training needed whereas aptitudes characterise potential for achievement given appropriate instruction. Consequently, achievement tests are tests "designed to evaluate a person's current state of knowledge or skill (ability), whereas aptitude tests are designed to evaluate potentialities for achievement (ideally) independent of current knowledge" (Reber 1985:6). This understanding of aptitude-ability distinction stresses the dynamic relationship between the two and fully articulates the idea of temporal priority of aptitudes in relation to abilities. What is more, by treating aptitude as a kind of latent capacity; a sort of dormant, potential ability which develops into fully fledged ability with appropriate training, this definition assumes psychological priority of aptitudes over abilities. However, this distinction is not ever so neat as we might hope because, as Reber remarks: "tests of aptitude are, in reality, tests of performance (ability) and interest; the distinction in usage comes from the notion of making prediction about future achievements" (1985:50). The author suggests that the two are basically the same construct and that the only viable difference between aptitudes and abilities lies in the function they perform: a diagnostic function in the case of abilities; a prognostic function in the case of aptitudes. Achievement, it seems, is the least vague of the three concepts. Some of these issues will be taken up later in the chapter.

1.2.3. Multilingual Glossary of Language Testing Terms

The *Multilingual glossary of language testing terms* (Milanovic 1998) prepared by ALTE members (Association of Language Testers in Europe) gives this definition of an aptitude test: "A test designed to predict or measure a candidate's potential for success within a particular area of learning, e.g. in learning a foreign language, or on a specific course of study" (ibid: 135). The definition emphasises again the predictive use of a measure and the assessment of the potential for learning on which an individual may capitalise. Other aspects of the concept are not included in this definition.

1.2.4. Carroll's views on *aptitude-ability-achievement* distinction

Carroll is one of the stoutest proponents of the separateness of the three concepts as independent notions. In his monumental survey of factor analytic studies he has this to say on the aptitude-ability-achievement distinction: "I regard the term *ability* as entirely neutral and even uninformative as to whether any given ability is an *aptitude* or an *achievement*" (Carroll 1993:16). He then goes on to say:

> *Here, I used the term aptitude in a relatively narrow sense, i.e., to refer to a cognitive ability that is possibly predictive of certain kinds of future learning success; I exclude the notion of aptitude as interest in and motivation for a particular activity* (ibid.)

Apparently, Carroll understands the two terms as overlapping to a certain degree. For him *aptitude* is *ability* used in order to predict future learning: "aptitude measures are always measures of some kind of achievement, because the responses of individuals taking these tests are dependent to some degree on past learning" (1981:85). Consequently, the two notions are logically distinguishable only in the context of a specific classroom situation i.e. teaching methodology, pedagogical aims of the curriculum, specific achievement requirements etc. (cf. *section 1.3* on an ideal context for prediction). Aptitude is defined by Carroll in cognitive, rather than in personal, affective or motivational terms. According to him, aptitude and achievement are clearly distinguishable on theoretical and logical grounds. On this he remarks:

> *I reserve the term achievement to refer to the degree of learning in some procedure intended to produce learning, such as a formal or informal course of instruction(...). Tests designed primarily to measure such degree of learning are measures of achievement, in addition to being measures of ability* (Carroll 1993:17).

Generally speaking, FL aptitude is understood to refer to a set of primary capacities, propensities at an individual's disposal available to him prior to learning and, to a certain degree influencing his potential level of achievement. Carroll (1973, 1981) speaks of the time needed to achieve a given level of competence. Viewing FL aptitude as a function of time the claim he is making is that practically anyone is predestined, or so to say, doomed to learn an L2 provided the instruction is appropriate, the learner is motivated and able to profit from instruction.

On the basis of the many studies conducted on FL aptitude by himself and others, Carroll was lead to the following conclusions regarding the characteristics of the construct he had so perseveringly advocated. According to him, FL aptitude:

a. is a (relatively) stable learner characteristic
b. is a composite of several relatively independent cognitive abilities
c. is not the same as verbal intelligence or academic ability (although it overlaps to some extent with these domains).

1.2.5. Seashore, 1951

As a point of interest it is noteworthy to quote a definition of aptitude formulated by Seashore as early as 1951. It seems that most scholars of nowadays would probably find it hard not to accept such a view of aptitude:

> *An individual's probable rate or possible ease of learning a skill. It is to be noted that:*
> 1. *aptitudes are ordinarily quite stable but may change as a result of training*
> 2. *aptitudes are relatively specific or, at most, are related only within small groups*
> 3. *high aptitude leads to ease in terms of low energy cost per unit of output*
> 4. *interest and satisfaction in the exercise of potential ability are easily developed*
> 5. *rapidity of learning is positively correlated with high ultimate capacity* (in Stevens 1951:1343).

Seashore's definition is remarkable for at least one reason. Namely, it encompasses virtually all aspects of aptitude which were to dominate much of the debate in the years to follow and its meaning is still current. Like Carroll (1962), Seashore also links aptitude with speed and ease of learning (points 3 and 5)[1]. In articulating an idea that the direction of causality might be from greater ease in learning brought about by high aptitude to personality factors such as interest, motivation or attitude[2] (point 4) he suggests that personality-based and cognitively-based individual differences might, in fact, be independent of each other (cf. *section 1.3.*). Seashore also signals relative specificity of aptitudes from other variables (point 2), and admits modifiability of, what otherwise, is a stable construct (point 1).

1.2.6. Sternberg's *developing expertise* view of abilities

In connection with the foregoing discussion it is noteworthy to mention a diametrically opposite approach to the aptitude-ability versus achievement distinction as represented by Sternberg (1985, 1988, 1997). According to him abilities (and aptitudes underlying them) are neither precursors to, nor causes of expertise (achievement). Rather than viewing abilities as prior to achievement (as predicting expertise in a variety of domains) Sternberg regards them as developing forms of expertise (achievement) themselves. On aptitude-ability (ability-expertise) distinction he notes:

> *Ability tests are typically temporally prior in their administration to the administration of measurements of various kinds of achievements but that what they measure is not psychologically prior.(...) In viewing the tests of abilities as psychologically prior, we are confounding our own typical temporal ordering of measurement with some kind of psychological ordering. But in fact, our temporal ordering implies no psychological ordering at all* (Sternberg 1998:15).

In denying any causal and linear relationship between aptitudes/abilities and achievement Sternberg opposes a traditional psychometric view of test scores as causes of school or job expertise and believes that: "the traditional notion of test scores as somehow causal is based on a confounding of correlation with causation [and that] differences in test scores, academic performance, and job performance are all effects of differential levels of expertise" (1998:14). Another important theoretical and practical consequence of Sternberg's model is that abilities are viewed as flexible and, in consequence, modifiable by training as opposed to the traditional view which treated them largely as fixed and/or innate (Sternberg 1998). What is not entirely clear from the above model of abilities as developing expertise is the role and position of aptitude in relation to abilities and achievement.

Sternberg's developing-expertise point of view on abilities served as one of the elements on which a new theory of FL aptitude: CANAL-F (Cognitive Ability for Novelty in Acquisition of Language – Foreign) was based (Grigorenko and Sternberg 1998; Grigorenko et al. 2000). This will be discussed in more detail in *section 1.3.3.2.*

In the context of language learning the concept of aptitude, despite its obvious lack of scholarly precision (as evidenced by the foregoing discussion of the concept within the field of general cognitive abilities), is relatively well-researched experimentally and well-established in the cognitive theorising mainstream of applied linguistics. As with studies in the area of human general cognitive abilities, language learning abilities were thought of as being dependant (at least partially) on language learning specific skills, and different from general cognitive abilities such as intelligence, however understood (cf. *section 1.5.* for the discussion of the relationship between intelligence and aptitude).

It seems that it is the very nature of the data required for aptitude research that makes FL aptitude a relatively unpopular object of scholarly investigation while the studies of motivation, anxiety and other affective domains strive. On this Carroll aptly remarks:

> *...There is much more difficulty in nailing down the nature of (...) aptitude and intelligence than there is, say, in describing the nature of motives and interests, for the latter are in some sense more accessible to the individual's awareness, (...) whereas in the case of aptitude he is largely limited to reporting whether foreign language learning is easy or hard for him (...) but he could not describe his aptitudes in any detail. Presumably aptitudes reside deep within the individual's biological constitution and manifest themselves only indirectly (...). Attempts to describe the nature of an aptitude demand subtle and penetrating psychological research and analysis.* (1981:86)

As a consequence, in much of the writings in applied linguistics (at least in this country) the role of aptitude has often been downplayed and neglected (or overtly misunderstood) as a factor influencing language learning success, with a concurrent emphasis put on more readily accessible affective factors.

1.3. Issues in L2 aptitude research – a selective review

This section discusses some of the controversies with regard to FL aptitude research and is intended as a brief presentation of the current state of knowledge in the field with regard to the issues considered. Unless indicated otherwise the following discussion is based on the writings of: Skehan (1989, 1998); Spolsky (1989, 1995) and Carroll (1968, 1979, 1981, 1990).

1.3.1. FL aptitude is an autonomous notion

The assumption of specificity of abilities required for successful language learning is a cornerstone of any research on FL aptitude. The study of aptitude makes sense only when it can be demonstrated that among many other individual differences the posited construct of aptitude is a separate factor which independently contributes to second language learning. It is generally agreed among researchers that the differences in the level of L2 achievement can be attributed to three broad factors: social (family background, parental education etc.); cognitive (aptitude, intelligence, cognitive style, risk-taking, memory etc.), and affective (motivation, anxiety, self-esteem, attitude, learning styles etc.)[3]. Below only the cognitive and affective variables will be briefly discussed.

FL aptitude is generally claimed to be autonomous from other personality (cognitive and affective) variables and as Ellis remarks: " [it] is one of the 'big two' individual difference factors (the other being motivation). Research based on tests such as the MLAT has revealed consistent correlations with language achievement in the order of .40 or higher" (2004:531). As to the extent of the independence of a specific language talent from other cognitive factors two views can be envisaged. One would emphasise cognitive similarity of all human learning situations and argue that language learning capacity is essentially no different from the abilities needed in the learning of any other skill and/or ability. According to this view, talented learners of languages would be talented learners of everything (Skehan 1998). This position is untenable in the light of numerous research showing incongruity between scores on aptitude measures and other cognitive variables (cf. discussion below). It is also contrary to everyday classroom observation to the extent that high and very high achievers in academic subjects do not necessarily make good/very good language learners. A moderate alternative of this position would allow for a differential involvement of particular sub-abilities within the area of general cognitive abilities to be responsible for differential levels of achievement in L2. The other view would be that language learning is an altogether different capacity distinct from other cognitive abilities.

There is a whole gamut of studies examining the effects of affective variables on second language proficiency. In almost all of the studies listed below positive correlations have been attested between the variables and measures of proficiency and/or achievement used. The following are examples of the most frequently researched affective variables: attitudes and motivation (Nelson and Jakobovits 1970, Gardner and Lambert 1972, Gardner 1985, Gardner 1990), anxiety (Horwitz et al 1986, MacIntyre and Charos 1996, Sparks and Ganschow 1991a), self-esteem (Heyde 1977) and personality (Brown 1973, Chapelle and Roberts 1986, Oxford and Ehrman 1989). However, only few studies have investigated the relationship between affective variables and aptitude in predicting foreign language proficiency.

One of the recent studies investigating just that is Ehrman and Oxford (1995). In their study of a variety of cognitive variables in a large sample (N=855) of adult learners (mean age=39) in intensive training they found that although a number of cognitive variables (motivation, self-confidence, cognitive flexibility and others) proved to be important factors, FL aptitude was the strongest correlate of end-of-training achievement levels in speaking and reading. Their aptitude measure (MLAT) correlated at 0.50; learning strategies at 0.21; anxiety indices at 0.40 and a variety of personality variables at 0.20 / 0.30. with the measures of reading and speaking. (cf. *sections 1.5.; 3.8. and 4.3.* for discussion of intelligence versus aptitude relationship). Rubin (1975) having studied how good language learners achieve success in a foreign language is of an opinion that much of the discussion concerning best prerequisite conditions for L2 learning is basically spurious as at least three variables make separate critical contributions to L2 learning success: aptitude, motivation and opportunity. Skehan (1989), reviewing criticism levelled against the concept of aptitude, points out that, paradoxically and contrary to all available evidence, the superiority of affective variables over aptitude has often been taken for granted by the researchers in the field. He then goes on to conclude that: "Aptitude is consistently the most successful predictor of language learning success, and failure to include some attempt to measure aptitude will render the interpretation of any research study very difficult" (1989:38).

1.3.2. FL aptitude is not fixed and innate

This is a well-known 'nature-or-nurture' controversy and there are several related questions bearing on the issue. One is related to Gardner and Lambert (1972) and Neufeld (1979) criticism of the concept of second language aptitude as 'misconceived' along the lines that it makes no sense in the light of the fact that: 1). everybody has aptitude for mother tongue and learns his/her native language unexceptionally (Neufeld 1979); 2). in bilingual settings everybody learns two languages to more or less equal degree (Gardner and Lambert 1972). This argument assumes that since there are no individual differences in how we learn our mother tongues the concept of aptitude for languages is misguided and the differences in attainment levels should be sought in affective and personal domains.

However, as demonstrated by Wells (1985) in a study done within the Bristol Language Project[4], there is considerable variation in the rate of L1 acquisition between children. If would-be native speakers exhibit differential L1 development then alike differences related to the speed of L2 acquisition might be posited for second language learning. This is in agreement with Carroll's view of aptitude as determining the rate of L2 learning rather than a *sine qua non* condition for learning (cf. also point 5 in Seashore's 1951 definition).

Another question is of a logical nature. Namely, if aptitude were to be of any interest to applied linguists its status would have to be equal to other features determining individual differences in learning i.e., motivation, anxiety, personality styles etc. In other words, it would have to constitute a psychologically valid set of characteristics i.e., what is known in psychology as a 'trait'. In order to be 'psychologically real' aptitude should manifest itself in a sustained and stable fashion, otherwise any claims concerning its effects or the direction of causality are not possible. As a logical result, foreign language aptitude would have to be either inborn to an individual (or stabilise very early in life) or be impermeable to change through training.

Available empirical evidence points to two directions. Skehan (1986b) in a follow-up study to the Bristol Language Project administered a set of FL aptitude tests and school achievement measures to 100 children traced down from 10 years earlier. He found a number of significant correlations between the early first language measures and the later aptitude indices. This could tentatively be interpreted as a link between L2 achievement and underlying capacities for language learning (a case for innateness) and against aptitude as product of past learning experience (a case against modifiability of aptitudes). Also a study by Politzer and Weiss (1969, cited in Skehan 1989) generally yielded negative results as far as the attempts to improve learners scores on aptitude tests were concerned.

Klauer (1993), in his study of the effects of an inductive reasoning training upon FL aptitude done on 12-year-old children (N=61) argued for an opposite hypothesis. His research proved that improving inductive reasoning skills on the Raven test positively transferred onto tasks calling for the same ability in a language aptitude test. Klauer concludes his article saying:

> *There can be no doubt that the learners who underwent training were in a better position to learn the skills offered by the test than those who did not undergo such training. Thus, for the first time, a transfer [of learning] has been demonstrated to affect the very process of learning and not only the achievement of a task* (1993:8, own translation).

Although the results of Klauer's research clearly show that limited modifiability of selective aspects of aptitude is possible and conducive to L2 learning, it is still to be established empirically if these results are generalisable to other age groups and/or skills. Klauer's study might be taken to indicate that modifiability of aptitude is more feasible than previously thought. As things are, the effects of trainability seem to be limited to chosen age groups, proficiency levels and/or types of abilities being modified.

McLaughlin (1995), who approaches the question within an information processing perspective, discusses working memory as one of the components of FL aptitude. He looks at how 'expert' versus 'novice' learners cope with FL learning tasks and concluding he remarks:

> *I believe that there are strategies that can be taught to increase the efficiency of working-memory processes [in order] to refine the repertoire of strategies of poor learners so that they may benefit from strategies used by 'expert' learners* (1995:383).

Apparently, the view that McLaughlin argues is that of aptitude as modifiable by previous learning and experience, at least with respect to the component he discusses. All in all, it is generally accepted that FL aptitude rather than being fixed and/or innate shows a fair amount of stability over longer periods of time and does not lend itself readily to any significant modification (improvement) through 'aptitude training'.

1.3.3. FL aptitude is componential

The issue of a non-unitary nature of FL aptitude can be viewed from two perspectives which are related to the history of research on the concept of aptitude. To be more precise, they can be linked to the purpose and aims of such research as envisaged by contemporary researchers. Namely, depending on whether the main purpose of aptitude research was prognostication of L2 learning success or the diagnosis of L2 learning taking place here and now, researchers focused either on the nature and structure of the L2 specific talent and on how to measure it (prognosis), or on how the structure thus uncovered contributes to how the learners go about the task of learning (diagnosis). Although prognostication was historically prior to diagnosis both perspectives form a viable basis on which much of the present-day research relies.

The foregoing discussion of the nature of language aptitude, has, in fact, implicitly assumed that FL aptitude is not a unitary ability but rather that it consists of a set of several aptitudes which can be isolated from the whole gamut of posited language-learning related abilities (Carroll 1981). Rather than being a monolithic capacity, FL aptitude is claimed to be made up of a range of sub-skills which may differentially contribute to second language achievement level. Despite the fact that the nature of these abilities is by no means at all clear and the structure of FL aptitude is often debated and questioned, "the proposition that second language acquisition is mediated by specific abilities raises many exciting possibilities" (Gardner 1985:24). These possibilities are connected with the diagnostic, rather than prognostic function of FL aptitude tests.

1.3.3.1. Diagnosis

Diagnosis, by being more interested in the 'now' of learning, tries to furnish language learners, teachers and administrators with relevant information concerning the learning profile of individual learners i.e., his/her strengths and weaknesses. A multi-componential view of FL aptitude offers a possibility of tailoring courses which would capitalise on learners' abilities, compensate for specific inabilities and/or orientate teaching toward students' strengths (Gardner 1985).

Pimsleur (1968) demonstrates how his Language Aptitude Battery (LAB) can be used: "in detecting beforehand the strong and not-so-strong points of each student, so he can be helped as early and as much as possible" (1968:102). He illustrates the diagnostic (and indirectly prognostic) value of his 'empirical theory' on the example of four cases of learners. Each learner profile has a differential level of scores obtained in the four parts of LAB: *Vocabulary, Language Analysis, Sound Discrimination* and *Sound-Symbol Association*. Pimsleur concludes:

> *These cases illustrate typical applications of LAB for diagnostic purposes. (...) I believe it can fairly be stated that the Battery is the only means currently available by which such predictive and diagnostic information can conveniently be obtained* (1968:104).

Wesche (1981) analysed the data gathered by the Public Service Commission of Canada from over 2000 adult learners over a period of several years. The learners, after having been tested for learner specific aptitude profile using Carol and Sapon's MLAT and Pimsleur's LAB, were matched to one of the three language teaching methodologies: audio-lingual, analytical or functional. Subsequent L2 success was tested using a standardised achievement test. The results were generally encouraging as it turned out that the groupings worked well. Wesche concluded her analysis by saying:

> *From these results it appears that the PSC system of matching students with methodological approaches, using language aptitude test profiles as basic criteria, encourages positive attitude and enhances achievement among highly analytical students* (1981:137).

Skehan (1986c) in his study of British armed forces personnel learning Arabic in an intensive, ten-week course applied cluster analysis to a set of scores obtained from a battery of aptitude tests and communicatively oriented end-of-course measures. His statistical analysis allowed him to isolate three groups of L2 learner profiles.

1. Generally unsuccessful: no particular strengths or weaknesses relative to achievement.
2. Generally successful: analytical-oriented – average memory, older learners, high in grammatical sensitivity
3. Generally successful: memory-oriented – good memory, younger learners, slightly above average grammatical sensitivity.

What all these studies show is that using multicomponentiality of FL aptitude along diagnostic lines proves to be much more relevant in educational terms than a mere prognosis which is interested more in how FL aptitude considered as a whole compares with other variables in prediction of L2 learning success.

1.3.3.2. Prognosis

The questions of prognosis of probable outcomes of second language learning constituted the main motive for searching for the most adequate and precise measures with the help of which success on language courses could be predicted. The approach adopted in order to maximise prediction was largely of a 'hit-or-miss' kind (cf. *section 1.1.*). The researchers would postulate a variety of factors, both cognitive and affective, and would assess their validity by looking at their joint or singular effect in predicting L2 learning success. With time, the abilities posited started to acquire more language specific integrity. However, the nature of these abilities was far from being explanatory of what the learning process was really about.

The critics of the concept of FL aptitude point to the fact that no universality can be claimed for the postulated construct which, clearly, is not to be interpreted in absolute terms. Rather, the evidence seems to suggest that the nature of the aptitudes that are isolated is dependent, to a certain extent, on factors such as: former language experience, type of L1 spoken, type and quality of instruction, formal vs. informal learning context, theories of L2 learning and/or teaching and, last but not least, research methodology and, specifically, the spectrum of abilities included for consideration[5]. What is interesting is that this confounding, as it might seem, evidence has had its reflection in models of FL aptitude that have been postulated in the last 50 years. Despite the uncertainty surrounding the precise definition of the abilities making up FL aptitude the search for its nature lead to the postulation of specific components of FL aptitude. Below, four FL aptitude models will be briefly presented.

Carroll and Sapon's (1959) Modern Language Aptitude Test (MLAT) is invariably associated with the concept of FL aptitude and is, by far, the best known and most widely used instrument to test the construct.

After an impressive number of 44 tests, based on the list of abilities thought to be important to L2 learning, had been correlated with the learners' achievement scores the authors factor-analysed their results and isolated four components of what was to become a standard view of FL aptitude. The postulated ingredients of FL learning ability were:

- <u>phonemic coding ability</u> – the ability to segment and identify (code) distinct foreign sounds, to form associations between them and graphemic symbols representing them so that they can be later remembered
- <u>grammatical sensitivity</u> – the ability to recognise the grammatical functions of words or other linguistic structures in sentences
- <u>inductive language learning ability</u> – the ability to infer linguistic rules and patterns from limited linguistic evidence
- <u>rote learning ability for foreign language material</u> – the ability to form associative links between form and meaning of a language material.

Oddly enough, Carroll and Sapon's MLAT, which consists of five parts, does not cover inductive learning ability (possibly because it seems similar to grammatical sensitivity) so, all in all, the test measures phonemic coding, grammatical sensitivity and mechanical memory. Despite being almost 50 years old now (it was designed at the times of audiolingualism), the four-component model of FL aptitude has withstood the test of time remarkably well.

Pimsleur's (1968) Language Aptitude Battery takes a roughly similar approach to the structure of FL aptitude, although Pimsleur's is less broad than Carroll and Sapon's in that he basically regards FL aptitude as being composed of a language analytic ability and an auditory ability. By including a question on the student's interest (Part 1) and by recording the student's grades in English, Mathematics, History and Science (Part 2), Pimsleur seems to be postulating that a cognitive factor of academic achievement and a personality factor of motivation/attitude are parts of FL aptitude. The remaining parts of PLAB are:

- Vocabulary (Part 3) – a knowledge component: a test of native language vocabulary
- Language Analysis (Part 4) – inductive language learning ability (rule inference)
- Sound Discrimination (Part 5) – an ability to attend to a new foreign language auditory material
- Sound-Symbol Association (Part 6) – a phonetic coding ability

MLAT and PLAB (or any other comparable test) have been the most frequently used tools in research and educational prognosis. Their predictive power expressed in multiple correlations with achievement measures under intensive training conditions with homogenous groups has stabilised between 0.40 and 0.65.

Skehan (1992, 1998) adapted the traditional MLAT structure of aptitude and suggested a three-factor model comprising the following:

- language analytic capacity
- memory capacity
- phonemic coding capacity.

Skehan's proposition can be viewed as an attempt to respond to the criticism levelled against MLAT to the extent that the model of FL aptitude that underlies it did not take into account new insights from the fields of cognitive psychology and linguistics (the nature of the language learning process), social psychology (the role of motivation, attitudes, personality etc.), educational psychology (individual cognitive styles, learning strategies), or applied linguistics (instructional methods or teaching methodologies).

Skehan's model did not produce a separate, new test of FL aptitude. Rather, he revamped the psychometric, four-component model arrived at on purely experimental basis by reformulating it in information-processing terms. Skehan managed to link his three-component view of FL aptitude to stages within a flow of information processing thus lending theoretical justification to the Carroll and Sapon's traditional model. This is summarised in *Table 1.1.* below.

Aptitude factor	Stage	Operations
phonemic coding	input	noticing
language analysis	central processing	pattern identification
		generalisation
		restructuring
memory	output	retrieval

Table 1.1. Aptitude and processing stages (after Skehan 1998:203)

In a section devoted to his three-factor model of FL aptitude Skehan concludes:

> *Aptitude tests are not simply predictive instruments which have no explanatory power. They have a rationale in relation to language learning processes, and then an even wider justification through general psychological processes. They are not arbitrary in any sense, even though the successful sub-tests from aptitude research have emerged rather inductively* (1998:205).

He also argues that his refurbishing of the Carroll and Sapon's traditional view on FL aptitude apart from rekindling interest in the concept itself lends it a cutting edge in the sense that FL aptitude expressed in information processing terms transpires as important for both, formal and informal learning contexts (Skehan 1998). Recently, a new and very interesting battery of aptitude tests that echoes Skehan's suggestion has been created and published commercially by a team of researchers from the University of Wales (Meara, Milton and Lorenzo-Dus 2001). The test (*Language Aptitude Tests*) is computer-based and includes the following five tests of FL aptitude on a CD-ROM:

- LAT A - a test of an ability to repeat unfamiliar language sound strings presented aurally; immediate repetition; short-term memory span for auditory material
- LAT B - a paired-associate test of visual memory for words
- LAT C - a test of inductive learning
- LAT D - a test of an ability to recognise unfamiliar words presented aurally
- LAT E - a test of the ability to form sound-symbol associations

The five tests address the following traditional components of aptitude:

phonemic coding:	LAT A, LAT E, and (roughly) LAT D
language analysis:	LAT C
memory:	LAT A and LAT B

Although the test's posited abilities look familiar enough, there are two things that distinguish it from earlier tests. Firstly, the test follows Skehan's three-component model of FL aptitude.

Secondly, it attaches more importance to the peripheral stages in the processing of language as seen by the number of tests assessing relevant abilities (input and output). Thirdly, thanks to the computer-format, the test truly simulates the learning situation as it would happen outside the test: test-takers learn as they go through the task (except for LAT A, all the parts contain elements of learning-while-doing). Fourthly, LAT is (potentially) simultaneously available in various language versions (at the moment: English, French and Hungarian).

Grigorienko, Sternberg and Ehrman (2000) developed an entirely new test of FL aptitude based on a particular cognitive theory of FL acquisition the so called CANAL-F Theory (**C**ognitive **A**bility for the **N**ovelty in **A**cquisition of **L**anguage – **F**oreign). The main tenet of the theory, as based on acquisition processes rather than on products of these processes, is that: "one of the central abilities required for FL acquisition is the ability to cope with novelty and ambiguity" (Grigorenko et al. 2000:392). The authors propose the following components of their CANAL-F Theory: 1). Knowledge Acquisition Processes; 2). Levels of Processing; 3). Modes of Input; 4). Memory Processes.

The processes that the learner makes use of during knowledge acquisition include the following:

- selective encoding – to distinguish between the relevant and irrelevant information in the stream of incoming data
- accidental encoding – to encode less important, secondary information outside the learner's conscious focus
- selective comparison – to determine the relevance of old information for the needs of a current task
- selective transfer – to apply decoded/inferred rules to new contexts/tasks (analysis and induction)
- selective combination – to synthesise new information with the old one.

In FL learning those five processes operate at four levels of processing: lexical, morphological, semantic and syntactic, (phonological? JR), in the visual or oral modes of input and rely on immediate and delayed memory processes.

The CANAL-F aptitude test was created to address and gauge the above mentioned four sources of individual differences. According to the authors, the test has at least two features that make it unique when compared to other traditional tests of FL aptitude. The test is designed to simulate a continuous learning situation and it taps the processes of knowledge acquisition at the time of the test. A comprehensive description of the test would go beyond the scope of the present brief review (cf. Grigorenko et al. 2000 for examples of tasks).

CANAL-F test was construct-validated against the following measures: Cattell's Culture Fair Test of *g* (Cattell and Cattell 1973); Terman's Concept Mastery Test; MLAT and prior language experience questionnaire. The test's construct validation yielded two broad first-order factors: an intelligence-related factor and a language-specific factor. While all the CANAL-F subtests loaded equally highly on the two factors, only two of MLAT's subtests (*Paired-associates* and *Spelling clues*) loaded on the language specific factor.

The study of the criterion-related validity of CANAL-F test against the subjects' (N= 63) instructors' ratings of relevant language skills did not produce outstanding improvement in the magnitude of correlation coefficients: ***r***=0,40. In this connection Ellis remarks:

> *While correlations with measures of language learning were of the same order as those reported for the MLAT, the test does afford the possibility of achieving a closer match between specific aptitudes and specific psycholinguistic processes* (2004:533).

If further research confirmed the theory then, as the authors argue, it would lead to a view of language aptitude as a form of developing expertise rather than a stable and largely fixed concept.

1.4. Second language proficiency: unitary or componential ?

The controversy between the proponents of the 'unitary competence' versus the 'completely divisible competence' hypotheses on the structure of SLP did not start until the publication of Oller's (1978) 'Pandora's box' article in which he argued, following Spolsky's (1973) rejection of the skills/components view of SLP, for a single global trait that included cognitive processing, rather than several distinct traits as proposed earlier by Lado (1961) and Carroll (1968) (cf. Purpura 1999).

Bachman and Palmer's (1981) article terminated the debate between the two extreme views on the organisation of L2 language skills. The authors conclusively demonstrated that the unitary trait hypothesis of language proficiency must be ruled out. In their study of native Mandarin Chinese speakers of English (N=75) Bachman and Palmer (1981) used the multi-trait multi-method procedure to construct-validate the FSI (Foreign Service Institute) Oral Interview examination. They found evidence for the existence of two distinct factors: speaking and reading, and concluded that:

> *(...) the model with two correlated traits is preferable to the model with a general factor and two distinct traits, since it is more parsimonious and provides factor loadings which are more interpretable and more consistent with the results of other analyses* (1981: 85).

The study's findings, however, were limited by the highly homogenous nature of the subjects' first and second language experience as well as their educational background.

Almost two decades prior to Bachman and Palmer's (1981) paper, Gardner and Lambert (1965), in their complex study of language aptitude, intelligence and language achievement, found evidence for the multidimensional model of foreign language ability. They identified three separate language proficiency factors, from seven posited, which they labelled:

1. *Linguistic Reasoning* – loaded by five aptitude subtests of MLAT and by achievement tests of grammatical knowledge, phonetic discrimination and reading comprehension[6]
2. *French Vocabulary Knowledge* – loaded by various vocabulary tests, one listening comprehension and one reading comprehension test
3. *School French Achievement* – loaded by a variety of teachers' evaluation of French achievement (Gardner and Lambert 1965:197).

Methodologically, Gardner and Lambert's (1965) study used factor analysis and the centroid method of factor extraction continued until little more common variance was left in the residual matrix. No hierarchical analysis was applied which could have yielded second-order factors likely to be interpreted as general proficiency factors.

Vollmer and Sang (1983) in their comprehensive survey of studies dealing with the problem of SLP structure conclude the section devoted to the studies supportive of the divisible hypothesis with the following remark:

> *The divisible hypothesis in its strong form, based on structuralist assumptions about the independence of the elements of a language, has already been replaced by the weaker form of this argument in most of the studies, namely, that at least the four integrated skills should be related to more or less independent dimensions of language ability* (1983:53).

The authors convincingly demonstrate by discussing numerous studies in support of the two extreme views of the structure of SLP that: "According to our critical assessment there seems to be no substantial support for either of these two extremes" (Vollmer and Sang 1983:68). Additionally, the authors note in passing that if language learning abilities are conceived of as an aggregate of skills all contributing to L2 learning success then it is somewhat of a paradox to claim that the product of these aptitudes should be one, unitary ability (cf. *section 4.2.2.*).

The introduction of more sophisticated statistical methods into applied linguistics in the 1980's and 1990's (hierarchical factor analysis, confirmatory factor analysis, structural equation modelling and others) offered a better explanation of the data. Specifically, the hierarchical structure of SLP, being a compromise between the 'unitary' and the 'divisible' hypotheses, gained wide support. Studies by Bachman and Palmer (1983), Carroll (1983), Upshur and Homburg (1983) and Sasaki (1996) lent support to the view of the general language factor(s) (as reflecting overall degree of advancement in different language skills) and a number of specific, lower-order factors associated with it and with one another. The nature of the specific, lower-order factors differed from study to study. The most common dimensions identified by researchers included factors described as: *active* versus *passive* performance (Carroll 1983); *structure/vocabulary* versus *integrated skills* (Oller and Hinofotis 1980); *literacy skill* (reading and writing) and *oracy skill* (listening and speaking) (Spearritt, 1979 quoted in Carroll, 1983:102); *composition* (writing), an ability to handle language in *short context* and an ability to effectively solve problems using vocabulary knowledge in a relatively *long context* (Sasaki 1996).

1.5. Intelligence versus L2 proficiency and FL aptitude

The results of research from the last 40 years on the relationship between, on the one hand, intelligence and L2 proficiency (von Wittich 1962; Gardner and Lambert 1965; Oller 1979, 1983; Genesee 1976, 1987; Flahive 1980; Boyle 1987 and Sasaki 1996) and, on the other, between intelligence and FL aptitude (Gardner and Lambert 1965, 1972; Wesche et al. 1982 and Sasaki 1996) is supportive of the 'compromise' view, namely such that while the two (intelligence vrs proficiency/aptitude) are related by no means are they identical.

1.5.1. Intelligence versus second language proficiency

The first group of studies examining the relationship between intelligence and second language proficiency (SLP) is briefly discussed below.

Von Wittich (1961) using multiple regression analysis found that IQ as measured by the *Ottis Mental Ability Test* was the poorest predictor of various criterion measures (school grades) in her sample of ninth grade students (N=230) with the correlation of 0.48 between IQ and the criteria. Gardner and Lambert (1965) who studied high school students (N=96) demonstrated that intelligence measures (Thurstones' *Primary Mental Abilities Test*) are relatively independent of L2 language skills as measured by a variety of French achievement tests. The correlations they report range from -0.17 to 0.49 (ibid. p. 194). Genesee (quoted in Spolsky 1989) suggested that there is a strong relationship between the scores in intelligence tests (measured by the *Canadian Lorge-Thorndike Test of Intelligence*) and formal tests of grammar as opposed to the learning of communication skills. Reported coefficients were 0.56 versus 0.39 for the respective skills. Flahive (1980) studying the relationship between non-verbal intelligence and SLP measures reports the correlations from 0.59 to 0.84. (quoted in Sasaki, 1996). Boyle (1987) in his factor analytical study of intelligence, reasoning and language proficiency also isolated distinct factors for language proficiency and reasoning abilities.[7] Sasaki (1996), applying confirmatory factor analysis, arrived at a hierarchical structure of her SLP, intelligence and FL aptitude variables and reported a correlation of 0.65 between her second-order General Cognitive Factor (consisting of first-order FL aptitude Factor, Verbal Intelligence Factor and Reasoning Factor) and General SLP Factor (consisting of the study specific, first-order factors).

In the present study the relationship between intelligence and L2 proficiency as evidenced in the correlation between the non-verbal intelligence test of Raven Progressive Matrices and the results on the criterion KET-A Test for the sample of 13-year-olds ranges from 0.38 for N=138 (cf. *Table 3.5.*) to 0.48 for N=200 (cf. *Appendix 1*), which is much lower than that reported by Flahive (1980) for the same non-verbal intelligence measure, but very much similar to the values reported in the remaining studies.

1.5.2. Intelligence versus FL aptitude

Below the results of studies examining the relationship between intelligence and FL aptitude are summarised. Gardner and Lambert (1965) demonstrated that FL aptitude (measured by MLAT) and intelligence (measured by PMA Test) although correlated (*r* from 0.06 to 0.38) were still relatively independent of each other. Wesche et al. (1982) using principal axis factor analysis followed by hierarchical solutions, examined the relationship between intelligence (Thurstones' PMA) and aptitude (MLAT) and found moderate-to-high correlations between them with the reasoning sub-test correlating the highest at 0.65. They also found three relatively independent factors which best accounted for the data: Verbal Knowledge, General Intelligence and Memory-within-Aptitude-Batteries. Hierarchical analysis, following Schmid and Leiman procedure (cf. *section 3.4.*) led them to postulate a higher-order general factor indicating that, at a more abstract cognitive level, aptitude and intelligence are related (Skehan 1989). Sasaki (1986) posited three separate but related first-order factors: Verbal Intelligence, Reasoning and Aptitude which are related to a higher-order General Cognitive Factor, the correlation coefficients being: 0.59, 0.56 and 0.82 respectively.

With respect to the relationship between intelligence and FL aptitude the results obtained in the present study are generally in line with all major research which has investigated the issue. The correlation between FL aptitude as expressed by a combined aptitude measure: *FL aptitude* (cf. *Table 3.1.* and *section 2.7.4.2.*) and non-verbal intelligence as measured by *SRPM* is 0.53 (cf. *Appendix 1*). This figure is lower than the correlation between the same aptitude measure and a test of verbal intelligence (*AVoc*) which is 0.70.

Somewhat in the margin of the current discussion the following methodological remark suggests itself in connection with the section's main topic. Namely, it is very interesting to compare how different authors understand and/or interpret other researchers' results of aptitude studies when they report them in their own publications. Bogaards (1991) and Skehan (1998) both reporting a study by Wesche et al. (1982) will serve as instructive examples. Concluding a section devoted to the notion of aptitude and commenting on Wesche's et al. correlation coefficient of 0.67 between MLAT and Thurston's test of Primary Mental Abilities (PMA) Bogaards states:

> *Factor analysis yielded a first-order factor which could be interpreted as 'general intelligence', and then three second-order factors of which the third could be interpreted as 'memory'. These data suggest that what [MLAT's] subtests exhibit is not just a simple relationship of overlap between the sub-tests, but that what the MLAT Total score measures is, first of all, intelligence. And the only specific factor that emerges is the one measuring mechanical memory. All in all, it must be stated that there is a connection between intelligence and aptitude but also that it is very difficult to situate one notion against the other with any degree of exactness* (1991:36; own translation).

Skehan, on the other hand, summarising his brief discussion of four studies concerned with aptitude-intelligence relationship remarks the following about the same study by Wesche et al. (1982):

> *It is also worth noting, in passing, that both Gardner and Lambert (1972) and I (Skehan 1982) report data on fairly unselected groups, while Wesche et al. (1982) and Sasaki (1991) report on subjects who were streamed in one way or another, and are not so representative of the population in general. (...). A further point to note is that the studies suggesting a stronger relationship between aptitude and intelligence require second-order factor analysis to bring this out, in each case reporting first-order factors which suggest some degree of separation. One can conclude, provisionally, that aptitude is not completely distinct from general cognitive abilities, as represented by intelligence tests, but it is far from the same thing* (1989: 209).

The difference is quite striking. Despite seemingly similar conclusions reached by the two authors, their understanding of the significance of the results and the limitations of the methodology adopted reveals that they completely differ in the interpretation of the same findings. Bogaards (1991) stresses the importance of the second-order level of factor analysis and is willing to take it as proof of virtual identity of aptitude subtests (MLAT) with intelligence measures (PMA) as demonstrated by the emergence of 'a general intelligence' second-order factor. For him FL aptitude is identical with general intelligence. Skehan (1998), on the other hand, by pointing out two possible limitations of the studies that report stronger relationship between aptitude and intelligence (sampling and hierarchical analysis) arrives at more tentative conclusions than Bogaards (cf. quote above). This difference in the perception of the relevance of a finding might itself be attributed to idiosyncratic predilections of a researcher for a particular hypothesis and/or explanation of empirical facts. It follows from further reading of Bogaards (1991) that he is much in disfavour of claiming any role for an aptitude factor in his model of L2 learning (cf. *Figure 6.2.* in Bogaards 1991:153).

1.6. FL aptitude research in Poland

In Poland the literature on the role of ability factors as determinants of learning success is scanty, and research on FL aptitude has practically been non-existent. The aspects of individual differences researched in Poland, and briefly reviewed in Komorowska (1987) include: motivation, self-evaluation, anxiety, speaking readiness, personality types and learning strategies. In particular, the latter aspect of individual learner differences has attracted much of researchers' attention in Poland over the last twenty years or so. Droździał-Szelest (1997) is a good example of an exhaustive presentation of the problem and a discussion of the research done by the author in the beginning of the 90's. The little that has been done so far in the area of cognitive factors is briefly reviewed below in chronological order.

Ozga and Tabakowska (1973) report a study in which candidates seeking admittance to the School of English at Jagielonian University in July 1972 took a test of language rule induction in addition to a regular proficiency entrance test (written and oral). The authors report a medium correlation between the aptitude test and achievement score of 0.60 and, somewhat surprisingly, state that the figure is low. Nowhere in their article do Ozga and Tabakowska state how numerous the sample was, what the *p* level is nor what type of correlation they use (rank or Pearson's). For this reason, it is practically very difficult to interpret their findings. Tabakowska and Ozga (1975) is, basically, a replication of the first study with a different test of inductive language learning component of FL aptitude. In conclusion the authors postulate that in order to prognosticate candidates' success with language at the English Studies a test of aural-oral language learning ability should be administered to candidates. They also note that prognosis should take place much earlier than the entrance exam stage to university.

All in all, it seems that, given the highly selected population that usually seeks admittance to university studies, what Tabakowska and Ozga'z tests did measure was prior experience with language learning rather than an ability to learn a language.

In the years 1969-1974 Lewowicki (1975) studied a very large representative sample (N=1820) of primary school learners in grades from five to eight (11-14 years of age) in 28 schools across Poland. The range of variables included in the study is impressive: social (parents' education, living conditions); affective (anxiety, motivation, personality type) and cognitive (intelligence, arousability). Lewowicki was concerned with how the individual difference variables used by him affected learners' school achievement as measured by school subject tests. L2 learning (Russian) did not receive any separate treatment and because no FL aptitude test was included his research can be described as focusing on general cognitive abilities. He reports a correlation of 0.36 between Raven Standard Progressive Matrices test and the criterion, which fits well with the results of other studies mentioned earlier in the chapter.

Niżegorodcew (1975, 1979) in a series of two articles reports on a longitudinal study carried out on a group of secondary school learners over the period of three years (1974-1977) in intensive and non-intensive L2 English language groups. The number of 78 subjects, with whom the author started, fell to 53 at the end of the study. The tests of the following abilities were considered for inclusion in the study:

1. inductive language learning
2. sound discrimination
3. grammatical sensitivity
4. paired-associate learning of words (Hungarian)
5. aural sequential memory – ASM (immediate recall of digits)
6. visual sequential memory – VSM (immediate recall of sequence of patterns)
7. immediate recall for meaningful verbal material
8. native vocabulary (synonymy)
9. sentence-completion (words-in-context)
10. sound-symbol association.

In the process of data acquisition tests 3 (*grammatical sensitivity*), 5 (*aural sequential memory*) and 7 (*immediate recall*) were removed from further consideration. The reasons the author gives for this are not entirely clear (cf. Nieżegorodcew 1979:33-34). The final battery then, sampled basically the same repertoire of abilities known from MLAT and/or PLAB with three abilities (*sound discrimination, ASM and sentence-completion*) rarely included in aptitude tests. Nieżegorodcew's findings are summarised below:

a. in an intensive group language analysis (*inductive learning*) had the highest *r* (0.62), while in a non-intensive group it was *rote memory* (0.22)
b. three highest correlations in an intensive group (*language analysis* 0.62, *sentence completion* 0.42 and *vocabulary* 0.34) displayed no correlation in a non-intensive group
c. Raven's test of non-verbal inductive abilities correlated at 0.28 with the English grade in an intensive and at – 0.13 in a non-intensive group
d. GPA (Polish, Mathematics, History and Chemistry) correlated at 0.66 with the test in an intensive and at 0.56 in a non-intensive group.

Analysing the results of her study she concluded that:

e. aptitude plays a role in intensive training only
f. sound discrimination is not an important ability in FL learning
g. the role of aptitude increases with learning
h. success in a FL correlates highly with general academic achievement

i. the best predictors of L2 success are: 1.) tests of paired-associate; 2.) VSM; 3.) language analysis; 4.) sentence completion

There are a few doubts that arise from her analysis. First of all, it is not at all clear why paired-associate and ASM are ranked among the highest predictors considering that language analysis has higher values of *r*, and no analysis is reported which would enable the author to gauge predictive powers of individual tests. In the light of this, the conclusion in 'i' seems unwarranted. Secondly, given that Raven correlates weakly with criterion English (cf. *'c'* above), and that criterion English correlates significantly highly with GPA (cf. *'d'* above) it would be insightful to find out what the correlation is between Raven and GPA. This however, is not reported. It might as well be the case that the analysis of partial correlations between the three variables (English, GPA and Raven) would reveal that it is the Raven test that accounts better for the variance in the GPA score. Without this analysis the conclusion in 'h' is trivial.

Finally, the very research design seems to be faulty in that the sample is small and, more importantly, the tests making up the battery were given to subjects in three successive test administrations over a period of three years, which must raise questions as to the reliability and validity of the results thus obtained. Despite all the insufficiencies, Niżegorodcew's research is, by far, the only serious and relevant attempt to address the issue of FL aptitude in its full right.

Figarski (1984), in his research of factors contributing to success in the learning of L2 Russian carried out in the years 1977-1979, studied a representative sample of primary (11-14 years of age) and secondary (15-18 year old) school learners (total N=806). As his independent variables Figarski considers the whole gamut of factors believed to influence L2 learning. He groups them into three categories (original order maintained):

1. <u>learner variables</u>: motivation, general intelligence, visual sequencing, musical ability/auditory acuity, general knowledge, attitude, self-evaluation
2. <u>social variables</u>: parents' education, economic status, degree of parental supervision of the learning process
3. <u>teaching process variables</u>: teacher personality, teaching methods (1984:49-51, own translation)

As criterion variables Figarski used a series of achievement tests constructed in co-operation with the school teachers. Reported psychometric properties of the tests (reliability and discrimination indices) are quite satisfactory (cf. Figarski 1984:52). The tests included:

- listening comprehension, reading comprehension, speaking and writing
- L2 vocabulary, L2 grammar,
- L2 orthography, reading aloud

- reproduction of sounds and words
- recognition of intonation and accentual patterns

There is a remarkable lack of FL aptitude factors in Figarski's group of learner variables. The only factor which could be claimed to be of importance for FL aptitude is musical ability/auditory acuity. However, because Figarski does not include the test, and because his description of the abilities the test measures are confounding[8] it is practically very difficult to see the importance of his findings for ability research.

The other cognitive factor investigated by Figarski is intelligence measured by the Raven test. The results (correlations with the criterion) are available only for the following group of learners: Grade five (N=102): 0.38; Grade seven (N=123): 0.56 and Grade nine (N=88): 0.75. While the first two correlations are within the range usually reported in this type of research, the last figure is astounding as no study is known to the present author in which this magnitude of association between L2 achievement and Raven test has been reported. But then, if the relatively small sample is taken into consideration (N=88), and the fact that nothing is known about the version of the test used the figure looks less disquieting.

Jankojć's (2002) unpublished doctoral dissertation is the most recent attempt to address the problem of individual differences in learning a second language. Not surprisingly, he also totally disregards FL aptitude in the list of factors he deems to be relevant to L2 learning success. In his study of 1st (N=62) and 2nd year students (N=56) learning English in the academic context, Jankojć includes an impressive list of factors on which he gathers data using the following measures:

a. general intelligence - four sub-tests of Wechsler verbal scale: Knowledge, Similarities, Comprehension, Vocabulary
b. personality - self-portrait questionnaire
c. speaking readiness
d. motivation - two measures
e. L2 self-evaluation test
f. teaching process variables
g. attitude scale
h. L1 word fluency
i. L2 word fluency measure.

The results obtained from these measures formed his independent variables, which were then independently correlated with the results from the end-of-year proficiency criterion tests[9].

With reference to the only cognitive factor used in the study which is of interest in the context of the present discussion i.e., verbal intelligence, Jankojć reports the following correlations between his Wechsler total score and L2 proficiency total score: 1st year students: 0.12; 2nd year students: – 0.17. The results are intriguing as they indicate that there is basically no correlation between FL proficiency and general intelligence measured by the four sub-scales of the Wechsler test. The reader is even more confounded when, in the last paragraph of the section (ibid. section 6.7.5 p. 318), it becomes clear that the test of general intelligence was correlated with Jankojć's measures of the spoken language rather than with the general L2 proficiency grade (which includes a test of grammar and of the four skills). This however a reader has to puzzle out by himself. While the low values of the indices obtained become clearer, one wonders about the author's rationale to seek relationship between the test of non-verbal inductive reasoning and speaking. All in all, Jankojć's research is beset with a few problems connected with research design, methodology used, validity and reliability of some of the testing tools and the statistical analysis employed[10]. And, what is even more important, the non-inclusion of FL aptitude measures in the range of independent variables leaves serious doubts as to the validity and generalisability of the results.

Other authors writing on the subject of learner differences in L2 learning or on the conditions determining the success in L2 learning either ignored FL aptitude altogether (Prejbisz 1973; Komorowska 1987) or mentioned it as an addition to 'better' predictors of L2 learning success (Komorowska 1984).

All in all, an examination of the available research and writings on cognitive variables in L2 learning in Poland leads one to conclude that the area has been almost completely ignored and disregarded, or at best the importance of the FL aptitude factor has been considerably downplayed in individual difference studies.

NOTES

1. Brandon (1985), discussing the issue of aptitude-ability distinction from a philosophical perspective, notes that apart form the ease and speed of acquiring a skill which are implicated in the idea of aptitude, other possible characteristics, such as accuracy and length of retention, can be equally well envisaged, though they have hardly been researched.
2. Studies on attitudes and language learning in children by Hermann (1980), Strong (1984) and others have demonstrated a positive influence of learning /acquiring a skill in language on attitudes towards the process of learning itself (McLaughlin 1995)
3. The categorisation used here is never so clearly cut. For example, learning style has both a cognitive and an affective dimension; motivation, on the other hand, is more of an affective than a cognitive factor while anxiety is predominantly an affective variable (Ellis 2004)
4. The Bristol Language Project examined the first language development of ca. 130 children aged between 2 and 10. Ten years later a follow-up study was conducted on 100 children during which their FL aptitude was assessed (Skehan 1986b)
5. Skehan, commenting on the use of factor analysis as a standard technique in aptitude research, remarks: "What is included, and what is not included in the analysis will have an immense impact on the likely results. Similarly, how the results [isolated factor loadings] are interpreted will also need careful scrutiny(...)" (1991:19). This suggests that considerable variety of possible aptitudes can be expected as there seems to be no 'general' aptitude to learn a language anywhere and anyhow but always an aptitude to learn a language in a particular setting, using a particular language pedagogy and with a particular idea of what it means to know a language i.e. with a model of second language proficiency structure. Although this all-sweeping generalisation seems to deny the validity of the concept, there is still room for the independence (relative) of the construct of aptitude from other, apparently important, but by no means identical, variables of motivation, anxiety and/or learning strategies (cf. Carroll 1968, 1981 and *section 1.3.1.*).
6. Interestingly, Gardner and Lambert's (1965) *Linguistic Reasoning Factor*, loaded by five subtests of MLAT and three proficiency subtests, points to an overlap between SLP and FL aptitude and indicates that while the two are distinct constructs they are still correlated. The data under consideration revealed a statistically insignificant merger between SLP, on the one hand, and FL aptitude and intelligence, on the other (cf. *section 4.3.*).
7. Unfortunately, Boyle does not provide correlation coefficients between the factors so it is impossible to judge whether the factors are orthogonal or oblique. If the latter is the case, then a hierarchical analysis might extract a new dimension to that suggested by Boyle.

8. At one point he describes the test as measuring the following abilities: imitation of individual sounds, simultaneity of sounds, listening (!?) and the perception of rhythm and tone (Figarski 1984:53). In reporting his results, however, he refers to the ability to recognise speech sounds, which he never mentions in his variable characteristic (ibid.87).
9. Students' proficiency was assessed through written as well as oral tests. It is the oral test that is most bewildering. Spoken language was assessed by Jankojć using the following five criteria:
   - tempo of delivery (number of words per minute)
   - mean length of a sentence measured by the mean number of words per sentence
   - length of the oral production (in minutes and seconds)
   - scope of L2 vocabulary
   - subjective judgement of the communicative aspect of the oral presentation

   It must be admitted that the list of variables constitutes a most unusual selection of criteria used in the assessment of spoken language. Jankojć gives no justification for the abandonment of, as he puts it: "the widely used method by Harris (Harris 1969) consisting in assessing individual aspects of spoken language on the scale from 0 to 5" (Jankojć 2002:214, own translation).
10. Jankojć relies exclusively on simple, bivariate correlational analysis of the relationships between individual variables taken one at a time. A much more powerful statistical tool of multiple regression would have allowed him to consider a simultaneous influence of all his variables in the prediction of L2 learning success.

CHAPTER 2

METHODOLOGICAL ASPECTS OF THE STUDY

This chapter is devoted to the presentation of the methodological aspect of the study. In *section 2.1.* a general research methodology of the present study is outlined thus helping to situate it in a particular approach to language research. *Section 2.2.* discusses sampling procedures and characterises the participants of the pilot stage and the study proper. In the next three sections the tests instruments used in the data collection stage of the study are described; in *section 2.3.* intelligence measures, in *section 2.4.* foreign language aptitude measures and in *section 2.5.* second language proficiency measure(s) used. Data collection and test administration procedures are discussed in *section 2.6.* which is followed by *section 2.7.* devoted to the presentation and description of the statistical tools used in the process of data analysis.

2.1. Research design of the study

The present study is grounded in the tradition of positivist philosophy which, in its epistemology, stresses the importance of observation, hypothesis testing and verifiability as tools in knowledge acquisition. In its data-acquisition and data-analysis process the study relies on objectivity and falsifiability. Because the aim of the study is to accumulate a substantial body of 'hard data' with a view to analysing it in the process of objective hypothesis verification the study is quantitative in nature.

The present study is also *ex post facto* as it seeks to "investigate possible cause-and-effect relationships by observing an existing condition (...) and searching back in time for plausible causal factors" (Cohen and Manion 1995:146). Typically, a truly *ex post facto* design investigates the effects of an independent variable or, more likely, variables, on a dependent variable. In this type of study the researcher works from the dependent variable - the only available, tangible 'proof' of the workings and/or influences of the independent, 'causal' variable(s). In other words, the independent variables lie outside a willed control of the researcher who has to limit himself to identifying 'causal' relationships in retrospect.

In the broader context of possible applications of an *ex post facto* design the independent variables will "have already occurred" (Cohen and Manion 1995:147) and thus are difficult to control through experimental designs. In the context of psychological or educational research the independent variables might be the learners' individual characteristics and/or traits (both cognitive and affective) which they bring into an L2 learning situation: FL aptitude (or rather aptitudes – as explained in *Introduction*), field in/dependence, memory capacity, anxiety levels, motivation, etc. The observable (dependent) variable studied for the effects of the independent variables would typically include some sort of a variability-displaying criterion in the performance levels (achievements) of individual learners. In the context of the present study the independent variables are: the postulated construct of FL aptitude as measured by the TPJ and PLAT tests, verbal intelligence skills as measured by Jurkowski's TJL (cf. *section 2.4.1.*) and intelligence in the form of eductive abilities (cf. *section 2.4.2.*) as measured by Standard Raven Progressive Matrices, Classical Form.

The present study is **descriptive**, as opposed to **experimental**, to the extent that it seeks to describe and quantify (using an array of analytical techniques) naturally occurring phenomena without experimental manipulation. As the investigation of the learners' language behaviour through the data gathered is done after the conditioning (independent) variables have revealed themselves to the researcher, the *ex post facto* design can be said to be both **theory-driven** (a feature of most experimental designs) and **data-driven** (a typical trait of a descriptive research). It differs from the former in the sense that the researcher does not begin with a specific and detailed question and/or interpretation. Rather he starts off with some general idea of what to expect from the internally derived data as opposed to determining whether the data derived externally through controlled manipulation in an experimental setting lend support for the initial hypothesis (Seliger and Shohamy 1990). By the same token, it differs from the data-driven research in the sense that it is not purely exploratory in nature, which can be seen in the inclusion of a limited set of ability-testing tasks in the two measures used (TPJ and PLAT), which reflects the authors' hypothesis as to the nature of the abilities making up FL aptitude.

At the same time, because the researcher can only focus on certain aspects of the data (the research question having been decided in advance) the study is more **analytic** than **synthetic** (holistic). Consequently, one of the limitations of the present study is that it only achieves the level of description and prediction through exploration rather than explanation. However, when carefully designed, an *ex post facto* research is likely to yield plausible explanations of the causal relationships among both observed and latent variables as Bachman (1989) argued.

The present study uses the following multivariate techniques of data analysis typical for *ex post facto* design: **exploratory factor analysis** (**EFA**) and **multiple regression (MR)**. Factor analysis (**FA**) is an analytic procedure which enables the researcher to reduce large sets of data to more manageable and less numerous factors which, it is hoped, underlie the original data. But data reduction is not the most important feature of FA. Through relating quite a large set of independent variables to one another, rather than to a dependent variable, FA operates on a multidimensional (multivariate) correlation matrix of each independent variable with any other used in the study. By examining the interrelations between a set of independent variables FA attempts to unravel an underlying pattern in the form of new dimensions (referred to as factors). As such it is exploratory in nature.

On the other hand, the analytical technique of multiple regression (**MR**), is used to "analyse the variability of a dependent variable (...) in relation to the variability in (several) independent variables" (Rietveld and Hout 1993:71). Using MR then, it is possible to gauge relative 'force' or contribution of each independent variable in the prediction of the dependent variable. Consequently, by combining a few independent (predictor) variables regression analysis hopes to achieve maximum prediction which any one of the variables would not achieve if used singly. More detailed discussion of the two data-handling techniques is offered in *Chapter 3*.

In the present study the constructs of two language aptitude batteries: TPJ and PLAT are validated using exploratory factor analysis *(section 3.6.)*. Then the results of another factor analysis run on the combined scale of TPJ and PLAT are entered, together with two remaining intelligence variables, into multiple regression analysis *(section 3.9.)* to seek the highest level of prediction among independent variables employed in the study.

2.2. Participants

2.2.1. Participants of the pilot study

As already mentioned in the *Introduction*, four measurement tools were used in the study proper to obtain data for the independent variables from the sampled population of 13-year-old L2 learners. The results of the following tests formed the study's independent variables:

A. Standard Raven Progressive Matrices (SRPM)

B. Verbal Intelligence Test by Jurkowki (TJL)

C. Language Aptitude Test by Kuliniak (TPJ)

D. author-designed Progressive Language Aptitude Test (PLAT)

Among the four measurement tools only the first three scales possessed stable psychometric properties. They had been extensively tried out in large-scale standardisation studies and in the case of two measures: TJL and SRPM both convergent and discriminant levels of test validity were established with relation to other scales purportedly measuring the same as well as different constructs. Consequently, it was assumed that the tests were reliable and valid measures of the constructs they aimed at (cf. *section 2.3.*)

As regards the fourth measure (PLAT), which was designed and developed by the present author it had to be put to test first before it could claim satisfactory psychometric properties and as such be used as a valid scale in the present research. In order to ascertain PLAT's reliability it had been piloted in a series of two successive applications of its preliminary versions in the year 2001 and 2002. The participants of the pilot studies were 6 Graders seeking admittance to one of the secondary schools in Poznań both at lower and upper level. The school, where English is the language of instruction, runs an intensive L2 English course (ca. 8 hours a week) for all its learners. Depending on the Grade, from 40% to 80% of the major 6 school subjects' instruction is in English. The fact that a candidate's prospective admittance to the school was contingent on the test, was a powerful motivating factor which provided for maximum concentration and effort invested in the test. Both trials took place in April on the school's premises. The test was administered and scored by the present author. There were 28 participants in the 2001, and 81 in the 2002 pilot group. Whereas the first trial group was homogenous as to age, the latter consisted of 34 candidates to the lower and 47 candidates to the upper school. The youngest candidate in the younger group was 12,5 and the oldest 13,1. In the older group, which was on average 3 years older, the figures were 15,4 and 16,0 respectively. There were no striking differences in the ratio of boy/girl candidates and the relevant percentages were:

2001 pilot –		F: 50% and M: 50%
2002 pilot –	younger group:	F: 55% and M: 45%
	older group:	F: 47% and M: 53 %.

In both trials 90 per cent of the candidates came from Poznań and the remaining 10 per cent from the schools around the city. No significant differences in the test scores were recorded between the two sub-samples. The results of the PLAT scale refinement procedure which involved item trimming for reliability enhancement are discussed in *section 2.7.* of the present chapter.

2.2.2. Participants of the study proper

A total of 242 grade 7 pupils from seven lower secondary schools *(gimnazjum)* in the Poznań area were involved in the study proper. All the pupils were in their 1st grade of a 3-year lower secondary school. All were 13 years of age at the time when the first of the measures was administered to them in early October 2002 (the youngest was 12;11 and the oldest 13;9). The number fell to 227 after the elimination of 15 pupils whose performance on any of the measures employed was judged as not-independent, neglectful or defyingly sloppy. In other words, pupils who cheated, displayed lack of interest and/or showed consistent signs of an inability to stay on task were removed from further consideration (cf. also *section 2.6.* on test administration procedure). The performance of the remaining 227 pupils on all the measures was considered sufficiently reliable for the purpose of the study.

At this point a qualification regarding the number of participants must be made. The five tools which provided data for the present study (four independent variables and one dependent variable) were administered to the participants in their schools in three stages over a period of eight months (cf. *section 2.6.*). Such organisation of the study made it especially vulnerable to the process of **data attrition** i.e., losing full data for any subject who, for this or other reason missed school on the day of test administration. The actual number of sets of scores obtained for any statistical analysis depended on the configuration of the variables examined in a given analysis. For example, descriptive statistics for the TPJ test is available for 213 subjects, PLAT was factor analysed on 153 subjects but multiple regression analysis which involved all the five measures was carried out for 'only' 139 participants. Consequently, there may be differences in the magnitude of statistics reported depending on the variables involved. Because of this reason an exact number of subjects (N) for each analysis or statistical report is always given.

The present research relied on **natural groupings** of subjects into school-classes. The groups were natural in the sense that they existed prior to the commencement of research (Seliger and Shohamy 1990:136). No specific selection of subjects to groups was attempted. However, sometimes the groups (classes) used in the study reflected school-internal groupings based either on ability (grades from primary school) or school-specific specialisation (one school was a sport-oriented school). In the case of one school the grouping of learners available for sampling into two different classes was done according to the L2 learned – one group of learners (N=14) had L2 English while the other (N=13) learned L2 German. This, incidentally, led to problems with the construction of the L2 proficiency test (cf. *section 2.5.*). To control for the school-typical selection bias, care was taken to always include a representative group-profile in each school singled out for the study.

The process of selecting **schools** for the study followed from the statement of purpose as described in the *Introduction*. As one of the study's main aims was to investigate the validity of the concept of aptitude as measured by the two aptitude tests (TPJ and PLAT) for 13-year-olds in the context of formal school instruction the choice of secondary lower schools became kind of obvious. Seven schools of the *gimnazjum* type were picked out at random to form a representative sample of learners in the Poznań area. There were three schools from a large city of 650.000 inhabitants (Poznań); two from towns of 15-25.000 inhabitants each within the radius of 35-40 km from Poznań and two schools from a rural setting - *gimnazja gminne* (within the radius of 40-50 km. from Poznań). The proportion of girls and boys in the whole 227-learner sample was more or less equal with girls slightly outnumbering boys and was not significantly affected by data attrition, i.e., both girls and boys missed out on successive test sessions more or less evenly. *Table 2.1.* below presents the relevant figures.

School type	Total	Girls		Boys	
City	63	37	*59%*	26	*41%*
Town	58	34	*59%*	24	*41%*
Rural	106	55	*52%*	51	*48%*
	227	126	*55%*	101	*45%*

Table 2.1. Demographic statistics of the study.

All the pupils had been learning an L2 for at least four years (from Grade 3) as part of the national curriculum requirement before they came to a lower secondary school. Their contact with L2, as declared by learners in a questionnaire, prior to the research differed considerably and ranged from four to eight years. The L2 learned was mainly English but since approximately one forth of the pupils reported having it for one year only it can safely be assumed that prior to that moment their L2 must have been another L2, probably German, which they had been learning in a primary school. All in all, given that three hours of L2 a week from Grade 3 (1 hour = 45 min.) is the minimal curriculum requirement for modern foreign languages in primary schools the learners had on average 400 hours of L2 formal training. The differences in the amount of prior L2 schooling were observable between the groups (classes) and were not linked to a school type i.e., its being city, town or rural.

2.3. The instruments – intelligence measures

In order to obtain a measure of the learners' intelligence two tests were selected. Practical considerations of the tests' applicability for the available testing format were of paramount importance in choosing the tests. As the study's research design relied exclusively on group rather than individual test application Standard Raven Progressive Matrices Classical Form (SRPM) and Verbal Intelligence Test (TJL) by Jurkowski were found appropriate for data acquisition. The SRPM test version used was 1989 Polish standardisation edition by Jaworowska and Szustrowa (2000). The TJL test, standardised by the test's author, was used in its 1991 edition (Jurkowski 1997).

The two tests are firmly grounded in the mainstream cognitive research which goes back to Spearman's two-factor model (underlying one general factor: *g*–factor) of intelligence. Spearman's *g* – factor manifests itself through two abilities:

1. experience-free, eductive abilities – predominantly non-verbal in nature and characterised by order-imposing reasoning skills
2. experience-bound, reproductive abilities – mainly recalled verbal knowledge.

Cattell's fluid and crystallised intelligence (Cattell 1971) is yet another distinction which operates along a similar dimension.

In order to perform its diagnostic function of a reliable and valid intelligence measure the Raven's test, being a technique diagnosing experience-free cognitive abilities (eductive/fluid abilities), must be supplemented by a measure of reproductive/ crystallised abilities which are culturally embedded and thus knowledge-based (Hornowski 1970; Nęcka 2003). Jurkowski's TJL is a long-awaited Polish adaptation of *Mill Hill Vocabulary Scale* developed by Raven to complement his Progressive Matrices Test. Only when applied together will the two tools gauge subjects' intellectual capabilities as argued by Court and Raven (1995).

As already mentioned in *section 2.2.* the two intelligence measures are assumed to be reliable and valid instruments of the construct they purport to measure. Specifically, they meet formal requirements of what may be referred to as *'test's goodness criteria'.* According to Brzeziński, a test can be used in research provided it is:

1. objective i.e., yields results which are independent of place, time and the person interpreting them
2. standardised with respect to testing, scoring and interpreting procedures
3. a reliable measure of a postulated construct
4. a valid measure as applied to a particular domain
5. normalised i.e., its scores are expressed on a standardised scale thus providing meaningful information in the interpretation of the results
6. culturally adapted - this refers to tests whose country of origin is different from that of the population on which it will be used (1984:9, 2000:25).

For details concerning psychometric characteristics and interpretation of the results in the standardisation study an interested reader is referred to the Raven Progressive Matrices and TJL test manuals mentioned earlier in the section (Jaworowska and Szustrowa 2000; Jurkowski 1997).

2.3.1. Standard Raven Progressive Matrices: Classical Form.

It is a widely known test of non-verbal reasoning abilities. The test consists of five series (A, B, C, D and E) of 12 tasks each. Total score is 60. Each task presents a non-verbal reasoning problem and the subject's task is "to notice the order in which materials are arranged in the matrix, or the trends or systematic changes in stimulus characteristics that take place over the rows and columns of the matrix" (Carroll 1993:212). The subjects must indicate, on an answer sheet, which one from the six (Series A and B) or eight (Series C, D and E) elements given completes the matrix. The SRPM is a power test i.e., there is no time limit on the length of a testing session. Also the testees are encouraged in the instruction to adopt a '*wandering*' strategy i.e., moving forward, skipping the more difficult items and going back to them at a later time. As the test measures 'language-independent' abilities there is no rubric in the test and instruction is given orally. The test's response format is of a multiple choice type and an average time taken to complete it (in the age-range sampled i.e., 13,11- 14,5) was 37 minutes.

2.3.2. Jurkowski's TJL - (Verbal Intelligence Test)

TJL has all the characteristics of a standard test of verbal intelligence. It consists of nine vocabulary-based tasks of min. 5 to max. 10 items. Total raw score is 60 points – 30 for each part. The test, with each task on a separate page, comes in the form of a booklet and testees write their answers in the spaces provided. The test has two parts: Passive Vocabulary (PVoc) – tasks I to V; and Active Vocabulary (AVoc) – tasks VI to IX. This fact though is not conspicuous to a testee as there is no delineation between the two parts. The PVoc and AVoc parts differ from each other in the amount and nature of processing involved in doing them. The former is based on word recognition, as all the words needed to perform the tasks are given, whereas the latter puts more productive demands on the subjects who have to recall the items needed for the tasks. In general terms, the nature of the response format of the PVoc tasks is of a 'closed' type whereas the AVoc tasks are of an "open-ended" type. The style, register and explicitness of the instructions to each task are generally appropriate but seem to be aimed more at a lower age-group and may sound too explicitly patronising and drawn-out to older learners. The tasks include the following aspects of lexical skills:

1. giving/finding synonyms
2. giving/finding words to definitions and definitions to words given
3. adjective – noun collocation
4. completing a sentence with an appropriate word.

TJL is a power test and in a 13,11 – 14,5 age-group took on average 50 minutes to complete. As the test is a relatively new measure it is included in *Appendix 2*.

2.4. The instruments - FL aptitude measures

As said before, of the two tests used in the study with the aim of obtaining measures of the learners' FL aptitude only Kuliniak's FL aptitude measure (TPJ) had been piloted prior to the commencement of the research. The other measure used (PLAT) was written by the present author and was piloted on a modest sample of 109 learners. What follows is a description of the constructs that TPJ and PLAT set out to measure. This is done through the examination of the content of the tests' tasks.

2.4.1. Test Predyspozycji Językowych – TPJ (Language Aptitude Test)

To date it is the only widely available test in Poland that purportedly measures FL aptitude. It was written by Kuliniak and was published in a booklet form with eight other similar tests. On several occasions the test was also reprinted in an educational supplement in a nation-wide daily newspaper (cf. *Appendix 3* for the full test). TPJ contains 40 test items grouped into 8 tasks (*Ćwiczenia*). The author of the test identifies five groups of abilities that he wants his tasks to capture:

Group 1. (Task 1) - sensitivity to register and to phrase (sentence) grammaticality

Group 2. (Tasks 2, 3, 4) - richness of lexis and of collocations

Group 3. (Tasks 5 and 6) - recognition of grammatical form in L1– analogy task

Group 4. (Task 7) - recognition of grammatical form in L3 – Latin

Group 5. (Task 8) - inferring rules from fictitious language material

(Kuliniak 2002:64, own translation)

At closer scrutiny, though, the test reveals some problems both with the author's delineation of abilities measured and terminology used. Even a cursory look at the test's content will reveal that:

- **Task 1** exclusively measures recognition of phrase and/or sentence grammaticality to the exclusion of register sensitivity (except for, marginally relevant, item 4A; cf. *Appendix 3*).
- The scope of collocations is not covered at all by **Tasks 2, 3 and 4**, instead the tasks focus on logico-semantic relations of synonymy and antonymy between lexical items, first in the context of a sentence (Task 2) and then in isolation (Tasks 3, 4).
- Similarly, the third group of abilities labelled *'recognition of grammatical form in L1'* does not characterise abilities that **Task 5** calls for and is only vaguely valid for **Task 6**. The former task may be misleading in that the testees may interpret the elimination principle along semantic, phonemic or graphemic lines rather than formal (grammatical) as posited by the author. It is to be observed that there is no principled reason why D should not be eliminated in 22 – all the remaining words contain the letter <o> in word-final position; or, alternatively, are two-syllable words. The problem of unexpected (and often surprising but logical) categorisations is a perennial problem of 'the-odd-one-out' tasks as noticed by Alderson et al. (1995) and Hoffmann (1962).
- The author is also suggesting that **Task** 7 encompasses the same ability as Task 6 whereas in fact, it relies on the same inductive reasoning abilities attributed to **Task 8** i.e., "identifying patterns of correspondence and relationships involving either meaning or syntactic form" or

both (Skehan 1989: 27). Incidentally, Task 8 is the only one in Kuliniak's battery whose 'declared' ability domain agrees with what it says the task measures.

Apparently, the problems with the identification and definition of postulated ability domains do not influence the subjects' performance on the test but could have a confounding effect on the analysis of the test's construct and the evaluation of its predictive strength in the subsequent statistical multivariate analysis (cf. the discussion of this issue in *sections 3.6.2.* and *4.2.1.*)

After the re-interpretation of the posited ability domains measured by TPJ in the spirit of Carroll and Sapon's classical view of language aptitude as embodied in MLAT (Caroll and Sapon 1959, Carroll 1979, Carroll 1993) and discussed in more detail in *Chapter 1*, the TPJ's five tasks could be understood as vaguely tapping the following two abilities (out of four) postulated by Carroll:

1. 'grammatical sensitivity' (Tasks: 1, 5 and 6)
2. 'inductive language learning ability' (Tasks: 7 and 8).

Alternatively, following Skehan's suggestion (Skehan 1986b, 1989, 1998) 'grammatical sensitivity' factor and 'inductive language learning' factor are collapsed into one 'linguistic ability' factor. As Skehan notes: "this melding is partly justified because the evidence requiring distinct factors here is not extensive" (1989:201). As will be seen in *sections 2.7.* and *3.6.2.*, where TPJ's validation procedure using factor analysis is described, only Tasks 7 and 8 consistently pattern together invariably loading on one underlying factor when TPJ' construct validity is verified. Prior to this discussion, it is assumed that Tasks 7 and 8, representing some sort of language analytic ability, proved to be the only ones from TPJ worth including in the present aptitude study.

The remaining three tasks (2, 3 and 4) which evidently measure L1 lexical competence are invariably used in most standard tests of verbal intelligence (cf. TJL) and in the form presented in TPJ are rarely included in FL aptitude measures. It follows then, that TPJ is partly a test of FL aptitude (although limited in scope) and partly a measure of crystallised intelligence (verbal skills). Consequently, the test suffers from the lack of clear definition of construct it purports to measure i.e., has weak construct validity. Furthermore, given the test's formal bias and its reliance on prior experience one might expect that the test's results would not necessarily show high correlations with L2 proficiency criterion measures. Indeed, one study in which mid-year school grades in Polish, English and GPA (academic subjects only) served as criterion measures reports higher correlation coefficients (**r**) for GPA ($r = .56$) than for Polish ($r =.47$) and English ($r = .49$) grades (Rysiewicz 2003). Although, all the correlations are statistically significant ($p = .01$), they are not high given the small sample, N=35. Consequently, the test's predictive validity of L2 learning success is less than moderate.

In the light of the above mentioned problems with adequate hypothesis formulation preceding the construction of items included in TPJ, it was considered necessary to examine the test's structure with the hope of determining a clearer internal factors lying at its bottom, rather than rely on doubtful and confounding construct description provided by the test's author. To this end exploratory factor analysis was carried out on a multiple correlation matrix obtained from the TPJ's eight tasks. On the basis of this analysis a decision was made as to which factors thus extracted were to be included in further analysis of the variables used in the study. The results of construct validation of TPJ by means of factor analysis are presented in *section 3.6.2.*

2.4.2. Progressive Language Aptitude Test – PLAT

PLAT (cf. *Appendix 4* for the full test) is an FL aptitude tool created for the purpose of the present study in view of the limited scope of abilities tapped by TPJ. As these were found to be only marginally on target (cf. the discussion above) the author was faced with a need to construct a test which would be a more reliable and valid measure of L2 learning potential. In attempting the task the present author tried to draw more on available theoretical conceptions of FL learning abilities rather than rely on his intuition and/or experience however large it might be. In constructing the test an effort was made to model the test after already existing and widely known FL aptitude tests as well as to adapt some of the ideas form the CANAL-F theory (Grigorenko et al. 2000). Specifically, it was considered appropriate to base the test's tasks on the premise that they should:

1. cover a **wider scope** of abilities demonstrated to be contributory to L2 acquisition (Skehan 1989, 1992, 1998; Carroll 1993)
2. reflect current **theoretical** views on the nature of L2 learning and/or acquisition (Skehan 1989, 1991; Ehrman and Oxford 1995; Ehrman 1996; Grigorenko and Sternberg 1998; Grigorenko et al. 2000)
3. try to replicate an incremental nature of L2 learning process i.e., **naturalistic** (Grigorenko et al. 2000)
4. be more **process** rather than **produc**t-oriented by allowing for a learning skill at the time of test i.e., **dynamic** rather than **static** (Grigorenko and Sternberg 1998; Grigorenko et al. 2000)
5. provide a **componential view** of an L2 learner abilities i.e., allowing for instructional decisions appropriate for a learner specific profile (Ellis 1994:498; Skehan 1989; Grigorenko et al. 2000).

Accordingly, PLAT included tests of various facets of memory (working and long-term, immediate and delayed recall, visual and auditory memory, accidental and intentional memory, in context and decontextualised) implicated in either visual and/or auditory comprehension or reproduction tasks. The tests included in the first four rows in the table below are broadly reminiscent of the memory-based abilities thought crucial to FL aptitude as originally formulated by Carroll (1962, 1981, 1993) and modified later by Skehan (1986b, 1989, 1998). Namely, it is believed, following Shehan (1981) and research on meaning-mediated learning which gives priority to context-embedded message extracting abilities, that text comprehension constitutes an important framework for the memory construct of FL aptitude. Following *Table 2.2.* below the discussion of PLAT's tests with reference to points 1 and 2 above is pursued in some detail. This, in turn, is followed by a brief characterisation of PLAT with reference to points 3, 4 and 5.

Task No*	***Task***	***Cognitive operations involved***	***Mode***	***Task/s description***	***Response format***
6	*INPUT*	unstructured/uncued reading	Visual	**Reading comprehension**	
7	**ofiarą**	immediate recall		memory for text (content,	MC
10	**handlarz**	delayed recall		visual-spatial organisation); con-	T/F
13	**układ**	delayed recall:accidental encoding + selective comparison		textual	Matching
	INPUT	unstructured/uncued listening	Aural	**Listening comprehension**	
3	**przed**	immediate recall		memory for text (content); con-	MC
5	**podczas**	delayed recall		textual	T/F
8	*INPUT*		Visual	**Paired-associate learning**	
9	**lubić**	immediate recall+ accident. encod.		decontextualised memory for	SR
11	**pamiętać**	delayed recall		new words	Matching
2	**dyktando**	immediate recall/reproduction + strategy differentiation	Aural	**Dictation** auditory memory; contextual (structured,cohesive text)	Write
14	*INPUT*	accidental encoding	Visual	**Symbol-meaning link**	
16	**wingdings**	delayed recall + selective compar. and selective transfer	Kinesth	kinaesthetic learning; decontextu-alised	Writing Matching
		inferred meaning from cues at:	Visual	**Lexical production (in context)**	
12	**pojezierze**	– word, phrase, sentence level		grammatical context	C-test
15	**sprawcy**	– word level		lexical context	Gap-fill
			Visual	**Lexical recognition (no context)**	
1	**zwolennik**	selective encod. (FI/D) + compar.		in excess information	Circling
4	**dbrdszn**	sound-symbol link		in impoverished context	SR

**Numbers follow the ordering of sub-tests in PLAT – see Appendix 4*

Table 2.2. Summary of PLAT's structure of tasks and abilities tested.

I. Tasks 7, 10 and 13 (listening comprehension) and 3 and 5 (reading comprehension) rely on the learner's postulated ability *to recall written and/or aural material after initial exposure to text* when no particular emphasis in the learning leads the learner to believe whether retention is going to be important for further task completion (Skehan 1989). The subjects' memory for text's content was tested: twice for aural, and three times for visual mode of presentation using, first, immediate and then delayed recall.

II. Tasks 9 and 11 include *a test of paired-associate learning* which represents one aspect of memory couched in associationist psychological theory dominant in the days when Carroll and Sapon wrote their MLAT (1959). The test was adapted from Meara et al. (2001) LAT-B. The subjects' task was to memorise a list of 10 words from a fictitious language and their Polish 'equivalents'. There was a 'cued' learning step included before the learners' memory was finally tested with a 'matching' task.

III. Task 2 is a dictation task. The dictation task is a test of subjects' ability to remember sentences presented aurally and hold them in their working memory (articulatory loop) for the duration of the task. The dictation consists of six sentences forming one cohesive, self-contained text. The length and syntactic complexity of the sentences varies progressively as the dictation proceeds.

IV. In tasks *14*, 16 (row 5 in *Table 2.2.*) an attempt was made to contrive a test that would take care of the dual structure of language (*'double articulation'*). The ability that the tasks hopes to address is *the capacity to segment incoming language* in its written form. This ability is assumed to function on the 'grapheme-to-meaning' plane in basically the same way as Carroll's 'phonemic coding' does on the 'sound-to-meaning' plane. Research on language learning difficulties by Ioup (1989), Sparks and Ganchow (1991a), Ganchow et al. (1998) and Sparks (2001) suggest an essential similarity between the incapacity to attend to 'letter-mediated' correspondences in coding/decoding language units (dyslexics) and the inability to impose structure on incoming auditory input so that it can be absorbed and retained for later recall, use or restructuring. Test 4 in the seventh row (*dbrdszn*) also addresses the 'sound-to-meaning' dimension at the same time functioning as a test of native vocabulary.

V. Recognising the central role of *the ability to cope with novelty and ambiguity* in FL acquisition (cf. the discussion of CANAL-F in *Chapter 1*) tests 12 and 15 were included in PLAT. The tests are 'impure' versions of the original test formats of C-test (Raatz and Klein-Braley 1981) and cloze procedure (Taylor 1953). What the two formats have in common is the integrative nature of the processing operations required. For the restoration of meaning and/or text cohesion, those processes depend on the amount of context available, parsing abilities and also, memory. Needless to say, there are other factors affecting the tasks' disambiguation which bring about variability in performance (cf. Bachman, 1990 and Brown, 1985). The sort of criticism levelled against the use of C-test as an integrative test of L2 competence, on the grounds that a test using messages with reduced redundancy does not necessarily test FL competence (Carroll 1987), seems to lend support to the choice of the format in testing the L1 abilities mentioned above. Given the analytical nature of the Polish language C-test seems to be very well suited as a measure of Polish subjects' capacity to operate on the macro-level of a phrase and a sentence. Despite the uncertainty as to the precise nature and interplay of the processes involved in both tests used for testing L2 competence they seem to be adequate formats for the kind of ability testing envisaged in PLAT. In short, the tests' theoretical rationale can be characterised as the ability to predict from context (immediate or distant) with reduced redundancy.

VI. Tests 1 and 4 were meant to measure the learners' abilities to process phonological-orthographic code of the native language through native vocabulary. It was hypothesised that the tasks would call for an ability to analyse the input into units of the code: visual and aural (letters and sounds) and test each hypothesis (formed 'word') for correctness against a semantic interpretation. Both tasks involve varying amounts of redundancy: *zwolennik* contains excessive information which acts as 'background noise', whereas *dbrdszn*, adapted from Carroll and Sapon's MLAT (Part III, Spelling Clues) operates on impoverished context. The tests measure both native vocabulary and phonemic coding ability (Carroll and Sapon 1959).

While the construct of *dbrdszn* can be safely assumed to be fairly well established, the construct posited for Test 1 (*zwolennik*) needs some explanation. Test 1 can be thought of as a verbal test of cognitive restructuring: an ability to disembed or restructure visual stimuli (Chappelle and Green 1992). According to Wittkin and Goodenough (1981), reliance on internal/external referents, cognitive restructuring skills and others is what makes up the construct of field-independence/dependence (FI/D). Typically, it is the embedded figures test (EFT) – a non verbal test and thus a test of fluid ability – that is used in L2 research as a test of FI/D. Following Chappelle and Roberts' (1986) claim that FI/D should be considered a component of FL aptitude and Chappelle and Green's (1992) suggestion "to extend research concerning the role of fluid ability beyond the specific restructuring ability as measured by EFT" (ibid. 77) it was considered appropriate to try to 'embed' verbal material into the FI/D guise.

Unlike traditional tests of FL aptitude such as Carroll and Sapon's MLAT, Petersen and Al-Haik's (1976) DLAB or Pimsleur's LAB the current test tries to adapt (in some of the tasks) the ideas suggested by Grigorenko et al. (2000) and Meara et al. (2001). (cf. *section 1.3.3.2.* for more detailed discussion). Accordingly, PLAT is structured as a test resembling **natural learning situation**. Namely, there is an input stage (IS) followed by a learning (rehearsal) stage (LS) which is then followed by a testing stage (TS). And so, in Task 6 the subjects are asked to read a text (cf. *Appendix 4*) on whose chosen content they are immediately tested in Task 7. This is then followed by Task 8 in which they are subjected to another learning situation (*paired-associate memory*) which gets in the way of the text-content memory task. In Task 10 the subjects are again tested on the text from Task 6 (more general questions; TRUE/FALSE response format). Finally, after further two tasks the subjects' spatial memory of the semantic aspect of the text is tested. Although this pattern is not as rigidly adhered to as is the case in the CANAL-F Test (Grigorenko et al. 2000) the sequence IS-LS-TS is applied to quite a variety of input which calls for changing involvement of different cognitive operations. PLAT's naturalness also manifests itself in the unpredictability of test-induced learning occasions which imitate real-world adaptive demands to cope with novel situations. Because of its 'progressive' design PLAT claims to be more than just a test of a series of ability constructs through which the outcomes of the process of knowledge acquisition are tested. Rather, it is the very processes of knowledge acquisition at the time of the test that is tested via the **dynamic paradigm** (Grigorenko and Sternberg 1998; Grigorenko et al. 2000). Consequently, PLAT, as partly modelled after CANAL-F Test, hopes to measure the potential/ability to learn rather than acquired knowledge.

2.5. The instruments – L2 proficiency test (criterion)

L2 Proficiency Test was modelled after UCLES Key English Test (KET) both in test/task format and scope. The KET is described in an UCLES test brochure as measuring candidates' skills to communicate in basic daily life situations. KET is an internationally recognised English language qualification providing assessment at ALTE/Cambridge Level 1 based on the Council of Europe's Waystage specifications. To achieve this level ca. 200 hours of English instruction is necessary.

One major difficulty that was encountered in constructing the test was the difference in L2 competence levels between the schools. Very early in the test construction process it became obvious that what looked like an appropriate L2 competence test for School A group of learners was either too difficult or too easy for School B learners. While total homogeneity of L2 proficiency levels in the sample had never been envisaged the differences seemed to be more the results of school-type and teacher-factors rather than cognitive factors. That is, the learners from rural schools had generally lower level of L2 than those from the city or town schools. Thus, it was decided that the test should make allowances for the apparent differences between schools. Accordingly, three versions of the test were constructed by the present author with the help and advice from the schools' L2 teachers. The fourth version of the test for L2 German group of 14 learners was prepared by the group's German teacher. The draft version of the test was given to two independent judges (L2 German teachers teaching the same level and age-group) and their remarks were used in preparing the test's final version. The German version was identical in form and length to L2 English Proficiency Test version. The versions differed in the number of test items included (51, 58 and 68 items respectively) and thus, indirectly, in length (cf. *Appendix 5* for four versions of the test).

The fact that it was not feasible to administer one, uniform test of L2 proficiency to all the subjects and that different versions of the test had to be used meant in practice that each version was, in fact, a different test (cf. *section 0.3.*). This fact might be a serious drawback for the validity and reliability of the results. However, since the aim of the study was to investigate variability within the same participant across the five measures used and not the kind of norm-related variability which obtains from the comparison of subjects' performance, the problem of scale comparison of the different test's versions was dealt with by using percentage values rather than raw scores. The use of percentages meant that both 'intra-' and 'inter- subject' dimensions of the tests were available with no loss of information about the differentiation of results on the scale as would have been the case had standardised scores (*z* scores) been calculated. Task format, skills and competencies tested in the four versions of the L2 proficiency measure did not differ. It is assumed that the three versions are parallel and equivalent to each other and that the results are comparable.

The 'original' UCLES KET consists of three Papers: Paper 1: Reading and Writing (70 min., 8 parts); Paper 2: Listening (30 min., 5 parts); Paper 3: Speaking (10 min., 2 parts). Considerations of practicality made the inclusion of the learners' speaking skills in the test impossible. Continuous assessment grade of this component of L2 proficiency was obtained, instead. Similar considerations led to the decision to shorten the test to 60-70 minutes. Depending on the school, seven or eight tasks were included in the final versions of the L2 proficiency measure which was divided into five subparts:

1. listening – 1 to 4 min long; TRUE/FALSE or short response format
2. reading – 2 texts; multiple-choice format; the lower groups had one reading text only (a dialog)
3. grammar in context – gap-fill with structural items; transformations
4. vocabulary – matching, elimination, word formation
5. writing – a postcard: max. 25 words; a letter: 100 words

All the tasks, with the exception of the listening part were either taken verbatim or adapted from *Cambridge Key English Test 2. Past Papers* (2001). It is assumed that despite 'heavy mutilation' the truncated KET still had both content and construct validity. Its appropriateness (face validity) for respective group of learners was judged by their English teachers who were asked to analyse the test's difficulty and 'doability' in a test evaluation questionnaire. Doubtful, difficult or unclear items were then deleted or re-written.

The table below contains a summary information on all the measures discussed in *sections 2.3. – 2.5.*

	Posited Construct	Variable's Function	Test type	Score in Analysis	Total Score
SRPM	intelligence (fluid)	independent (predictor)	power	raw score	60
TJL	intelligence (verbal)	independent (predictor)	power	raw score	60
TPJ	FL aptitude	independent (predictor)	power	raw score	40
PLAT	FL aptitude	independent (predictor)	power/ speed	raw score	138
KET – A	L2 proficiency	dependent (criterion)	speed	raw score (percentage)	51/58/68

Table 2.3. Summary of the measures used in the study

2.6. Data collection and tests administration procedure

2.6.1. Stages of data collection

The data for the present study were collected in 4 test-stages and one test-free stage:

Stage 1: late October 2002: TPJ – the first FL aptitude tool

Stage 2: late November 2002: PLAT – the second FL aptitude measure

Stage 3: January 2003: SRPM and TPJ – two intelligence measures

Stage 4: June 2003: KET-A – L2 Proficiency Test (after ca. 100 hrs of ESL)

Stage 5: learners' mid-year and final school grades were obtained

The results from the first three stages formed the independent variables of the study whereas the data obtained in stages 4 and 5 formed the dependant variables. There were a few practical considerations for spreading the test-stages over three visits to the schools rather than administering them all on one visit. One such consideration was learners' fatigue with an unknown and relatively demanding test formats which undoubtedly would have affected their performance. Another, had to do with minimising a disruptive effect of the external interference with the schools' activities.

As already mentioned in *section 2.3.* TPJ, SRPM and TJL were principally power tests i.e., the subjects were given ample time and they could go back to more difficult items they had left out. The criterion test (KET-A), understandably, was timed (but not speeded) and subjects could go back to items they had omitted (if time and the task's nature allowed). In this respect PLAT was a mixed type with, on the one hand, a fixed sequence of tasks (at many of which learners were not given a second go) and, on the other hand, a few sub-tests were rigorously timed (speeded tests) allowing for no back-up.

The tests in stages 1, 2 and 3 were administered and supervised by the present author. When there was more than one group to supervise at any given time the help of the teachers in doing test proctoring was sought. The teachers' role was limited only to maintaining order in class, no part of tests' administration procedure was relegated to them. In the case of 15 pupils the scores obtained by them on the test measures were considered unreliable (cf. *section 2.2.*) and were removed from further considerations. This had to be done if the scores were to meet the assumption of scores independence (cf. *section 3.2.2.*). Grading and scoring of the tests in stages 1 to 3 was done by the present author. The procedure was different for Stage 4 of L2 German test. It was administered and supervised by a class EFL teacher, who also corrected and scored the tests with the key and instructions provided to her. The test's total score with component results were then reported to the author on a Test Report Sheet.

2.6.2. Scoring the tests

The scoring procedure of KET-A was relatively error-proof as the majority of the test items used objective response formats (multiple choice, limited selection, elimination, matching). In scoring the written component care was taken to provide as explicit a mark scheme as possible, which was not difficult given the elementary level of the writing task. In the case of 'fill-the gap' task due to heavily reduced structural ambiguity the task became virtually objective as there was only one possible word to go into each gap.

The scoring of the SRPM was straightforward and did not leave any margin for error. As to TJL its Passive Vocabulary part was unproblematic as all the tasks were objectively scored. The Active Vocabulary part was, quite understandably, 'open-ended' but with a little cueing, though. It was scored against the key however, acceptability criterion was also used at times. This was considered necessary because the key did not make allowance for alternative answers. For example, in task VI item 9 the key only accepts *'nawiązywać'*, whereas *'sprawdzać'* is also possible. There are a few examples of the sort in the test.

With reference to PLAT's scoring procedure four tasks required separate treatment from the remaining nine which were all objectively scored. The four tasks were: cloze (*sprawcy*), C-test (*pojezierze*), dictation and 'page layout memory'. The cloze task was scored using the acceptable-word method, however the number of possible options was not large. The exact-word method was used in scoring the C-test and this seemed the only viable method given the theoretical rationale behind the C-test measured construct i.e., "the ability to predict from context with reduced redundancy" (Jafarpur 1995). The dictation task was scored using the procedure adopted by Irvine et al. (1974), whereby the number of undistorted words appearing in the original sequence was counted. Misspelled words were accepted as long as they made sense within the sentence. One or two departures from the original sequence per sentence still won a point. A combined total score for six sentences in the dictation was awarded. In the 'page-layout memory' task (cf. *section 2.4.2.* for the task's description) a full point was awarded for placing the sentences in an appropriate section of the blanked text. No exact positioning of the sentence was required.

Two last issues still need mentioning here. One concerns the dictation task and more precisely the manner of its delivery to learners on each of the 9 testing occasions (in two schools two groups of learners participated in the study). Alderson et al. (1995) remark that to maintain the same standard of presentation dictation should be pre-recorded and played-back to testees, or if the use of a tape recording is impossible the person doing the dictation must be aware of how to control speed of delivery, positioning of pauses, voice emission and the like. In the present study it was the author who tried to live up to these standards.

The other issue touches upon the delicate nature of mental ability testing in general and tests of intelligence (SRPM and TJL) in particular. In psychology this problem is known as 'test's invasiveness' i.e., the potential of a test to harm a subject's emotional and/or mental well-being if used by a non-expert or in a way showing ignorance of test-specific procedures. To guard against this danger the author co-operated closely with a trained psychologist who advised him on how to run the test sessions using the two measures mentioned. By the same token, the subjects' right to be informed about their test results was respected. After each test was scored and corrected the learners received reports with their raw score, percentage score and class average. These reports were delivered to learners personally to secure minimum privacy and to avoid peer ridicule and derision in case the scores were very low. In the rare cases of exceptionally bad performance (5% out of total 1044 tests scored) i.e., when the learners low scores against class average might have affected their self-esteem, a 'soothing' explanation of the result accompanied the figures. Also the learners' parents' were asked written permission to agree to their children taking the tests.

2.7. Statistical Analyses

This section will present the statistical instruments used in the analysis of the data obtained in the study. First, item analysis of PLAT will be briefly discussed (*section 2.7.1.*). Then the procedure of the tests' reliability estimation will be presented (*section 2.7.2.*). This will be followed by the description of how PLAT's low reliability sub-tests were handled (*section 2.7.3.*). Another statistical procedure employed in this section is factor analysis which is used as an analytical tool in construct validation of FL aptitude tests (TPJ and PLAT) and the proficiency measure (KET-A) and then in structure exploration of the nature of the abilities measured by the test variables used in the study (*section 2.7.4.*). Finally, the use of multivariate analytical tool of multiple regression will be discussed in the prediction of L2 proficiency from a set of independent measures (*section 2.7.5.*). No results of the statistical analyses will be given in this section. They, together with an ensuing discussion, will be a subject matter of *Chapters 3* and *4*.

All statistical analyses carried out in the present research were performed using a data analysis software system STATISTICA of StatSoft, Inc. (2003). version 6. licensed to School of English, Adam Mickiewicz University, Poznań, Poland.

2.7.1. Item Analysis

Before PLAT could be used in the research part of the study as a test of postulated L2 capacities, it had to be piloted on a smaller sample of the target population of 13 year-olds as described in *section 2.2.* Because of the comfortable piloting situation (ample time, motivated learners, strict supervision) the pilot version of the test covered a wider spectrum of abilities purportedly pertinent to FL acquisition than the research version. And so, the piloted version additionally included six tasks (cf. *Appendix 4b*): three tests of language analysis, two tests of contextual inference (one involved reading a fairy tale in Czech and answering comprehension questions), and an additional learning stage task for *Akareke* paired-associate memory (the other half of items than in 9). As it will be recalled from *section 2.2.* the results were obtained for two groups of candidates: the younger and the older, respectively 13 and 15 years of age.

Before PLAT's reliability was estimated, item analysis was performed on the test's results. Facility and discrimination indices for each of the 211 test's items were obtained. Facility value (FV) is a measure of the level of difficulty/facility of a test item and is simply expressed as a ratio of learners who got the item right to the total number of learners taking the test. FV index assumes values from 0 (impossibly difficult) to 1 (trivially easy). Discrimination index (DI), in turn, is a measure of "the extent to which the results of an individual item correlate with the results from the whole test" (Alderson et al. 1995:80). In other words, an item has a good discriminatory power if more of the top-scoring learners on the whole test also answer the item correctly. DI assumes values from 0 (no discrimination) to 1 (full discrimination). As all the items in PLAT, TPJ and KET-A were scored using a dichotomous scale i.e., the item was either answered correctly (1 point awarded) or not (0 points awarded), the point biserial correlations were calculated to yield DIs. In addition to the correlational type of the DI, an index based on the difference between the extreme groups of learners (D_{30}) was also calculated and the two indices compared to achieve maximum accuracy. No major differences were found between them.

In the subsequent procedure of item elimination the test was screened for 'suspicious' FV and DI indices. Whenever those were found below generally acceptable levels for a particular item (Niemierko 1999:154,158), the item was considered for deletion or modification. However, the decision to delete an item was not a mechanical one based only on the two statistical parameters. Items considered important for content validity of the test but with very low or low DIs were still retained. Also the fact that due to the specificity of the piloting scheme two different age groups were available helped in the decision. The older group acted as a kind of 'control group' in the examination of the test's facility value. Items which had low facility value (FV) for the younger group but were more difficult for the older group became likely candidates for deletion from the test.

The same procedure was repeated for TPJ. Its aim however was to 'confirm' the test's parameters rather than to 'explore' them. As already mentioned in *section 2.4.*, TPJ had undergone large scale piloting and was available as a ready-to-use test with the parameters known at the time the present research started. Consequently, indices of item facility and discrimination obtained in the research (N=227) were compared to those from the test's pilot (Kuliniak 2002:64). They were found to be highly comparable [1].

2.7.2. Reliability and validity of tests

To put it at its crudest, a test's reliability is its "overall capacity to define levels of knowledge or ability among candidates consistently" (McNamara 2000:61). In other words, before any generalisations about a cognitive or affective domain can be made a researcher must be sure that the performance on a test (observed in the test's scores) is due to an unobserved dimension measured rather than to the imperfection of the tool used i.e., measurement error (Bachman 1997). Test reliability is expressed as a reliability coefficient (index) which can assume values from 0 (no reliability) to 1 (perfect reliability). Test validity, on the other hand, can be described at its simplest as "the degree to which a test measures what it claims to be measuring" (Brown 1988:101). There are various types of validity and ways in which it can be demonstrated for a test[2]. Two types of validity, content validity and construct validity, have already been briefly touched on earlier in this chapter (cf. *section 2.4.*). Below, the validation of TPJ and PLAT using the technique of factor analysis will be discussed.

Depending on the number of measurements available to a researcher, the range of abilities tested and the nature of the scoring procedure i.e., whether a test is objective (raters cannot vary in the scores they give, for example multiple-choice) or subjective (raters will display variability of scores, for example written compositions) different methods can be used to estimate reliability of a test. Accordingly, methods relying on either two applications of a test ('test–re-test') or on alternate forms of a test ('parallel tests') could not be used in the present study as only one application of each test was available for measurement. In other words, it was not possible to retest individuals after some time or to administer to them an identical version of the target test to see whether the two sets of results give acceptable correlational values. This is why internal consistency estimates which rely on "sources of error from within the test and scoring procedures" (Bachman 1997:172) had to be used. There is one problem though. The internal consistency method works well on the assumption that items included in the test are homogenous i.e., all measure more or less the same ability/skill/capacity. Apparently, this condition cannot be met in the case of cognitive ability testing where the heterogeneity of the measured universe that a researcher wants to describe (here: individual differences in language learning abilities) presupposes heterogeneity in the tools meant to explore it. Both PLAT and TPJ are examples of multidimensional tests.

Reliability of a multidimensional test must be estimated for each dimension in the test which is then treated as a separate test (a sub-test). Internal consistency reliability estimates are independently calculated and reported for each sub-test. This is a viable solution provided the test is a valid measure of a posited construct. In other words the test must be shown to represent constructs it purports to measure (Bachman 1997; Hornowska 2003). Consequently, structure confirming validation of a test must be carried out first. The analytical technique of FA used as an exploratory tool will provide a researcher with a firmer basis to identify the ability-structure of his test battery than otherwise afforded by his intuition or hunches. If the factors that have been isolated through FA lend themselves to an interpretation as representing ability dimensions of the test, internal consistency reliability indices can be estimated for each dimension (factor) as an independent sub-scale. (cf. *Chapter 3* for the discussion of the interpretation of factors extracted from TPJ and PLAT). The importance of the approach: *validity first – reliability estimates second*, is obvious. Without a clear definition and/or isolation of the construct the test targets, there is little comfort in the knowledge that the test reliably measures a dimension of which researchers are only mildly aware.

At this point it is necessary to justify the method of validation used for TPJ and PLAT. The type of construct validity used and discussed above looks for confirmation of the test's construct inside the test. Another source from which the confirmation of the test's validity can come lies outside the test i.e., is external to its structure. Two types of external construct validity can be envisaged. One seeks justification of the test's construct in its significant and high correlations with other independently existing tests of a similar construct (convergent validity), the other hopes to demonstrate that because the test in question displays significant and low correlations with related but dissimilar tests, its construct is different from those compared to (divergent validity). In short, to claim construct validity for a test, both convergent as well as divergent validity have to be demonstrated against some external measure. As no such measures are known to have been either adapted to the Polish context or written for the purpose of research there was no externally available measure for this type of construct validation and, instead, the method of internal validation had to be used.

While reliability and validity indices of the FL aptitude measures are discussed in *Chapter 3* one type of validity, content validity, needs to be briefly mentioned with relation to TPJ. The key terms here are: representativeness and content domain (subject matter content). Hatch and Farhady (1982:251) define content validity as "the extent to which a test measures a representative sample of the subject matter content". And they go on to remark: "The focus of content validity is on the adequacy of the sample and not on the appearance of the test (face validity)" (ibid.). This invariably involves making decisions about what skills and/or components make up a particular domain one wants the test to quantify. In other words, test items must reflect in 'miniature' the universe of the sampled content domain (Hornowska 2003:87). According to the same author (ibid.), a test has content validity if it meets the following two criteria:

1. all the test's items belong to the defined content domain
2. the test as a whole is proportionally representative of the content domain

Content validation then, typically involves the judgements of experts in the field with respect to the above mentioned criteria. Such judgements are relatively easy and unproblematic to make even by non-experts when the content to be sampled is well established and/or defined, which is the case with most achievement and proficiency tests. More expert judgements and caution must be applied when less well defined and more ambiguous content domains are sampled for the purpose of a content valid test, which is the case, for example, in all types of ability testing (Hornowska 2003).

TPJ's content validation involved five experts (all applied linguists with many years' experience) who were asked to express their judgements on the content validity of the test's items. The values of CVR (content validity ratio) for eight test's tasks ranged from – 0.60 to +0.30 for tasks 2, 4 and 5 (the majority of judgements were negative); and from +0.40 for tasks 1, 5, 6 to +0.80 for tasks 7 and 8. Thus tasks 7 and 8 were the only ones that the experts found relevant for the postulated sample of the content (i.e., they were close to being statistically significant at α = 0,05).

In the case of PLAT, where FA yielded an interpretable structure, the values of Cronbach's α (Cronbach's coefficient 'alpha') general formula for estimating internal consistency are reported separately for each postulated sub-scale of the test whose items do not violate the independence assumption of classical reliability estimates. The items in three PLAT's sub-tests: Dictation, C-test *Pojezierze...* and a gap-fill *Sprawcy...* could not be assumed to be independent of each other as they are dependent on the context beyond the clause and cannot be said to 'stand alone' (Bachman 1997). In the remaining sub-tests it is assumed that the items meet the condition of item independence despite the fact that in a few tasks (reading and listening-based sub-tests) some dependence of items on the same passage is apparent. This, however is not a major problem since the nature of the abilities tested minimises this dependence (the tasks concerned are more memory-based than reading comprehension-oriented). The Guttman split-half reliability index was calculated for the three independent sub-tests.

In the case of TPJ the results of FA did not yield factors which could be interpreted in terms of posited abilities. The only sub-tests which displayed a conspicuous patterning were *TPJ 7* and *TPJ 8*. It was these components of TPJ which were used in further analysis together with PLAT.

The reliability estimates for the L2 proficiency measure, KET-A were made for each sub-test i.e., listening comprehension, reading comprehension, grammar and vocabulary. Because the first two sub-tests violated the assumption of independence of items, Guttman split half estimate was calculated for them. For the remaining two scales (grammar and vocabulary) Cronbach's α are reported.

The Speraman-Brown split-half estimates for SRPM are cited after Jaworowska and Szustrowa (2000) and for TJL after Jurkowski (1997). The tests' reliability is assumed and no separate reliability estimates were made from the data obtained during the present study. The reliability coefficients are high and generally stable throughout the sub-groups of individuals.

2.7.3. Item trimming to enhance PLAT's reliability

As noted above, 211 test items making up a preliminary version of the test battery were screened for low facility and discrimination indices. Apart from item facility and discriminatory potential, considerations of test practicality were of importance in deciding on the final version of PLAT. As it will be recalled from the introductory chapter, one of the aims of the study was to construct a test of FL aptitude which apart from being a reliable and valid measure of a posited construct in the age group studied would, in the first place, be convenient for use in the context of a school. This meant that the test should be manageable timewise and take no longer than 45 – 50 minutes to administer. All things considered, 62 items were eliminated from the pilot version so the research version had 149 items.

As reliability values for some of the sub-tests were low, items found responsible for this state of affairs were trimmed in order to enhance internal consistency indices. Item trimming was done across the board rather than in step-by-step manner. All the items which were either negatively correlated with the total scores or whose 'part-whole' correlations were below 0.2 were eliminated. Resultant recalculations of reliability followed each item-trimmed scale. To safeguard against 'blind dropping' of items to improve scale reliability at the expense of its validity, the candidates for deletion were examined for potential 'destructive' effect to the test's content/construct. The 'truncated' test was not thought to have any serious effect on the PLAT's validity. All in all, after the recalculations of reliabilities 11 items became candidates for deletion and were finally dropped in further analyses (as it was to be expected they were the items that had the lowest DI values).

The sub-tests on which reliability enhancement was performed were the two 'comprehension' components of PLAT. Their reliability indices were lower than those of the remaining sub-tests. The reliability improvement on the two sub-scales was not impressive but was worth the trouble: 0.007 and 0.038. Out of the three sub-tests for which the assumption of independence of items was not met, *dictation* and a gap-fill *sprawcy* had fairly low reliability indices, however no item was considered for deletion. Dropping items from the two scales would have substantially undermined their content validity. The last 'dependent-item' sub-test i.e., C-test *pojezierze* displayed quite satisfactory reliability value.

2.7.4. Exploratory Factor Analysis (EFA)

The correlational technique of FA is used several times in the analysis of the data. Firstly, as a preliminary step in addressing **RQ 1** as formulated in the *Introduction*, FA is used to construct-validate the TPJ and PLAT tests (*section 2.7.4.1.*) and then to determine the structure of FL aptitude (*section 2.7.4.2.*) thus addressing **ARQ 1** as formulated in *section 0.2.* Secondly, EFA is applied to all the variables to address the **RQ 1** addressing the issue of equivalence or independence of cognitive and linguistic abilities (*section 2.7.4.3.*). The nature of SLP ability structure, as measured by KET-A, is dealt with in *section 2.7.4.4* thus addressing **ARQ 2**.

The method of factor extraction used throughout the study is that of principal axis factoring (unless otherwise stated) with the varimax method of rotation. For each analysis, factors' eigenvalues are given first, optionally followed by 'scree' plots and unrotated factor loadings. Rotated factor solutions are reported last. In the cases when the resultant factors are not orthogonal (uncorrelated) but display a considerable degree of correlation (around and above 0.65) a hierarchical exploratory factor analysis (Schmid and Leiman 1957) is run to seek second- and first-order uncorrelated (orthogonal) factors[3].

2.7.4.1. Construct validation of KET-A

In order to address **ARQ 2** the structure of second language proficiency, as measured by KET-A criterion test, was analysed with the help of factor analysis. The analysis revealed one first-order (general) factor and two second-order (specific) factors. The emergence of the general language proficiency factor seemed to be in accord with previous research (cf. the discussion in Carroll 1983). The specific factors coincided with two traditional dimensions of multi-dimensional four-skill view of language proficiency and were labelled *SLPCompetence* and *SLPSkills* respectively. In further analysis they were used as dependent variables (criteria) to address RQ 2b (cf. *section 2.7.5.*).

2.7.4.2. Construct validation of TPJ and PLAT

As already mentioned, TPJ's content was found to be at odds with the results of closer scrutiny carried out by the present author. To see what the test's ability structure really looked like an analytical procedure which would reveal it was required. This was afforded by a multivariate technique of factor analysis. The result of TPJ's construct validation was a pattern of factor loadings that could not be reconciled with the description postulated by Kuliniak (2002:64) and discussed in *section 2.4.* Two patterns emerged from FA as potential solutions to the problem of ability dimensions purportedly measured by TPJ. One was a three-factor and the other a two-factor solution. Although a two-factor solution yielded satisfactory reliability indices for two isolated scales (thus assumed to be unidimensional) no viable interpretation of these factors/ability dimensions in terms of common, underlying abilities could be offered. Consequently, only two variables (sub-tests) which most consistently patterned together at all stages of FA and whose reliability indices were the highest were selected for further analyses from an overall number of eight variables (sub-tests) proposed by Kuliniak. They were the test's Task 7 and 8, from now on referred to as *TPJ 7* and *TPJ 8* (cf. *section 2.4.*).

With regard to PLAT, the analysis of its latent ability structure using the technique of factor analysis revealed three factors which lent themselves to a much easier interpretation in terms of postulated abilities than was the case with the TPJ measure (cf. the results in *section 3.6.3.*).

2.7.4.3. Factor Analysis of FL aptitude

Exploratory factor analysis was performed on the combined scale of fifteen variables: 13 PLAT and 2 TPJ variables, thus addressing **ARQ 1** (cf. *section 0.2.*). A four-factor solution was chosen as lending itself best to interpretation in terms of common abilities characterising each of the four factors. Those factors were tentatively labelled: *Aptitude 1* to *Aptitude 4*[4] and were later relabelled as General Verbal Skills (*GenVrbl*), Inductive Reasoning (*IndReas*), Paired-Associate Memory (*AssocMem*) and Memory for Text Comprehension (*TextComp*). The potential of FL aptitude factors extracted through FA to predict the subjects' L2 proficiency was subsequently tested using multiple regression analysis (cf. *section 2.7.5.* and *3.9.*).

2.7.4.4. FA of SLP, FL aptitude and intelligence variables

As mentioned in *section 2.7.4.*, in order to address **RQ 1**, factor analysis was used to test the hypothesis that intelligence and linguistic skills are equivalent (or independent) of each other. All in all, thirteen variables were factor analysed:

1. four FL aptitude variables factored out from 15 variables in *section 2.7.4.3.*:

   GenVrbl - General Verbal Skills

   IndReas - Inductive Linguistic Reasoning

   AssocMem - Paired-Associate Memory

   TextComp - Memory for Text Comprehension

2. three intelligence variables (cf. *sections 2.3.2* and *2.3.2.*):

   PVoc - Passive Vocabulary of TJL : verbal intelligence

   AVoc - Active Vocabulary of TJL : verbal intelligence

   SRPM - Raven's Progressive Matrices : non-verbal reasoning

3. six second language proficiency variables obtained from KET-A: *Listening Comprehension, Reading Comprehension, Grammar, Vocabulary, Writing* and *Speaking*.

After the inspection of latent roots (eigenvalues) and a scree plot a principle component factor extraction was varimax rotated to yield a solution which lent itself best to linguistic interpretation in terms of ability structure sought. The results of factor analysis showed that neither the 'unitary' nor the 'completely divisible' hypothesis about the organisation of the three ability domains could be tenable. The pattern that emerged from the analysis suggests a considerable degree of independence of SLP from both FL aptitude and intelligence faculties, which is in line with the results of many an earlier studies done on this issue. On the other hand, the pattern for FL aptitude and intelligence variables, while turning out to be factorialy more complex than in the case of SLP, did not offer any support for the 'completely divisible' view of cognitive abilities. The full results are reported in *section 3.8.*

Summarising this section of the study, it can be remarked that construct validation of PLAT and KET-A revealed a more readily identifiable structure of L2 abilities in PLAT and of second language proficiency in KET-A than was the case with L2 abilities purportedly measured by TPJ.

2.7.5. Multiple Regression

While the procedure of FA is commonly applied to one set of different variables with a view to describing, organising and interpreting the multiplicity of intercorrelations between the variables, the technique of multiple regression is applied to two different sets of data (dependent and independent variables) with a view to predicting behaviour or test performance on the basis of a limited set of variables. As such, multiple regression is interested in making as accurate predictions about the unobservable, future behaviour as possible. In other words, it functions as time and effort-saving device and seeks compromise between maximum levels of prediction and minimum levels of expenditure.

The regression method used in the present study is that of stepwise selection of variables for the regression equation. The method was considered most appropriate to the data-driven nature of the study. As noticed by Rietveld and Hout: "Stepwise selection is by far the most widely applied method, especially when the regression analysis is more explorative." (1993:97). Its advantage over other regression methods, such as forward or backward selection, is that a variable, once entered in the regression equation, does not have to stay in it when other variables are added, but can be removed if its characteristics change due to the dynamic nature of interrelations among the remaining variables (cf. *section 3.9.*). For each multiple regression analysis a correlation matrix is given followed by reports of multiple correlation coefficient R and R^2 (square) analysis of variance and the contribution of each variable to the prediction.

2.7.5.1. KET-A (Total) – Multiple Regression

In order to address **RQ 2a** as formulated in the *Introduction*, four aptitude and two intelligence variables (*SRPM* and *AVoc*[5]) were entered into multiple regression equation using a stepwise method (cf. stage 3 in *Table 2.4.* at the end of this chapter).

The decision as to which solution is preferable was made on the basis of the percentage of explained variance (the value of R^2). It was found that the solution combining the following independent variables: Active Vocabulary (*AVoc*), Inductive Reasoning (*IndReason*), General Verbal (*GenVrbl*) and Memory for Written Text Comprehension (*TxtComp*), gave in effect the highest value of the multiple correlation coefficient (R). Eventually, this subset of variables was accepted as best predicting the results of second language proficiency KET-A Total measure in the sample studied. The implications and limitations of the prediction model were then discussed.

2.7.5.2. KET-A (Competence and Skills) – Multiple Regression

In order to address **RQ 2b**, the relationship of all six independent variables was further investigated in two regression analyses: first with *SLPCompetence* and then with *SLPSkills* as criteria. As already mentioned in *section 0.2.* and *1.3.3.* and *1.4.* this was done to see if two broad language domains (*Competence* and *Skills*) factored out from KET-A criterion test could be differentially linked to different sets of variables in a linguistically valid, non-accidental way. In this connection Carroll remarks:

> *Certain kinds of regression analysis results (...) further suggest that foreign language skills are to some extent differentiable, and even differentially predictable (...). For example, in the study of college foreign language majors, canonical regression analyses showed that the relative superiority of students in reading and writing as opposed to listening and speaking could be predicted from Part 4, Words in Sentences, of the Modern Language Aptitude Test* (1983: 95).

However crude the delineation of the SLP abilities into the two broad areas was (no fine-grade prediction between, let's say, speaking-writing or reading-listening were made in lack of sufficient number of variables) it still produced some interesting results concerning the nature of the prediction and, indirectly, the structure of second language abilities.

Namely, as the results in *Chapter 4* will show, *SLPCompetence* was best predicted by the verbal intelligence measure, *AVoc* and by two aptitude variables: *TextComp* and *AssocMem*, to the exclusion of *GenVrbl*, *IndReas* and *SRPM*. The other SLP component – *Skills,* relied on two FL aptitude variables (*GenVrbl* and *IndReas*) and on *AVoc*. to a noteworthy exclusion of a rote memory variable (*AssocMem*), *TextComp* and *SRPM.*

The stages of the study, their aims and the analytical tools used and described in the sections of this chapter are summarised in the *Table 2.4.*

	Test	**Analysis**	**Statistic**	**Aim of analysis**
Stage 1	PLAT	1. item analysis 2. reliability 3. item trimming	FV and DI Cronbach's α ; Guttman	to specify **reliability** of the tests' sub-scales
	KET-A	reliability estimates	Cronbach's α ; Guttman	
	SRPM and TJL	reliability estimates	Speraman-Brown	to compare **reliability** with existing indices
	TPJ	factor analysis	factor loadings	to extract test's sub-scales for **reliability** estimates
Stage 2	KET-A	factor analysis	factor loadings	**ARQ 1**– construct valid.
	TPJ	factor analysis	factor loadings	**ARQ 2** – construct valid.
	PLAT	factor analysis	factor loadings	**ARQ 2**– construct valid.
	TPJ 7 and 8 + PLAT	factor analysis	factor loadings	**ARQ 2** – to extract L2 aptitude factors
	SLP+L2 Apt. +Intelligence	factor analysis	factor loadings	**RQ 1** – structure of cognitive abilities
Stage 3	4 Aptitude +SRPM +AVoc	multiple regression Criterion: KET-A	multiple correlation – R squared mult.corr. – R^2	**RQ 2a:** to seek the highest level of **prediction** of general L2 proficiency (KET-A)
Stage 4	4 Aptitude + SRPM +AVoc	multiple regression Criterion: SLPComp and SLPSkills	multiple correlation – R squared mult.corr. – R^2	**RQ 2b:** to seek the highest level of **prediction** of two components of SLP: (*Comp* and S*kills*)

Table 2.4. The summary of the design of the study

NOTES

1. FV and DI values from Kuliniak's standardisation study are not reported here.
2. For more details on the issue of test validity cf. any book on tests design i.e., Bachman 1997; Niemierko 1999; Hornowska 2003.
3. More detailed explanation of the technique and its terminology together with a worked example of factor analysis done on Raven Progressive Matrices is included in *Section 3.5.*
4. The factors are interpreted and labelled in *sections 3.6.3. and 4.2.*
5. Of two TJL sub-tests only *AVoc* (Active Vocabulary Test) was included in MR rather than *PVoc* (Passive Vocabulary Test). This was done to avoid the problem of collinearity between two strongly correlated variables (r=0.78) had both *PVoc* and *AVoc* been entered into MR (cf. assumptions for MR in *section 3.2.*). Another reason was that given a nearly perfect correlation of both sub-tests with the total TJL score: PVoc x TJL = 0.95; AVoc x TJL = 0.94 (cf.. *Table 4.5*) either could be regarded as representing the whole TJL score, and since passive vocabulary was well represented in PLAT (cf. the two tasks: *dbrdszn* and *zwolennik*) it was the active sub-scale of TJL that was chosen.

CHAPTER 3

RESULTS OF STATISTICAL ANALYSES

In this chapter descriptive statistics of the 22 variables (5 test totals and 17 sub-tests) are reported first in *section 3.1.* This will be followed, in *section 3.2.* by a short discussion of the assumptions (preconditions) underlying the statistical analyses used in the study and described in *section 2.7.* The reliability estimates of the test variables used will be presented in *section 3.3.* Initial correlation matrix for the test measures and the criterion are given in *section 3.4.* The technique of factor analysis is briefly introduced in *section 3.5.* on an illustrative example of the intelligence variables. The results of the validation of the TPJ, PLAT and KET-A testing tools are discussed in *section 3.5.* The next two sections give the results of factor analysis used in the exploration of internal structure involved in: 1. FL aptitude (*section 3.7.*) and 2. SLP, FL aptitude and intelligence (*section 3.8.*). Finally, *section 3.9.* presents the results of multiple regression analysis seeking the optimal prediction of the criterion variables.

3.1. Descriptive Statistics

Descriptive statistics as described by Brown is "a numerical representation of exactly how each group (of subjects) performed on the interval scale measures" (1988:67). Two groups of statistics are customarily given for any scale: central tendency and dispersion measures. The tables below additionally report figures describing the distribution characteristics for the variables (skewness and kurtosis). *Table 3.1.* reports relevant statistics for the test instruments used in the study.

	N*	Mean	St. Dev.	Skewness
KET-A	210	41,5	20,77	0,24
Compet	210	0,1	0,84	0,27
Skills	210	0,1	0,82	0,20
SRPM	212	45,3	7,48	-1,18
TJL	216	24,8	9,20	-0,16
TPJ	216	22,0	7,85	-0,09
PLAT	153	75,7	24,61	-0,05
*L2Aptitude***	149	80,7	28,06	-0,26

** - cf. section 2.2. about the number of subjects in each test*
*** - combined aptitude measure of the results of two TPJ tasks and PLAT (cf. section 2.7.4.2.)*

Table 3.1. Descriptive statistics for 5 test totals

Table 3.2., in turn, contains descriptive statistics for 15 variables on which factor analysis was run and 8 variables which entered multiple regression analysis (cf. discussion in *section 2.7.*).

Variables in Factor Analysis	N*	Mean	St Dev	Skew
TPJ 7	216	3,7	2,29	-0,38
TPJ 8	216	3,1	1,75	-0,02
PLAT ofiara	153	3,8	1,65	-0,41
PLAT handlarz	153	3,3	1,30	-0,68
PLAT układ	153	6,4	3,32	-0,77
PLAT przed	153	1,6	1,01	-0,26
PLAT podczas	153	4,5	1,31	-0,94
PLAT lubić	153	3,8	1,64	-1,19
PLAT pamiętać	153	5,3	3,74	0,02
PLAT dyktando	153	3,4	1,64	-0,29
PLAT wingdings	153	3,9	1,54	-1,45
PLAT sprawcy	153	6,6	2,83	-0,68
PLAT pojezierze	153	15,7	7,98	0,04
PLAT dbrdszn	153	6,9	4,02	0,01
PLAT zwolennik	153	9,8	4,04	-0,07
Multiple Regression				
GenVrbl	146	0,1	0,90	-0,06
IndReas	146	0,1	0,84	-0,14
AssocMem	146	0,1	0,80	-0,56
TextCompr	146	0,1	0,87	-0,50
SRPM	212	45,3	7,48	-1,18
TJL	216	24,8	9,20	-0,16
PVoc	216	16,1	5,05	-0,41
AVoc	215	8,6	4,72	0,18

* *cf. section 2.2. about the number of subjects in each test*

Table 3.2. Descriptive statistics for FA and MR variables.

3.2. Assumptions

In most statistical procedures of data analysis certain assumptions are always made with reference to the arrangement of scores along the continuum as well as to the mutual intercorrelation of scores or set of scores. Depending on the statistical tool used various conditions must be fulfilled lest serious errors resulting in invalid generalisations from the data will be made. In the present study the data were examined whether they met the following assumptions (Brown 1988, Rietveld and Hout 1993).

1. a high measurement level for all the variables i.e., interval or ratio
2. independence of scores from all other scores across the observations
3. normal distribution of variables (scores from test instruments used)
4. linear relationship between any two set of variables
5. constancy of error across observations (homoscedasticity condition)
6. no intercorrelation between the independent variables (no multicollinearity)

While conditions 1 to 4 must be assumed for even the simplest of statistical analyses, the assumptions of homoscedasticity and lack of multicollinearity are typically made with the correlational technique of multiple regression. A discussion of the assumptions for the present study follows.

3.2.1. Measurement scales.

At its simplest, data collection process in psychological and educational research consists in assigning numbers to observable tokens of measured behaviour. This results in the data being organised on scales of measurement. Depending on the degree of precision afforded by the measurement, different types of scales obtain. Arranged from least to most precise measurement levels the scales are:

1. nominal scales – reflecting categorical and grouping characteristics of data, also labelled as discrete or discontinuous i.e., Female / Male.
2. ordinal scales – reflecting a relative position of value on a scale i.e., ordering data from lowest to highest value of a parameter
3. interval scales – as ordinal but including information of equidistance between points on the scale
4. ratio scales – as interval but including a comparison of points on the scale to an absolute zero.

In the present study the use of a ratio scale makes little sense as one cannot claim with any degree of certainty that a particular behavioural parameter for example, anxiety level has a zero value for any subject studied. Consequently, the second strongest measurement level is applicable i.e., interval scale expressed either in raw scores or percentage values. The assumption of the interval scale of measurement for variables is a prerequisite for the statistical techniques of correlation, factor analysis and multiple regression carried out in the present study.

3.2.2. Independence of scores.

There are two aspects of this assumption. The first requires that two sets of scores, for one or more than one individual, do not depend on each other. For example, if an individual A copied from an individual B the two scores would be dependent. Similarly, if an individual A was given the same test twice within a short time period their two sets of scores will probably not be independent. The data of the present study meet the assumption of score independence as: 1). no test instrument was administered twice to the same group of individuals, and 2.) individuals suspected of malpractice were eliminated from further analyses (cf. *sections 2.2.* and *2.6.*). With regard to suspicion of malpractice, a possibility cannot be ruled out that a limited set of test results may display dependence of scores due to occasional copying practices which might have gone unnoticed by the proctors despite their 'heavy policing' during the tests. However, every effort was made that those be isolated cases, affecting only a margin of scores. The second aspect assumes that no set of scores is subsumed in any other set of scores. For example, no analysis is carried out on a correlation matrix involving a test's total scores and any of its components. This assumption is particularly important for factor analysis.

3.2.3. Normal distribution.

This assumption requires that the frequency distribution of scores in each of the variables follow the pattern of a bell-shaped curve with a 50/50 split in scores around its central point. While such ideal symmetry is hard to come by, it has been observed that the more numerous the samples and the greater the range of scores, the greater the likelihood of normal distribution. Typically, the normality of score distribution is revealed by the examination of its skewness (cf. Brown 1988 for the discussion of skewness and kurtosis) and/or mean and median score.

The examination of skewness in *Tables 3.1.* and *3.2.* above reveals 4 variables with values above or approaching 1 (printed in bold). It is customary to consider the data as deviating from normality if the value for skewness and/or kurtosis exceed 2 (Sasaki 1999:67). Accordingly, the four 'suspicious' variables are considered to have normal distribution. This is additionally confirmed by examining the values for median and mean. They very much coincide, which is a sign of a normal distribution. However, an inspection of a graphical representation of the score distribution in all 22 variables reveals that two variables: TPJ *Analysis 1* (Task 7) and PLAT *pamiętać* have a distinctly bimodal pattern of score distribution. Typically, a bimodal curve is a sign that there is something unusual in the sample tested with respect to the feature investigated. Alternatively, it may be a signal to a researcher that this particular test or sub-test needs substantial modification. In the context of the present study this violation of normality of distribution for the two variables is acknowledged. The values of the Shapiro-Wilk *W* statistic (S-W *W*) test of normality are reported in *Table 3.3.* below (cf. the histograms for the first measures in *Appendix 6*).

Variable	N	SW-W	p level
KET-A	210	.977	.001
SRPM	212	.923	.001
TJL	216	.989	.083
TPJ	216	.973	.001
PLAT	153	.982	.047
TJL PVoc	216	.978	.001
TJL AVoc	216	.980	.003

TPJ Analysis 1	216	.817	.001
TPJ Analysis 2	216	.943	.001
PLAT ofiara	153	.925	.001
PLAT handlarz	153	.902	.001
PLAT układ	153	.862	.001
PLAT przed	153	.867	.001
PLAT podczas	153	.865	.001
PLAT lubić	153	.737	.001
PLAT pamiętać	153	.871	.001
PLAT dyktando	153	.941	.001
PLAT wingdings	153	.725	.001
PLAT sprawcy	153	.930	.001
PLAT pojezierze	153	.980	.023
PLAT dbrdszn	153	.958	.001
PLAT zwolennik	153	.989	.298

Table 3.3. Shapiro-Wilk *W* test for normality of distribution for 5 tests and 17 sub-tests.

3.2.4. Linearity.

This assumption stipulates that the relationship between the set of scores on the scatterplot (a graphical representation of the relationship) be linear i.e., any incremental change in a variable X must be accompanied by an increase in a variable Y in the same direction. To check the data against possible violations of this assumption, bivariate scatterplot diagrams were checked for signs of nonlinearity and/or restriction of the range of scores. No such violations were found. The scattergrams were also inspected visually for potential outliers i.e., cases which lie far away from the main regression line and which have a 'deflating' effect on the magnitude of the correlational relationship. Seven potential outliers were identified, temporarily removed and the data re-plotted and recalculated to see what effect the removal would have on the correlations. Since the deflating effect of the outliers was considered small (from 0.008 to 0.01) they were retained for further analyses and dealt with at the later stage when multiple regression analysis for the variables was run (cf. *section 3.9.*). Thus the data were assumed to meet the linearity condition. The following two assumptions apply only to multiple regression.

3.2.5. Homoscedasticity (homogeneity of variance).

As noticed by Brown "the problem of heteroscedasticity occurs when the error variance is not constant over all observations" (1988:149). In practice it suffices to inspect the variable's histogram to detect serious deviations from homoscedasticity as it appears when a distribution is strongly skewed (Guilford 1960:177). The examination of descriptive statistics for the variables used in multiple regression in *Table 3.2.* reveals that neither of the seven variables entered in the analysis violates the assumption.

3.2.6. Multicollinearity (or: collinearity).

This requirement specifies that the independent variables used in predicting the most variation in a dependent variable should not be highly correlated. Otherwise, when there is a dependency between the variables to be entered in the regression equation "it becomes almost impossible to assess the separate contribution of the independent variables to the prediction of the criterion" (Rietveld and Hout 1993:159). Through progressive manipulations of the Pearson's ***r*** between two independent variables in MR the critical value of collinearity has been set at approximately 0.700. The problem can be easily spotted if correlation matrices of independent variables are examined. In the present study *TJL PVoc* and *TJL AVoc* are the only variables which show very high dependence on one another: ***r*** = 0.776 (N=211) in *Table 3.6.* and ***r*** = 0.797 (N=112) in *Table 3.2.3.* Consequently, their inclusion in MR would result in uninterpretable solution. The problem has been solved by including only one TJL sub-test (*AVoc*) into MR with four aptitude factors extracted in exploratory factor analysis (cf. *section 3.7.*). The examination of the correlation matrix in *Table 3.26.* on page 115 shows that no two variables excede the critical value for multicollinearity.

3.3. Reliability estimates

Table 3.4. below gives reliability estimates for the L2 proficiency measure (KET-A) and two aptitude measures: PLAT and TPJ. As described in *Section 2.7.*, two types of reliability indices are used: Cronbach's α - for independent scales, and the Guttman split-half estimate - for dependent scales (marked with **).

Scale (variable)		Reliability	No of items
KET-A*	N=210		
Listening		.867**	9/9/9/14
Reading		.869**	12/12/12/15
Grammar		.927	19/19/16/10
Vocabulary		.880	10/10/8/18
PLAT***	N=153		
ofiara+handlarz		.712	11
układ		.883	10
przed+podczas		.527	9
lubić+pamiętać		.914	15
dyktando		.627**	6
wingdings		.798	5
sprawcy		.756*	10
pojezierze		.912**	36
dbrdszn		.846	16
zwolennik		.802	20
TPJ	N=227		

Group 1	.219	4
Group 2	.765	16
Group 3	.651	10
Group 4	.871	6
Group 5	.684	6
*Group 4&5*****	*.862*	*12*

* For KET-A averaged values from four versions of the test are given for each component
** Guttman split-half reliability index for dependent scales (cf. *section 2.7.*)
*** all the values after item trimming for reliability enhancement (cf. *section 2.7.*)
**** cf. end of *section 3.5.* for the discussion of this point

Table 3.4. Reliability estimates sub-tests of KET-A, PLAT and TPJ

The estimates were made for each unidimensional scale rather than for the whole test (cf. the discussion in *section 2.7.*). In the case of KET-A the scales were the four tested components of language proficiency. In the case of PLAT reliability estimates were made for each sub-test (with the exception of two very closely-knit tests of the 'progressive learning' type (cf. *section 2.4.2.*). PLAT's estimates were the results of the scale enhancement and low-reliability item trimming (cf. *section 2.7.3.*). In the case of TPJ, reliability estimates are quoted for the dimensions declared by Kuliniak (2002). The problem with the validity of his delineations was already briefly touched upon in *sections 2.4.* and *2.7.* where it was said that it is not enough to claim that a test measures a given ability. What needs to be done is to demonstrate through the analysis of underlying traits the validity of a postulated construct. This will be discussed in the next section. For the moment suffice it only to emphasise again that Kuliniak's Group 4 and Group 5 abilities when reinterpreted in terms of Carroll (1979, 1993) as one ability i.e., 'inductive language learning ability' has the highest reliability index: $\alpha = 0.862$ of the whole TPJ test. All in all, the reliabilities reported in *Table 3.4.* above are more than satisfactory for KET-A and satisfactory for PLAT.

3.4. Initial correlation matrix.

Table 3.5. contains the correlations of the four test instruments (independent variables) with the results of the achievement/proficiency final L2 test (dependent variable) as well as the intercorrelations between the measures.

N = 138				
	SRPM	**TJL**	**TPJ**	**PLAT**
TJL	0.42			
TPJ	0.48	0.66		
PLAT	0.52	0.79	0.72	
KET-A	**0.38**	**0.61**	**0.62**	**0.60**

all correlations significant at p<,01

Table 3.5. Initial correlations between independent variables and the criterion

The values of the Pearson's ***r*** for TJL, TPJ and PLAT against the criterion are almost identical and, given the sample, can be characterised as moderately high. On the other hand, the value of the coefficient for the Raven's test of non-verbal intelligence and KET-A is medium low.

In this connection an important methodological remark must be made. It is not an isolated fact that when investigating relationships between phenomena studied researchers all too readily rely on simple correlational statistics. In doing so they run the risk of claiming relationships between variables when in fact no, or insignificant, links appear. The problem of positing non-existing relationships becomes even more aggravated when poorly conceived criterion tests are employed or worse, when unreliable and irrelevant measures are used as predictors. Another thing is that generalisations made on the basis of simple correlational evidence are not likely to yield any stable and valid predictions of the criterion. Accordingly, one might be led to believe by inspecting the correlations in *Table 3.5.* that the three testing tools make equally valid contributions to the prediction of the criterion. As will be seen in *sections 3.9.* and *4.4.* this is an entirely unjustified claim.

3.5. Factor analysis – an illustrative example

Before the analytical procedure of factor analysis is applied to FL aptitude measures of TPJ and PLAT its workings will be discussed on a straightforward example of the two intelligence measures i.e., SRPM and TJL.

As it was said in *Chapters 1* and *2*, where the notion of general cognitive abilities was discussed, the SRPM and TJL tests address different types of abilities thought to make up the construct of intelligence (Spearman 1927, Cattell 1971, Raven 1988). Accordingly, it has become customary to distinguish two intelligences: one, experience-based (crystallised intelligence) and the other, experience-independent (fluid intelligence). Raven's test of *Progressive Matrices* and his *Mill Hill Vocabulary Scale* (cf. *section 2.3.*) are just one of a few operationalisations of the concept of the two types of intelligence. If the two tests really measure two different aspects of a higher-order construct one can expect these facts to be reflected in the relevant statistics. And indeed, this is the case. *Table 3.6.* below presents correlations between the tests' totals and tests' components.

N= 211								
	SRPM	TJL	PVoc	AVoc	SRPM A	SRPM B	SRPM C	SRPM D
TJL	0.52							
PVoc	0.51	0.95						
AVoc	0.47	0.94	0.78					

SRPM A	0.65	*0.39*	*0.39*	*0.35*				
SRPM B	0.73	*0.36*	*0.35*	*0.32*	0.45			
SRPM C	0.82	*0.43*	*0.46*	*0.35*	0.50	0.57		
SRPM D	0.80	*0.39*	*0.40*	*0.34*	0.46	0.53	0.57	
SRPM E	0.77	*0.41*	*0.37*	*0.40*	0.42	0.40	0.51	0.53

all correlations statistically significant at p=0,01

Table 3.6. Correlation matrix for TJL, SRPM and components

A cursory look at the part of the matrix showing correlation between SRPM and TJL totals will lead one to conclude that the two tests are significantly correlated (p=0.01) and that, given the size of the sample, the strength of the correlation (0.52) is moderate. This is only to be expected since, according to the two-factor model mentioned above, the two types of intelligences converge on one 'general' intelligence factor without being very closely intercorrelated themselves.

A closer examination of the matrix reveals a more complex nature of the relationship. What one cannot fail to notice is that the correlations between the components within a test (**in bold**) are higher than between the components across the two tests (*in italics*). Accordingly, Pearson's ***r*** for TJL PVoc x TJL AVoc = 0.78[1], and within SRPM's components it ranges from 0.40 to 0.57, whereas the correlations PVoc and AVoc x SRPM (A, B, C, D, E) range from 0.32 to 0.46. Thus a clear pattern seems to emerge from the matrix above: there are two groups of variables inversely correlated. In other words, it looks as if there are two latent groups of abilities (factors) which underlie the correlational relationships among the 7 variables. Obviously, this multivariate pattern is difficult to disentangle without a tool more powerful than just a simple examination of the correlation between two variables only. Factor analysis is just the procedure that enables a researcher to handle a multiple set of correlations between more than two variables (factor analysis is in fact a more complex variety of a simple correlational analysis).

Quite a few psychometricians contributed to the development of factor analytical technique of data analysis but the names of Spearman, Thurstone and Hottelling are invariably mentioned when the acknowledgements are made. Generally speaking, there are two types of factor analyses depending on what method of factor extraction is used. The two methods of factor extraction are: *principal component* (PC) and *principal factor* (PF). The methods differ in how they treat the variance obtaining between the variables. The PC analysis aims at explaining the total variance by means of its components, while the PF analysis aims at explaining common variance between the variables. In other words, PC treats variance as indivisible unity, whereas the PF splits it into common and unique variance (unique variance = variables' specific variance + error). There are many ways in which the total variance can be analysed into common and unique variance, each being a separate PF method. The PF method used in the present study is the *principal axis* factor extraction (unless otherwise specified). In more practical terms, the PC method is used when the aim of the researcher is to reduce a large body of data to a smaller, manageable portion, while the PF method is used when the structure underlying the variables is sought.

Two concepts are basic to the understanding of factor analytic methods: **dimensionality reduction** and **simple structure**. The former concept is in effect a **reduction of a two-dimensional space** (formed by plotting two variables against each other on a scattergram) to a one-dimensional space represented by the best-fitting straight line, as can be seen in *Figure 3.1.* This regression line, called also principal

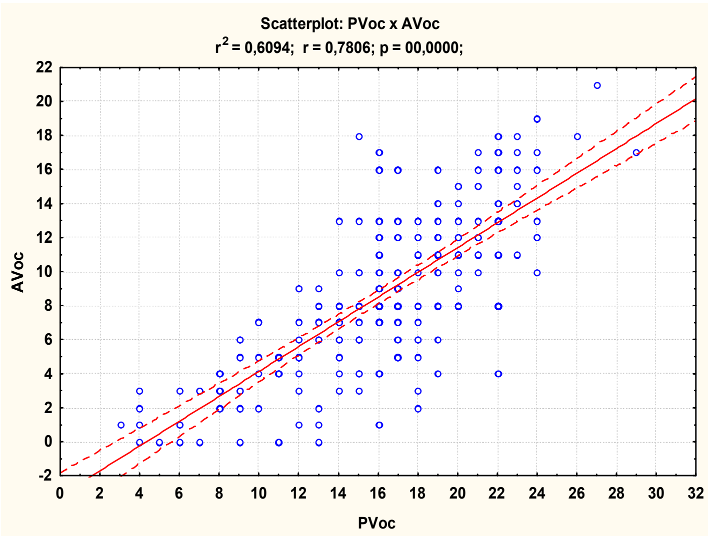


Figure 3.1. Scattergram for AVoc x PVoc with a regression line

component or axis, becomes thus the only representation of the two variables. In a way, it represents, in a 'condensed' form, the original variables. The projections of the original (two-dimensional) scores on the regression line form the 'new' (one-dimensional) scores. The procedure of dimensionality reduction becomes increasingly more complex when the number of variables increases. Extraction of factors for each pair of variables in a multidimensional (multivariate) space leads to **structure** exposure, another interesting property of factor analytic procedures. Namely, through the analysis of extracted components[2] (regression lines) and their manipulation on an X / Y two-dimensional space it is possible to find regularities in a large number of confusingly intercorrelated variables. In other words, the technique offers the opportunity to see beyond and through the multiplicity of intercorrelations between variables and to discover a simpler structure than could otherwise be afforded by the raw data.

Apparently, a certain loss of information about the original variables is unavoidable in any regression. The degree to which it happens is reflected in the 'strength' of the new dimension which determines a variable-specific contribution to each factor, the so-called **factor loading**. The procedure of factor extraction, using either PC or PF, invariably produces a prominent first factor which is loaded by most variables. Since it is virtually impossible to interpret the structure thus obtained factor extraction is followed by factor rotation. Factor rotation aims at the discovery of a pattern of higher-order factors underlying the raw data in such a way that the factors are maximally distinct (uncorrelated) from each other.

Returning to our illustrative example, we will proceed with factor analysis along the following steps (Carroll 1992):

1. preparation of a correlation matrix for variables
2. condensation of the matrix either by principal component or principal factor method. This invariably involves a decision on the number of factors to be extracted[3].
3. rotation of factors to simple structure
4. (optional) application of a hierarchical factor analysis to seek a second-order structure solution as proposed by Schmid and Leiman (1957) in case of high correlations between factors yielded in step 3.
5. interpretation of a step 3 or step 4 factor solution

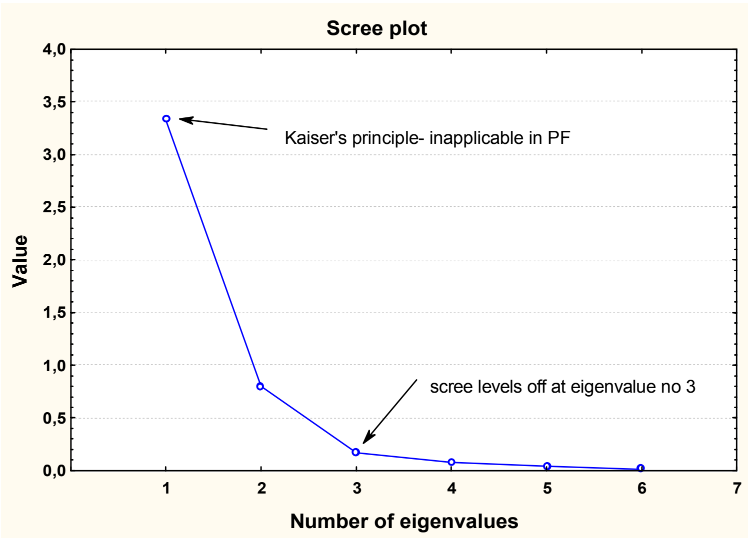
Because our aim in this illustrative example is to 'discover' a structure underlying the variables, PF will be used as the method of factor extraction. The correlation matrix in *Table 3.6.* represents **step 1** in the procedure. In **step 2** the decision on how many factors/components to extract yields information on a 'self' value (technically known as: eigenvalue) of each component. A component's eigenvalue is to be interpreted as the amount of variance explained by this component. *Table 3.7.* shows eigenvalues and percentage of explained variance for each component in the six-factor solution. As can be seen in the tabulation the seventh factor has not been extracted because of indefinitely small eigenvalue. At this point of FA a decision has to be made about how many factors to extract.

N=227	Eigenvalue	% of total variance	cumulative eigenvalue	cumulative %
1	3,34	47,64	3,34	47,64
2	0,80	11,39	4,13	59,03
3	0,17	2,37	4,30	61,40
4	0,08	1,07	4,37	62,47
5	0,04	0,61	4,42	63,08
6	0,01	0,19	4,43	63,27

Table 3.7. Eigenvalues for RPM and TJL sub-tests

The well-known rule, referred to as the 'Kaiser-Guttman' rule (Kaiser 1960) advises us to accept only those factors for extraction whose eigenvalues are greater than unity[4]. But, as Rietvald and Hout notice, Kaiser's principle makes sense only in the context of PC extraction method where "the variance of a variable is not split into different parts and all the variables have the communality of 1" (1993:273). Another principle which helps in the decision is Cattell's 'scree test'. It is a graphic representation of eigenvalues where the magnitudes of successive eigenvalues are plotted against component numbers to form a 'scree'. The rule advises us to extract the factors whose values are located **before** the break point in the scree, i.e., where it 'elbows' into a more or less straight line. Cattell's scree test is more helpful in the case of PF and points to two factors to be extracted in our illustrative example as shown in *Figure 3.2.* below.

Figure 3.2. Cattell's scree test for number of components to be extracted



Very often the two tests point to two different solutions. In such a case the decision of which solution to choose for further analysis is an arbitrary one as there is no principled way to opt between either. Neither of the two principles should be followed blindly as both have their advantages and limitations (cf. Rietvald and Hout 1993; Carroll 1999). In practice, another criterion has to be relied on i.e., the criterion of interpretability of extracted components. This is why it is not unusual in FA to try several solutions to see which lends itself best to interpretation in terms of posited (measured) construct. *Table 3.8.* below shows loadings for 6 factors extracted using

Extraction: principal axis (PF); N = 227

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	h^2
PVoc	**-0.749**	0.493	-0.107	-0.103	0.072	-0.015	*0.651*
AVoc	**-0.698**	**0.545**	0.073	0.111	-0.055	0.023	*0.625*

SRPM A	**-0.614**	-0.141	-0.031	0.011	-0.069	-0.099	*0.342*
SRPM B	**-0.652**	-0.266	-0.184	0.159	-0.017	0.027	*0.405*
SRPM C	**-0.748**	-0.261	-0.126	-0.151	-0.054	0.043	*0.512*
SRPM D	**-0.706**	-0.270	0.094	0.043	0.156	-0.006	*0.458*
SRPM E	**-0.654**	-0.159	0.299	-0.042	-0.050	0.015	*0.339*
expl. variance	47.6%	11.4%	2.4%	1.1%	0.6%	0.2%	

significant loadings in bold (cf. note 6)

Table 3.8. Loadings for six extracted factors. Unrotated solution

the principal axis method of PF extraction. All variables load significantly on the first component (negative signs are to be disregarded) as could be expected given that factor extraction is done in such a way that each subsequent component extracted accounts for progressively smaller amount of variance.

The last column gives the values of communalities *(zasób zmienności wspólnej*). Extraction communalities are estimates of that part of the variance of a variable which is accounted for by the common factors in the factor solution (Rietveld and Hout 1993). In other words, h^2 tells us what proportion of a variable's variance can be reproduced/accounted for given a particular factor solution. Very low communalities (h^2 < .200) for one or two variables in the analysis (out of many) indicate that those variables are not adequately explained by a given factor solution (STATISTICA 2003).

As has been noticed earlier, the initial factor extraction does not lend itself easily to interpretation in terms of common properties of the construct underlying the variables. In **step 3** the process of axis-rotation of uninterpretable factors aims at discovering a factor pattern simple enough to be interpreted by a researcher. Rotation to simple structure produces high correlations of the original variable with some factors and low correlations with others. The method of rotation used in the present study produces MAXimalization of VARiance of the factor loadings and is called VARIMAX. Rotation to simple structure is carried out only on the factors indicated by the Kaiser's or Cattell's rule[5]. The table below shows the result of the rotation of two factors as suggested by the scree test in *Figure 3.2.* above.

Extr.: PA	3 factor solution			2 factor solution		
	Factor 1	**Factor 2**	**Factor 3**	**Factor 1**	**Factor 2**	***h^2***
PVoc	0.262	**0.858**	0.041	0.252	**0.846**	*0.651*
AVoc	0.131	**0.856**	0.174	0.176	**0.860**	*0.625*
SRPM A	**0.529**	0.287	0.188	**0.562**	0.294	*0.342*
SRPM B	**0.674**	0.226	0.100	**0.652**	0.229	*0.405*
SRPM C	**0.716**	0.288	0.183	**0.728**	0.294	*0.512*
SRPM D	**0.626**	0.245	0.336	**0.707**	0.258	*0.458*
SRPM E	0.444	0.281	**0.527**	**0.576**	0.308	*0.339*
expl. variance	27.5%	26.1%	7.2%	31.4%	26.3%	

*significant loadings in bold (cf. note 6)

Table 3.9. Varimax rotated 3 and 2 factor solution*

For illustrative purposes a three-factor 'solution' is also included in the table. As can be seen the pattern in the three-factor solution while generally producing an interpretable structure has one variable (SRPM E) which moderately[6] loads only one factor. In two-factor solution there is a clearer factor structure with each variable displaying a more 'orchestrated' pattern of loadings. Consequently, if not for any other then for this reason, a two-factor solution is preferred.

In **step 4**, Schmid and Leiman (1957) procedure helps to determine a hierarchical structure in case the factors extracted by an earlier method prove to be highly correlated (oblique)[7]. This procedure extracts second-order (general) factors *(czynniki wtórne)* and first-order (specific) factors *(czynniki pierwotne)* by factor analysing an oblique factor matrix, which results in orthogonal (uncorrelated) factor structure (Sasaki 1999, Wesche et al. 1982). The procedure works in such a way that it: "analyses clusters of marker variables (with high unique factor loadings, and low cross-loadings) and rotates axes through those clusters. If no such clear clusters of 'marker variables' can be identified for each oblique factor, the hierarchical analysis will not be performed." (STATISTICA 6, 2003).

The next step, **step 5** involves the interpretation of the factors underlying the variables in a given *n*-factor solution. In practical terms the interpretation means seeing whether the factors thus extracted make any sense at all in the context of the researched domain. The nature of a factor can be determined by examining which variables in the analysis have the highest loadings (correlations) on this factor. The labelling of the factor involves finding a common name which would capture the nature of the contributing variables. In our example the two factors extracted pattern very neatly with the two tests and are labelled *Verbal Factor* (factor 2) and *Reasoning Factor* (factor 1). The total amount of variance explained by the solution is 57.7% and each factor's amount of explained variance is given in *Table 3.9.*

What remains to be done next is the inspection of the reproduced correlation matrix (cf. *Appendix 7*) and residual correlations. This tells a researcher how closely the extracted solution approximates the original correlation matrix and how big the difference is between the observed values of the correlation and those predicted by the factor solution. Customarily it is considered that residuals greater than 0.10 are a signal that the variance unexplained by the solution is too large and that the factor solution should be reconsidered or reinterpreted. *Table 3.10.* shows residual correlations for our illustrative example. The residuals are generally low and very low, well below the critical point.

N=211	PVoc	AVoc	SRPM A	SRPM B	SRPM C	SRPM D	SRPM E
PVoc	0,22						
AVoc	0,00	0,23					
SRPM A	0,00	-0,00	0,60				
SRPM B	-0,01	0,01	0,01	0,52			
SRPM C	0,03	-0,03	0,00	0,02	0,38		
SRPM D	0,00	-0,01	-0,02	0,01	-0,03	0,43	
SRPM E	-0,04	0,03	0,00	-0,05	0,00	0,04	0,57

Table 3.10. Residual correlations.

It may be concluded then that the two-factor solution represents well the original variables with very little variance being unexplained. In the situation of a real research though, such an elegant factorial analysis hardly obtains. Usually, the situation is more complex and sometimes, as it will be seen in *section 3.6.2.*, when no feasible interpretation of resultant factors can be made the analysis can not be reckoned successful.

To conclude this succinct introduction to factor analysis, two points need to be mentioned. First, in order for factor structure to be effectively interpretable it is preferable that there should be more than two variables potentially determinate for one factor (Carroll 1992:89). Otherwise, the amount of variance and covariance afforded by the interplay between akin variables might not be high enough to allow a researcher to claim that a postulated factor has been extracted from the underlying data. Second, as already mentioned, in the present study the limitations of the principal component (PC) method of factor extraction (*model składowych głównych*) as formulated in Carroll (1983,1992); Vollmer and Sang (1983); Rietvald and Hout (1993); Ferguson and Takane (1999) and Zakrzewska (1994) are acknowledged[8] and the principal factor (PF) method (*model czynników głównych*) is used for structure discovery.

The discussion of more detailed points of factor analysis including its various types and their limitations is obviously beyond the scope of this cursory introduction. More on factor analysis can be found in, for example, Gorsuch (1974); Hatch and Farhady (1982); Carroll (1993); Rietveld and Hout (1993) or Zakrzewska (1994).

3.6. Results of construct validation of KET-A, TPJ and PLAT.

3.6.1. KET-A's construct validation.

Following the discussion of the structure of second language proficiency (SLP) in *section 1.4.* it was found necessary to factor analyse the test of language proficiency used in the study to determine the construct of SLP. Specifically, it was hoped that the statistical procedure of exploratory factor analysis would uncover an underlying, 'true' structure of the criterion test (KET-A) thus enabling the researcher to define L2 competencies measured by the test. The extracted and interpreted factors would then be used as dependent variables in multiple regression analysis (together and/or apart from the total KET-A score) to investigate how the independent variables predict more specific competencies unravelled by EFA.

KET-A's validation analysis starts with the correlation matrix in *Table 3.11.* Initial factor extraction using the principal axis extraction method was run on the matrix

N= 206					
KET – A	**LC**	**RC**	**GR**	**VC**	**Wr**
ListComp					
ReadComp	0.63				
Grammar	0.44	0.60			
Vocab.	0.45	0.63	0.69		
Writing	0.56	0.64	0.54	0.54	
Speaking	0.40	0.47	0.52	0.40	0.56

all r significant at $p<0.01$

Table 3.11. Correlation matrix for KET-A variables (sub-tests)

to determine the number of factors to be extracted. This was done by examining the factors' eigenvalues and a scree plot. Two factor solution was finally chosen with 3,32 eigenvalue for Factor 1 and 0,30 for Factor 2, and 60.6 % of explained cumulative variance of KET-A. *Table 3.12.* gives factor loadings for the two-factor rotated solution and the communalities for each variable.

N= 206	Extraction: principal axix		
	Factor 1	**Factor 2**	h^2
ListComp	0.293	**0.663**	*0.440*
ReadComp	0.494	**0.668**	*0.602*
Grammar	**0.804**	0.317	*0.568*
Vocab.	**0.717**	0.356	*0.559*
Writing	0.421	**0.687**	*0.542*
Speaking	0.402	0.467	*0.388*
expl.variance	30.5%	30.1%	

factor loadings > 0.500 in bold

Table 3.12. Varimax rotated 2-factor solution for KET-A's 5 variables

The first factor receives heavy loadings from two variables: *Grammar* (0.804) and *Vocabulary* (.717) and is also significantly loaded by three variables (the critical value of factor loading for N=210 is 0.300). Similarly, Factor 2 while receiving heavy loading from the variables

ListComp (0.663), *ReadComp* (0.668) and *Writing* (0.687) is also significantly loaded by the remaining three variables. The variable *Speaking* loads the two factors almost equally which indicates that there is quite an amount of shared variance between the two factors. Accordingly, the first factor can be interpreted as language knowledge or language competence factor: *SLPComp,* while factor 2, comprising the four skills variables, is interpreted as language skills factor: *SLPSkill.* The pattern of factor loadings suggests that there is a substantial amount of unexplained variance which neither of the two factors can account for. This is further confirmed by the communality (h^2) values which oscillate around 0.500 for each variable with the exception of *Speaking*. This suggests that the two factors are not independent of each other (orthogonal) but correlated (oblique). And indeed, the correlation between the two factors is 0.837. *Table 3.13.* reports the results of Schmid and Leiman's hierarchical FA applied to the matrix in *Table 3.11.*

KET - A	**2nd order**	**1st order**	**1st order**
	Factor 1	**Factor 1**	**Factor 2**
ListComp	**0.646**	-0.042	**0.327**
ReadComp	**0.784**	0.086	**0.260**
Grammar	**0.757**	**0.411**	-0.076
Vocab.	**0.724**	**0.340**	-0.020
Writing	**0.748**	0.032	**0.298**
Speaking	**0.586**	0.097	**0.162**

Significant loadings in bold

Table 3.13. Hierarchical factor analysis of KET-A's 6 variables

There appears a new, higher-order factor which is loaded by all six variables. Two primary (1st order) factors are significantly loaded by the same variables as in *Table 3.12.* thus maintaining the factorial structure of the SLP variables. The second-order factor (to be interpreted as a General SLP factor) is almost perfectly correlated with each of the two primary factors: r = 0.915. The correlation for each factor is the same because the amount of explained variance by *SLPComp* and *SLPSkill* is equal. The discussion of the results is postponed till *Chapter 4.*

3.6.2. TPJ's construct validation.

As it was said in *sections 2.4.1.* and *2.4.7.2.* TPJ's construct validity needed to be confirmed in the light of the analysis of the test's tasks which revealed problems with its content validity. Factor analysis as described in *section 3.5.* can let the test 'speak for itself' by analysing intercorrelations between its components. If TPJ is a valid measure of the abilities it declares to measure then we would expect the test's variables (tasks) to overlap (at least partially) with 5 or

4-factor solution. The analysis starts with the correlation matrix in *Table 3.14.* on which principal axis extraction method was run.

N = 226							
	TPJ 1	**TPJ 2**	**TPJ 3**	**TPJ 4**	**TPJ 5**	**TPJ 6**	**TPJ 7**
TPJ 2	0.22						
TPJ 3	0.14	0.36					
TPJ 4	0.15	0.45	0.37				
TPJ 5	0.18	0.38	0.35	0.41			
TPJ 6	0.27	0.42	0.28	0.47	0.35		
TPJ 7	0.11	0.40	0.43	0.48	0.49	0.33	
TPJ 8	0.16	0.38	0.30	0.46	0.42	0.37	0.58

significant r in bold; p = 0.01

Table 3.14. Correlation matrix for TPJ's 8 tasks

The matrix looks promising as at least three degrees of magnitude of ***r*** can be distinguished suggesting a potential pattern of underlying factors. There is one high ***r*** (0.58), a cluster of moderate correlations in the range of 0.30 to 0.49 and four low-magnitude correlations (0.18 to 0.28).

The results of eigenvalue estimation are shown in *Table 3.15.* which together with the scree test in *Figure 3.3.* offer suggestions as to how many factors are to be extracted using the PF method of extraction. Again, as in our example, Kaiser's 'greater-than-unity' rule is of no help (cf. *section 3.5.*) and the factor solution is followed which is suggested by Cattell's 'scree test'. This points again toward a 2-factor solution.

However, before this solutions is presented, let us see how TPJ looks when a 5-factor solution is attempted. As it will be recalled from earlier in this section it has been suggested that if TPJ addressed stable construct then its posited abilities would coincide with five factors (cf. *section 2.4.1.* where Kuliniak's posited abilities are discussed) extracted by factor analysis, as was the case in our illustrative example. In other words we would expect one factor to coincide with Group 1 ability i.e., *sensitivity to register*

N=226	Eigenvalue	% of total variance	cumulative eigenvalue	cumulative %
1	3,10	38,80	3,104	38,80
2	0,39	4,90	3,496	43,70
3	0,19	2,32	3,682	46,02
4	0,13	1,58	3,808	47,60
5	0,02	0,26	3,829	47,86
6	0,01	0,06	3,833	47,92
7	0,01	0,02	3,835	47,94

Table 3.15. Eigenvalues for TPJ's sub-tests

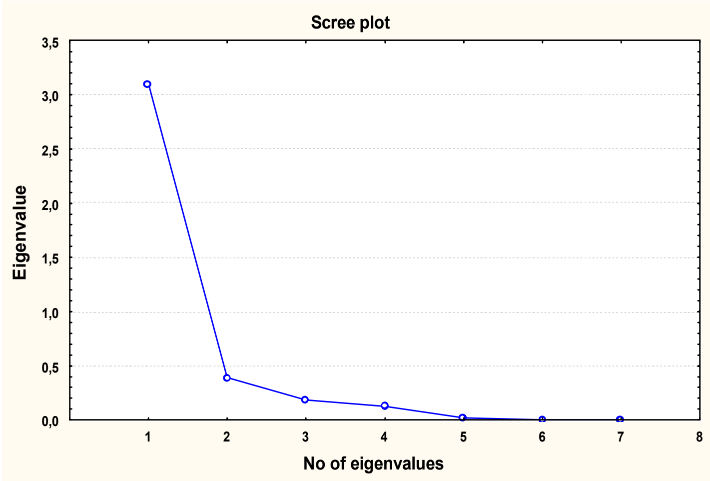


Figure 3.3. Scree test for number of components to be extracted for TPJ

and to grammaticality, another one with Group 2 ability *richness of lexis and collocations*, still another one with Group 3 *recognition of grammatical form in L1* and so on. This, however, is not the case as can be seen in *Table 3.16.* below which gives results for a five-factor rotated solution for TPJ. Only two abilities postulated for TPJ 'coincide' with two separate factors (cf. the discussion below).

N = 226	loadings >0.400 in bold					
	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**	***h^2***
TPJ 1	0.088	0.250	0.008	0.402	0.018	*0.104*
TPJ 2	0.339	**0.498**	0.160	0.147	0.091	*0.323*
TPJ 3	0.369	0.315	0.361	0.024	0.004	*0.251*
TPJ 4	**0.439**	**0.606**	0.079	0.010	0.045	*0.419*
TPJ 5	**0.580**	0.219	0.189	0.210	0.110	*0.366*
TPJ 6	0.283	**0.574**	0.053	0.261	-0.059	*0.330*
TPJ 7	**0.799**	0.210	0.149	0.010	-0.022	*0.491*
TPJ 8	**0.664**	0.309	-0.149	0.131	0.034	*0.419*
expl. variance	24.5%	16.2%	3.1%	3.9%	0.3%	

Table 3.16. Five-factor rotated solution for TPJ

Among the tasks making the *richness of lexis and collocations* factor only TPJ 2 and TPJ 4 significantly load the same factor. TPJ 3, purportedly measuring the same ability, insignificantly loads three different factors. Similarly, the loadings of TPJ 5 and 6 (*recognition of grammatical form in L1*) are spread over two factors. One factor (Factor 5 in *Table 3.16* above) reveals only marginal loadings by all the variables. What this means is that, apparently, an entirely different picture of ability structure emerges as a result of factor analysis, at odds with the pattern intended by the test's author.

Continuing the search for the structure of the abilities underlying TPJ, two factors have been extracted and rotated to yield a two-factor solution. The results, in the form of the varimax rotated factor loadings are shown in *Table 3.17.* Again, as in *section 3.5.* a three-factor solution is included for the sake of comparative illustration.

N = 226	loadings >0.400 in bold					
	3 factor solution			2 factor solution		
	Factor 1	**Factor 2**	**Factor 3**	**Factor 1**	**Factor 2**	***h^2***
TPJ 1	0.101	0.383	-0.066	0.082	0.364	*0.104*
TPJ 2	0.362	**0.501**	0.184	0.367	**0.523**	*0.323*
TPJ 3	0.383	0.277	0.374	**0.413**	0.314	*0.251*
TPJ 4	**0.473**	**0.505**	0.168	**0.469**	**0.537**	*0.419*
TPJ 5	**0.571**	0.287	0.121	**0.569**	0.316	*0.366*
TPJ 6	0.301	**0.626**	0.053	0.281	**0.638**	*0.330*
TPJ 7	**0.814**	0.158	0.159	**0.857**	0.172	*0.491*
TPJ 8	**0.672**	0.313	-0.124	**0.592**	0.335	*0.419*
expl. variance	25.5%	16.6%	3.3%	25.3%	18.0%	

Table 3.17. Varimax rotated three and two-factor solution for TPJ

It is immediately to be seen that with respect to the simplicity of achieved structure neither solution could compare with the neatness of the example from *section 3.5.* But even if that was a 'tailored' case for the sake of the illustration, we would expect, in the example under consideration, a more pronounced pattern of factors underlying TPJ's construct. This however, is not the case. Because of the violation of the Cattell's factor extraction principle referred to above, the 3-factor solution is to be rejected as it does not lead to simple structure as defined in *section 3.5.* Observe that with one 'idle' factor in the solution the requirement of a simple structure obtained after rotation is not met. Apparently, the three-factor solution from *Table 3.17.* is not in any way simpler than, for that matter, an unrotated solution.

The other reason is more compelling and is the corollary of the 'simple-structure-condition'. As it will be recalled from the discussion of the illustrative example the aim of factor analysis is to discover a pattern of the interdependencies between the multiplicity of variables. The ultimate decision about the pattern underlying the variables is left to a researcher's interpretation rather than imposed by the 'structure-oriented' technique. Factor analysis, as any statistical procedure, is only a tool and is bound to produce a pattern out of practically any dataset. It will expose a clear structure where no structure can feasibly be expected[9]. A random selection of variables as uncorrelataed with one another as the make of a car one drives, hair colour and L2 proficiency measured by standard tests will yield a structure when factor analysed[10]. Thus the interpretation of the results of any factor analytic study must, on the one hand, take account of what variables were analysed and on the other, be compatible with the current state of expertise in the field.

On the count of the two aspects of what, in principle, is one criterion neither solution offers a reasonably interpretable ability structure of FL aptitude as measured by TPJ. While structure imposing properties of FA suggest 2-factor solution for TPJ, it does not produce easily identifiable ability structure. On the contrary, a two-factor varimax rotated solution runs counter the delineations declared by the TPJ's author. An examination of *Table 3.17.* reveals that 6 variables (TPJ 3, 4, 5, 7 and 8) in the two-factor solution are correlated with/load Factor 1. One of those, TPJ 3, barely approximates the level of statistical significance i.e., 0.413; p=0.420 for N=226 (cf. *note 6* in this chapter). The variables that pattern together are Kuliniak's *richness of lexis and collocations* [TPJ 4], *recognition of grammatical form in L1* [TPJ 5] and (rephrased in Carroll and Sapon's terms) *inductive language learning ability* [TPJ 7 and 8] (cf. *section 3.4*). Factor 2 is loaded by *recognition of grammatical form in L1* [TPJ 6] and by *richness of lexis and collocations* [TPJ 2 and 4]. The last variable has also significant loading on Factor 1. The *sensitivity to register and to grammaticality* ability [TPJ 1] does not load any of the two factors.

The picture does not look any way better for TPJ even when Schmid and Leiman procedure for correlated (oblique) factors is followed. Despite the fact that the two factors extracted in *Table 3.17.* are highly correlated (***r***=0,822) a hierarchical factor analysis does not yield second and first-order factors which would be any easier to interpret than the two-factor solution. Thus, it may be concluded that TPJ looks disappointingly week as far as its postulated construct is concerned.

Concluding the validation of TPJ's construct the following two things can be said. The only ability dimension that withstood the validation procedure is *inductive learning* [TPJ 7 and 8]. The variables TPJ 7 and TPJ 8 making up the ability were found to pattern together consistently enough in several solutions tried to be considered a separate ability dimension – *inductive language learning ability* (Carroll and Sapon 1959, Carroll 1979, Carroll 1999). Consequently, they were included as the only reliable TPJ's variables into further analysis with PLAT's variables (cf. *Table 3.4. – Group 4 and 5* reliability index).

3.6.3. PLAT's construct validation.

The analysis of PLAT's construct validity has proceeded along the same lines as for TPJ. However, for the lack of space only steps 1 and 5 (optionally 4) will be discussed. *Table 3.18.* contains intercorrelations between PLAT's variables.

N = 153	ofiara	handl	układ	przed	podcz	lubić	pamię	dykta	wingd	spraw	pojez	dbrds
ofiara												
handlarz	**0.50**											
układ	**0.60**	**0.30**										
przed	**0.27**	**0.19**	**0.17**									
podczas	**0.34**	**0.20**	**0.29**	**0.27**								
lubić	**0.35**	0.06	**0.29**	0.11	**0.16**							
pamiętać	**0.38**	0.14	**0.27**	**0.18**	0.16	**0.52**						
dyktando	**0.35**	**0.34**	**0.40**	**0.26**	**0.25**	**0.19**	**0.27**					
wingdings	**0.22**	**0.21**	**0.37**	0.08	0.12	**0.16**	**0.24**	**0.25**				
sprawcy	**0.41**	**0.29**	**0.32**	0.07	**0.29**	**0.16**	**0.28**	**0.45**	**0.25**			
pojezierze	**0.49**	**0.40**	**0.54**	**0.20**	**0.29**	**0.31**	**0.34**	**0.48**	**0.35**	**0.45**		
dbrdszn	**0.39**	**0.30**	**0.46**	**0.23**	**0.29**	**0.35**	**0.41**	**0.55**	**0.37**	**0.43**	**0.64**	
zwolennik	**0.44**	**0.29**	**0.42**	**0.19**	**0.22**	**0.26**	**0.36**	**0.50**	**0.32**	**0.44**	**0.64**	**0.69**

significant correlations in bold; p <0.05

Table 3.18. Correlation matrix for PLAT's 13 variables

All the correlations are positive and those in bold are significant at p<0.05. An examination of the matrix suggests that FA is likely to disclose an underlying structure of intercorrelations between the variables. When compared to *Table 3.14.* the matrix for PLAT shows much greater scope in the magnitude of the correlations. While for TPJ it ranges from 0.11 to 0.50 in the case of PLAT the range is from 0.07 (*sprawcy* x *przed*) to 0.69 (*zwolennik* x *dbrdszn*).

The examination of the scree plot (cf. *Appendix 8*) suggested a three-factor solution. The communalities and residual correlations were examined and found acceptable. *Table 3.19.* below presents the results of factor rotation following factor extraction using the principal axis extraction. The three-factor solution was then interpreted. The amount of variance explained by the three factors totals 45.7%. Each factor's contribution to the amount explained is given in the bottom row of *Table 3.19.*

Extr. PA; N = 153; loadings > 0.450 in bold				
	Factor 1	**Factor 2**	**Factor 3**	***h²***
ofiara	0.245	0.206	**0.868**	*0.574*
handlarz	0.304	-0.112	**0.493**	*0.334*
układ	0.433	0.162	**0.486**	*0.494*
przed	0.184	0.061	0.260	*0.164*
podczas	0.250	0.053	0.337	*0.194*
lubić	0.178	**0.721**	0.188	*0.338*
pamiętać	0.311	**0.579**	0.199	*0.367*
dyktando	**0.615**	0.024	0.266	*0.421*
wingdings	0.394	0.106	0.159	*0.217*
sprawcy	**0.500**	0.038	0.299	*0.356*
pojezierze	**0.691**	0.139	0.359	*0.573*
dbrdszn	**0.809**	0.231	0.168	*0.617*
zwolennik	**0.737**	0.147	0.217	*0.572*
expl.variance	23.2%	8.1%	14.4%	

Table 3.19. Varimax rotated 3-factor solution for PLAT's 13 variables

Generally, the pattern which emerges coincides with postulated ability domains discussed in *section 2.4.2.* and summarised in *Table 2.2.* Starting from the top of the table, the first three variables (*ofiara, handlarz, układ*) comprising reading comprehension tasks of the PLAT test all load significantly on Factor 3. The communalities for the variables are moderately satisfactory (average h^2=0.467). The next two variables, *lubić* and *pamiętać*, comprising rote-memory-for-words tasks load Factor 2 (average h^2= 0.353). The last five variables, with the exception of *wingdings*, all exhibit significant loadings on Factor 1 (average h^2= 0.670).

The loadings for three variables: *przed*, *podczas* and *windings* do not reach statistical significance and the variables' communalities are low. This suggests that the variables' variance is not well accounted for by the factor solution. Still, the variables associated with the listening comprehension tasks i.e., *przed* and *podczas* contribute the little common variance they account for to Factor 3 by loading it at 0.260 and 0.337 respectively. In other words, the variables have 26% and 33.7% of common variance respectively with all the remaining PLAT variables. Also the variable *wingding* minimally 'misses' on Factor 1 with a loading of 0.394. With respect to some of the variables loading other than their 'principal' factor it is worth noticing that the 'spatial memory for verbal material' variable *układ* (cf. *Table 2.2. section 2.4.*) almost reaches significance on Factor 1 and could be said to contribute almost equally to Factor 1 and 3. Another interesting pattern obtains in the case of Factor 3 which is loaded from two 'contextual lexical production' variables *sprawcy* and *pojezierze*. Also, two reading comprehension variables i.e., *ofiara* and *hadlarz*, display some loading on Factor 1.

As some of the factors turn out to be correlated Schmid and Leiman hierarchical analysis is applied to the original matrix in *Table 3.18.* The factors from *Table 3.19.* reveal a high correlation of 0.787 between Factor 1 and 3; and moderate correlations between Factors 1 and 2 (0.559) and 2 and 3 (0.516). This suggests that a second-order, more general factor exists apart from more specific domains of FL aptitude. The results of hierarchical analysis in *Table 3.20.* below reveal a general FL aptitude factor as shown by factor loadings in the first column.

N = 156	significant loadings in bold			
	2nd order Factor 1	**1st order Factor 1**	**1st order Factor 2**	**1st order Factor 3**
ofiara	**0.747**	-0.097	0.018	**0.536**
handlarz	**0.443**	0.088	**0.252**	**0.283**
układ	**0.620**	0.146	0.025	**0.207**
przed	0.293	0.048	0.027	0.128
podczas	0.377	0.073	0.062	0.166
lubić	**0.488**	-0.017	**-0.590**	-0.002
pamiętać	**0.526**	0.090	**-0.433**	-0.016
dyktando	**0.552**	**0.351**	0.147	0.010
wingdings	0.380	**0.217**	0.010	-0.013

sprawcy	**0.506**	**0.260**	0.118	0.066
pojezierze	**0.697**	**0.364**	0.074	0.041
dbrdszn	**0.691**	**0.488**	-0.022	-0.143
zwolennik	**0.644**	**0.435**	0.050	-0.076

Table 3.20. Hierarchical factor analysis of PLAT's 13 variables

This general factor which is significantly loaded by all but three variables: *przed*, *podczas* (aural-text-memory factor) and *windings* (symbol-manipulation factor), underlies all the remaining specific three factors which exhibit the same pattern of loadings as in *Table 3.19.*

A more detailed discussion of the interpretation of the factors will be postponed till the next chapter. At this point it will only be remarked that the three-factor solution confirms the assumptions of independent sub-scales that lay behind reliability estimates discussed in *section 3.3.* Consequently, the construct that PLAT purports to measure emerges as more viable and interpretable than that of TPJ.

3.7. Results of factor analysis of FL aptitude

After the application of FA in construct validation procedure described in *section 4.6* above exploratory factor analysis was used to uncover the structure of abilities underlying the combined scale formed of PLAT and the two TPJ sub-scales found to be the most reliable and, at the same time, displaying the most stable construct. As in the previous application of FA, only the results of steps 1, 5 and optionally 4 (cf. *section 3.5.*) will be discussed. The matrix of correlations in *Table 3.21.* repeats the values for PLAT[11] from *Table 3.18.* and additionally contains correlations for the two TPJ variables (the first two columns).

N = 146 significant correlations in bold; p <0.05

	TPJ 7	**TPJ 8**	**ofiara**	**handl**	**układ**	**przed**	**podcz**	**lubić**	**pamię**	**dykta**	**wingd**	**spraw**	**pojez**	**dbrds**
TPJ 8	**0.56**													
ofiara	**0.35**	**0.37**												
handlarz	**0.31**	**0.42**	**0.50**											
układ	**0.34**	**0.36**	**0.58**	**0.30**										
przed	**0.28**	**0.29**	**0.25**	**0.16**	0.15									
podczas	**0.17**	0.16	**0.33**	**0.21**	**0.28**	**0.26**								
lubić	**0.21**	**0.19**	**0.32**	0.04	**0.28**	0.09	0.16							
pamiętać	**0.34**	**0.24**	**0.36**	0.11	**0.27**	0.15	0.14	**0.52**						
dyktando	**0.31**	0.16	**0.31**	**0.32**	**0.39**	**0.23**	**0.24**	0.16	**0.25**					
wingdings	**0.25**	**0.20**	**0.18**	**0.20**	**0.35**	0.07	0.13	0.12	**0.23**	**0.24**				
sprawcy	**0.33**	**0.20**	**0.40**	**0.29**	**0.32**	0.07	**0.29**	0.16	**0.29**	**0.44**	**0.25**			
pojezierze	**0.44**	**0.40**	**0.49**	**0.39**	**0.54**	**0.18**	**0.29**	**0.32**	**0.33**	**0.48**	**0.35**	**0.46**		
dbrdszn	**0.52**	**0.36**	**0.37**	**0.29**	**0.45**	**0.22**	**0.29**	**0.33**	**0.40**	**0.54**	**0.34**	**0.43**	**0.64**	
zwolennik	**0.46**	**0.35**	**0.43**	**0.28**	**0.41**	**0.18**	**0.23**	**0.25**	**0.34**	**0.50**	**0.28**	**0.46**	**0.63**	**0.69**

Table 3.21. Correlation matrix for 15 FL aptitude variables (PLAT + TPJ 7 and 8)

It is interesting to observe that the magnitude of the correlation between the two TPJ variables is higher than with any of the PLAT's variable (with the exception of relatively moderate-to-high correlation with *dbrdszn*). The pattern of the intercorrelations among the PLAT's variables is the same as described in *section 3.6.* It may thus be expected that principal axis factoring followed by the varimax rotation will yield an additional factor which would capture the inclusion of the two TPJ's variables into the FL aptitude battery. And indeed, this is the case. *Table 3.22.* shows factor loadings for the four-factor solution.

N = 146	Extrct.: PA loading >0.450 in bold				
	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	h^2
TPJ 7	0.419	**0.569**	0.147	0.126	*0.455*
TPJ 8	0.177	**0.794**	0.077	0.222	*0.441*
ofiara	0.222	0.198	0.205	**0.818**	*0.565*
handlarz	0.244	0.324	-0.157	**0.480**	*0.400*
układ	0.401	0.177	0.159	**0.481**	*0.489*
przed	0.148	0.274	0.048	0.172	*0.199*
podczas	0.250	0.065	0.057	0.325	*0.199*
lubić	0.162	0.064	**0.709**	0.173	*0.340*
pamiętać	0.288	0.148	**0.603**	0.156	*0.386*
dyktando	**0.641**	0.023	0.004	0.241	*0.433*
wingdings	0.356	0.128	0.080	0.137	*0.208*
sprawcy	**0.523**	0.040	0.055	0.314	*0.365*
pojezierze	**0.651**	0.225	0.150	0.354	*0.583*
dbrdszn	**0.789**	0.240	0.222	0.115	*0.631*
zwolennik	**0.719**	0.206	0.136	0.189	*0.578*
expl.variance	20.3%	9.4%	7.3%	11.6%	

Table 3.22. Varimax rotated 4-factor solution for PLAT's 13 variables

The examination of reproduced and residual correlations revealed only two residuals > 0.100 and thus it is assumed that the solution represents well the original matrix. The pattern of factor loadings shown in *Table 3.22.* is the same as in the case of PLAT in *Table 3.19.* The small changes in the magnitude of the loadings are due to TPJ 7 and 8 having been included in the correlation matrix. As expected, an additional factor appeared (Factor 2) which receives strong loadings from the two variables included. It accounts for 9.4% of the total variance, higher than factor 3 which explains 7.3% of the total variance. All in all, the four factors explain 48.6% of the total variance. It is to be reminded at this point that the 'percentage of explained variance' index in the PF method is always smaller than in the case of the PC method due to the fact that the former, by subtracting error and specific variance from the total variance, accounts for common variance only, unlike the PC method which explains the total variance without splitting it into common and unique variance (cf. Rietvald and Hout 1993; Zakrzewska 1994). The inspection of the non-significant factor loadings from *Table 3.22.* reveals basically the same pattern as in *Table 3.19.* and described in *section 3.6.* Among the new variables added, only TPJ 7 shows quite a substantial (but still not significant) loading on Factor 1 (0.419).

The considerations of factor loadings, both significant and non-significant, may lead to some interesting observations regarding the 'coverage' of factors by variables and what may be called a 'robustness' of factors. Both notions link with the concept of simple structure which, as it will be recalled, results from factor rotation and produces, ideally, a structure where a variable or variables strongly load(s) one factor only while negligibly/minimally affecting other factors. In such a case a factor may be said to be distinctly, if not exclusively, identified by the construct measured /characterised by the variables. In other words, the factor shows 'robustness' or stability in capturing the quality and/or feature attributed to the variables and only this quality and feature. The higher the values of the correlations between the factors and the original variables (factor loadings) the more 'robust' the factors are. Now, if all variables used in the analysis have significant factor loadings on all extracted and rotated factors, the factors are said to have good 'coverage' by the variables (Rietvald and Hout 1993; Ferguson and Takane 1999). Good factor coverage by the variables lends additional support to a factor solution being considered. The presence of 'idle' variables i.e., variables not contributing to any factor, while not necessarily an outright disaster for the factor model, needs to be explained and interpreted and calls for careful re-examination of the posited construct or its operationalisation.

If examined with the discussion from the previous paragraph in mind, *Table 3.22.* reveals the following. Out of the four factors in the solution, Factor 3 exhibits the highest degree of 'robustness'. Only the variables *lubić* and *pamiętać* have high loadings on the factor: 0.603 and 0.709 respectively, and all non-significant loadings from other variables are very low and range from 0.004 to 0.222 (average: 0.115). This means that the factor represents well the postulated ability dimension (here: *paired-associate memory*) and it can be safely assumed that this construct accounts for the explained variance attributed to the factor. The second best is Factor 2 strongly loaded by two TPJ variables. The factor's 'unwanted' loadings from thirteen PLAT's variables are again very low and range from 0.023 to 0.324 (average: 0.162). For Factors 1 and 4 the relevant figures are: 0.148 to 0.419 (average: 0.266) and 0.115 to 0.354 (average: 0.210) respectively. For the purpose of the discussion to follow, the factors will be labelled: F1=*General Verbal* (GenVrbl); F2=*Linguistic Inductive Reasoning* (IndReas); F3=*Paired-Associate Memory* (AssMem) and F4=*Written Text Comprehension* (TextCom). The discussion of the results and the interpretation of the extracted factors is postponed till *section 4.2.*

3.8. Results of factor analysis of L2 proficiency, FL aptitude and intelligence

In order to address RQ 1 (cf. *section 0.2.1*) a correlation matrix was prepared for all the relevant variables in *Table 3.23.* Read horizontally, the first four are FL aptitude variables (factor analysed earlier), the next three are intelligence variables (verbal and non-verbal) and the last six are L2 proficiency variables. With the exception of most

N = 136	significant correlations in bold; $p < 0.01$											
	Gen Vrbl	**Induc Reas**	**Ass Mem**	**Txt Com**	**TJL PVoc**	**TJL AVoc**	**SRP M**	**SLP LC**	**SLP RC**	**SLP GR**	**SLP VC**	**SLP WR**
IndReas	0.12											
AssMem	0.12	-0.03										
TextComp	0.07	0.09	0.03									
TJL PVoc	**0.55**	**0.25**	0.08	**0.55**								
TJL AVoc	**0.65**	**0.22**	0.17	**0.45**	**0.80**							
SRPM	**0.44**	**0.39**	0.17	0.21	**0.37**	**0.43**						
SLP LC	**0.37**	**0.36**	0.07	0.17	**0.40**	**0.44**	**0.29**					
SLP RC	**0.42**	**0.33**	0.10	**0.31**	**0.50**	**0.49**	**0.35**	**0.52**				
SLP GR	**0.38**	**0.27**	0.19	**0.28**	**0.50**	**0.51**	**0.29**	**0.46**	**0.63**			
SLP VC	**0.36**	**0.25**	0.13	**0.24**	**0.39**	**0.48**	**0.29**	**0.46**	**0.67**	**0.69**		
SLP WR	**0.50**	**0.35**	0.02	**0.30**	**0.54**	**0.53**	**0.30**	**0.54**	**0.66**	**0.54**	**0.55**	
SLP SP	**0.51**	**0.28**	**0.26**	**0.23**	**0.47**	**0.54**	**0.40**	**0.33**	**0.39**	**0.53**	**0.37**	**0.53**

Table 3.23. Correlation matrix for FL aptitude, intelligence and L2 proficiency

correlations for FL aptitude variable *AssocMem* (third column) and two correlations for *TextComp*, all the remaining ones are significant at $p<0.01$[12]. The examination of the magnitude range of the correlations suggests a wealth of possible patternings to be extracted by factor analysis. The lowest positive correlation is 0.07 for *AssMem* x *SLP LC* (insignificant) and the highest is 0.80 for *TJL AVoc* x *TJL PVoc* (significant).

Factor analysis with varimax rotation was run on the matrix above. After the examination of the values of latent roots (eigenvalues) and a scree-plot it was the four-factor solution which was decided on. The two principles commonly used as guidelines in deciding on the number of factors to be extracted (Kaiser's and Cattell's principles in *section 3.5.*) gave in effect conflicting results. The greater-than-unity principle suggested four factors while Cattell's scree test suggested five (the scree elbowed at eigenvalue no. 6). Both solutions were tried and the four-factor solution was chosen as lending itself best to linguistic interpretation. *Table 3.24.* below presents the results of the analysis.

N = 136	Extraction: PC significant loadings>0.400 in bold				
	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	***h^2***
GenVrbl	**0.482**	0.229	0.358	**0.431**	*0.604*

IndReas	-0.027	0.296	-0.257	**0.743**	*0.314*
AssMem	-0.012	0.098	**0.857**	-0.047	*0.139*
TextComp	**0.776**	0.155	-0.193	-0.177	*0.431*
TJL PVoc	**0.834**	0.299	0.042	0.204	*0.727*
TJL AVoc	**0.770**	0.321	0.216	0.263	*0.745*
SRPM	0.312	0.076	0.227	**0.707**	*0.365*
SLP LC	0.148	**0.628**	-0.028	0.363	*0.397*
SLP RC	0.280	**0.789**	0.011	0.189	*0.615*
SLP GR	0.270	**0.783**	0.211	0.074	*0.601*
SLP VC	0.178	**0.839**	0.120	0.060	*0.598*
SLP WR	0.378	**0.663**	-0.034	0.300	*0.606*
SLP SP	**0.403**	0.318	**0.414**	**0.403**	*0.483*
expl.variance	21%	25%	10%	13.9%	

Table 3.24. Varimax rotated solution for FL aptitude, intelligence and SLP

The interpretation of the factor solution will start from the factor which accounts for the largest percentage (25.0%) of the total variance (F.2). This factor is almost exclusively loaded by five, out of six, SLP variables all displaying very high loadings on the factor. It also receives substantial but non-significant loadings from *SLP Speaking* and *TJL AVoc* variables. Reading down the second best factor column (21% of explained variance), we see that three variables *L2 TxtCom.*, *TJL PVoc* and *TJL AVoc* load highest on this factor and that two other factors *L2 GenVrbl* and *SLP SP* contribute to it significantly. F.4 which is markedly loaded by *SRPM* (non-verbal inductive reasoning) and *L2 IndReas* (linguistic inductive reasoning) variables (0.707 and 0.743 respectively) is also significantly loaded by *SLP Speaking* and *L2 GenVrbl.* It accounts for almost 14 percent of the total variance. The last factor (F.3) is loaded very highly by *L2 Paired Associate Memory* variable and significantly, but less pronouncedly, by *SLP Speaking* variable. Its factorial structure is difficult to interpret not only because of an 'under-representation' of other variables, but also because of its very low communality value ($\boldsymbol{h^2}$=0.139) which indicates that the variable's unique variance is not well accounted for by the extracted factors. The communalities of all the remaining variables are well above the critical values.

The correlations between the first-order factors were then examined for obliqueness. It turned out that Factors 1, 2 and 4 were correlated to the degree which would suggest a second-order level of relationships missed by the first-order solution[13]. Consequently, a hierarchical factor analysis (cf. *section 3.4.*) was run on the second-order correlation factor-matrix. *Table 3.25.* below contains factor loadings for the hierarchical factor solution.

	Extraction: PC; loadings > 0.300 in bold					
	2nd order Factor 1	**2nd order Factor 2**	**1st order Factor 1**	**1st order Factor 2**	**1st order Factor 3**	**1st order Factor 4**
GenVrbl	**0.574**	0.197	0.251	-0.008	**0.314**	0.260
IndReas	**0.520**	-0.239	-0.167	0.141	-0.208	**0.538**
AssMem	0.014	**0.342**	-0.065	0.052	**0.789**	-0.011
TextComp	**0.384**	0.233	**0.592**	-0.030	-0.252	-0.279
TJL PVoc	**0.677**	0.262	**0.548**	0.006	-0.022	0.006
TJL AVoc	**0.685**	0.295	**0.479**	0.023	0.147	0.066
SRPM	**0.547**	-0.010	0.130	-0.117	0.229	**0.521**
SLP LC	**0.592**	0.076	-0.076	**0.395**	-0.049	0.170
SLP RC	**0.656**	0.213	0.007	**0.508**	-0.042	-0.010
SLP GR	**0.589**	**0.313**	0.005	**0.514**	0.138	-0.090
SLP VC	**0.567**	0.272	-0.074	**0.583**	0.055	-0.103
SLP WR	**0.694**	0.164	0.101	**0.377**	-0.076	0.083
SLP SP	**0.568**	0.224	0.171	0.079	**0.364**	0.236

Table 3.25. Hierarchical factor analysis of FL aptitude, intelligence and SLP

Two second-order factors and four first-order (specific) factors emerged from the analysis. A 2nd order Factor 1 is heavily loaded by virtually all the variables (with the exception of FL aptitude *paired-associate memory*) suggesting that there is a General Cognitive Factor which underlies all the specific abilities. The other 2nd order Factor 2, while being decisively less pronounced than the first one, exhibits quite clear factorial structure. The factor is significantly loaded by FL aptitude memory variable and SLP grammar test. Three other variables (all involving knowledge of vocabulary) are just below statistical significance with their loadings on the factor. The analysis of the specific ability factors (1st order factors) shows generally the same pattern and factorial structure as in *Table 3.24.* The discussion of the results and their implications will be presented in *section 4.3.*

3.9. Results of multiple regression analyses

As a statistical technique of data analysis multiple regression (MR) relies basically on the same mathematical procedure described in connection with factor analysis in *section 3.5.* It consists in arriving at the best fitting line representing two sets of data or scores on an X / Y co-ordinates plot. But unlike factor analysis which interprets the regression line as a new reduced dimension, regression analysis is interested in estimating the values for variable X on the basis of known values of variable Y. In other words, on the basis of the results obtained from the sample, researchers hope to predict patterns of performance and/or behaviour for the population, with specified degree of certainty (Pedhazur 1982). In short, regression analysis allows researchers to generalise their findings from the sample to the whole population[14].

Despite an obvious limitation of the technique such as the impossibility of attributing causation to predictor variables (a limitation which it shares with other techniques) multiple regression analysis is a powerful analytic tool which can eliminate quite a substantial amount of guesswork in the process of interpolation from the sample to the population; from the observable to the unobservable. Because MR analyses the relationships between predictor variables on the one hand, and a criterion variable on the other, it can give a truer picture of the position of each variable in predicting the criterion variable. Specifically, in accounting for the variance of the dependent variable the individual contributions of the predictor variables are estimated in relation to each other and thus "the role of a predictor variable can be assessed more precisely by ruling out that part of the variation in the criterion variable which can be assigned to other predictor variables" (Rietvald and Hout 1993: 96).

In practical terms it is possible, using fewer variables, to make equally accurate predictions of the criterion as when a larger set of predictors is used. Thus the technique of multiple regression allows a researcher to identify a subset of predictors within a larger set whose R^2 value is not significantly lower from the R^2 of the whole set thus rendering the prediction task more economical for test administrators and less strenuous for the learners. In the present study the stepwise method of multiple regression[15] is used to identify the most efficient variables.

3.9.1. Correlation matrix for MR.

Table 3.26. below contains an initial correlation matrix for six independent variables used in the present study on which multiple regression was run[16].

N = 137	significant r in bold; p<0.01									
	Gen Vrbl	**Ind Reas**	**Ass Mem**	**Txt Comp**	**SRPM**	**AVoc**	*TJL Total*	*L2 Aptit*	**KET-A**	**SLP Compet**
IndReas	0.12									
AssoMem	0.12	-0.03								
TextComp	0.07	0.09	0.03							
SRPM	**0.44**	**0.39**	0.17	0.21						
AVoc	**0.65**	**0.23**	0.17	**0.44**	**0.42**					
TJL Totl	***0.63***	***0.25***	*0.13*	***0.52***	***0.42***	***0.95***				
L2 Aptitude	***0.83***	***0.39***	***0.34***	***0.47***	***0.55***	***0.76***	***0.78***			
KET-A	**0.48**	**0.39**	0.13	**0.34**	**0.38**	**0.58**	**0.61**	**0.63**		
Compet	**0.37**	**0.24**	0.17	**0.27**	**0.29**	**0.50**	**0.50**	**0.49**	**0.87**	
Skills	**0.46**	**0.40**	0.03	**0.28**	**0.34**	**0.48**	**0.55**	**0.61**	**0.79**	**0.40**

Table 3.26. Correlation matrix of six predictor variables with the criterion variable(s)

The examination of the correlations between the predictors and the criterion on the one hand, and among the predictors themselves on the other, allows one to make the following observations.

With the exception of *paired-associate memory* all the correlations between the predictors and the criterion are positive and statistically significant at a high probability level of 0.01. In order to assess the magnitude of the correlations the factors known to influence it, such as sample size (N=137) and the range of scores (cf. *Table 3.2.*) were taken into account. The scatterplots of the predictors against the criterion were also examined for possible 'outliers', i.e., isolated results lying far outside the main 'regression area' and having a potential deflating effect on the magnitude of the correlation. Few were found (cf. *Appendix 9*) and temporarily removed from the plot to see if the resultant Pearson's product moment ***r*** changed drastically. As it did not, the cases was retained and dealt with in the multiple regression analysis, together with other outstanding cases[17].

Researchers differ in their interpretations of the magnitudes of correlation coefficients but according to most authors[18] the values of the coefficients for the variables *IndReas.:* 0.39, *TxtComp:* 0.34 and *SRPM:* 0.38 can be interpreted as average while those for *GenVrbl:* 0.48 and *AVoc:* 0.58 as high. Stopping at this simple correlational level of analysis a researcher might be tempted to make claims about the relative importance of the independent variables in predicting the scores of the dependant one[19]. Accordingly, basing his decisions on the matrix in *Table 3.26.* he would most likely say that this influence follows the ordering of the coefficients from the strongest to the weakest i.e., 1. *AVoc*, 2. *GenVrbl*, 3. *IndReas*, 4. *SRPM*, 5. *TxtComp.* This, however, produces a false picture of individual contributions of each of the variables to the prediction of the criterion variable as will be seen in *sections 3.9.2.* and *3.9.3.* below.

There are three points worth noticing about the correlation matrix in *Table 3.26.* before the results of multiple regression are presented. Firstly, the correlations of the *paired-associate* predictor do not achieve statistical significance with any of the remaining predictors. While the lack of statistical significance for intercorrelations within the FL aptitude variables is understood[20] the fact that the variable in question does not correlate significantly with SRPM and AVoc is telling. The next observation to be made is connected with the assumption of non-collinearity discussed in *section 3.2.6.* of this chapter. To avoid the violation of this assumption, which would yield an uninterpretable regression analysis output due to a high intercorrelation between *PVoc* and *AVoc*, only the latter was chosen. The inspection of the matrix in *Table 3.26.* reveals highest correlation of 0.65 between *GenVrbl* and *AVoc*, which is below the critical value of 0.70 for collinearity. The last point concerns the *SRPM* variable. It is interesting to observe that this non-verbal reasoning test correlates only marginally higher with verbal variables *AVoc* (0.42) and *GenVrbl* (0.44) than with language inductive reasoning variable *IndReas* (0.39). This is an intriguing fact since the opposite configuration was expected.

3.9.2. Results of MR for KET- A (Total)

In order to address **RQ2a** as formulated in *Introduction* the independent variables (cf. *Table 3.26.*) were entered into multiple regression analysis to estimate which combination would yield the highest level of explained variance (R^2).

First, the correlation matrix for the predictor variables and the criterion was calculated for each model. Second, multiple regression analysis was run on the matrix giving: 1). multiple correlation coefficients R; 2). squared R; 3.) standardised regression coefficients (BETA); 4.) *t*-test value; 5.) regression sum of squares (SS_{reg}); 6.) residual sum of squares (SS_{res}) and 7.) *F* ratio significance test of SS_{reg}[21]. Finally, the analysis of the residuals between the observed and the predicted scores was carried out. This was done by inspecting the scatterplots and the list of outliers. In order to 'improve' the predictive power of the model being considered the cases whose residuals were found to be greater than 2,5 *sigma* (units of a standardised residual deviation) were removed form the analysis and the regression equation recalculated again[22]. The removal procedure for outliers was continued until no substantial gain in R^2 was offered by the recalculated data. After each successive outlier was removed from the data the multiple regression analysis was carried out on a recalculated correlation matrix for a new N. The statistics reported below are the result of data-recalculation following the removal of 15 cases of outliers. Full results of multiple regression analyses for the KET-A Total criterion are included in *Appendix 10*. *Table 3.27.* below gives the correlation matrix and the summary of the results.

N = 125	p < 0.01					
	GenVrbl	IndReas	AssMem	TxtCom	SRPM	AVoc
IndReas	0.12					
AssoMem	0.12	-0.03				
TextComp	0.10	0.12	0.00			
SRPM	**0.50**	**0.40**	**0.25**	**0.29**		
AVoc	**0.65**	**0.27**	0.14	**0.46**	**0.55**	
KET-A	**0.56**	**0.39**	0.15	**0.41**	**0.56**	**0.70**

	Variable	BETA	t (118)	R	R^2 AdjR^2	Regres. sum of squares	Resid. sum of squares	KET-A Mean	KET-A St.D.	SEE
N=125	F(6,118)=27,860			0,766	0,586	24857,6	17547,2	42,0	18,49	12,19
	GeVrbl*	**0.233**	**2.699**		0,565					
	IndReas*	**0.203**	**3.040**							
	AssMem	0.048	0.777							
	TxtCom*	**0.166**	**2.340**							
	SRPM	0.107	1.317							
	AVoc*	**0.348**	**3.620**							

* BETAs statistically significant - variables entered in the regression equation

Table 3.27. Summary of MR analysis for KET-A Total

The first column gives the values of the standardised regression coefficients: BETA. BETA regression coefficient states the contribution of each independent variable in the regression equation. The second column contains *t* values, a statistic that tests the significance of each predictor's unique contribution to the criterion (tests whether BETA values are significant statistically). The next two columns contain multiple correlation coefficient R (and associated with it R^2) i.e., the correlation between the observed and the predicted scores. It is to be interpreted in the same way as a simple, bivariate Pearson's ***r***. Thus R^2 indicates the amount of variance in the criterion variable explained by a particular set of weighted predictors. The higher the value for R the greater the strength of the relationship between the regression model and the criterion. The next two columns: regression sum of squares and residual sum of squares, test the model fit against the data. The regression column displays information about the ratio of the variation accounted for by the model, whereas the residual column specifies how much of the variation is not accounted for by the model. Because the value of the *F* statistic exceeds the critical values for the given degrees of freedom the null hypothesis (that there is no relationship between a given predictor and a criterion) has to be rejected.

The next two columns provide another test of the 'fit of the model'. The comparison of the standard error of the estimate (SEE) with the standard deviation for the criterion variable shows that our best guess of a KET-A's score on the basis of the data gathered would be 42 (Mean) with a standard deviation of 18,49. The SEE value indicates how considerably lower the error of our estimate will be if a given subset of predictors is applied (cf. Hatch and Farhady, 1982:224). Apart form the statistics in *Table 3.27*. the normality of unstandardised residuals was checked visually using a histogram and a plot. No violations of the normality assumptions were recorded.

The examination of the statistics in *Table 3.27.* leads to the following observations. The analysis identified four best predictors from the total number of six. The *AVoc* variable contributes the strongest to the prediction of the criterion with its BETA=0.348, the second is *GenVerbl* (0.233), the third *IndReas.* (0.203) and the last one *TextComp*. (0.166). The BETAs for the *SRPM* and *AssocMem* variables did not prove high enough for them to be entered into the equation. The total amount of variance that we hope to account for in the criterion using the four predictors is 58.6% (adjusted to 56.5%). The value for the *F* ratio statistic is high: $F(6,118)=27,860$, which indicates a non-accidental relationship between the subset of predictors and the criterion. The comparison of the SEE (standard error of estimate) value (12,19) with that of St.Dev. for the criterion (18,49) indicates that we are 30 % more likely to make an error in predicting the criterion from the six variables than when basing our prediction on the subset of four predictors identified in the course of multiple regression. The discussion of the results is postponed till *Chapter 4*.

3.9.3. Results of MR for KET-A's components: *Competence* and *Skills*

In order to address **RQ2b** as formulated in *section 0.2.* two additional multiple regression analyses were made against each of the criterion variables factor analysed from the total second language proficiency variable (cf. *section 3.6.1.*). The two SLP factors identified were: *SLPCompetence* (knowledge) and *SLPSkills*. If it turned out that particular predictors (subtests) had differential weights in predicting each of the criterion variables than they can be interpreted in terms of different aspects of foreign language aptitude being involved in different aspects of criterion performance. This, in turn, would corroborate the hypothesis that separate dimensions of foreign language aptitude exist and make independent contributions to the prediction of foreign language learning success (Carroll 1981:100-101). The analyses followed the same procedure as in the case of KET-A Total in *section 3.9.2*. *Table 3.28.* gives the results.

A. SLP Competence

N = 125	p < 0.01					
	GenVrbl	IndReas	AssMem	TxtCom	SRPM	AVoc
IndReas	0.13					
AssoMem	0.16	-0.00				
TextComp	0.14	0.13	-0.02			
SRPM	**0.54**	**0.40**	**0.24**	**0.28**		
AVoc	**0.67**	**0.28**	0.15	**0.48**	**0.56**	
SLP Compet	**0.46**	**0.27**	**0.23**	**0.40**	**0.43**	**0.60**

	Variable	**BETA**	**t** (118)	**R**	**R^2** **$AdjR^2$**	**Regres** sum of squares	**Resid.** sum of squares	**SLPCompet** Mean	**SLPCompet** St.D.	**SEE**
N=125	F (6,118) = 14,943			0.657	0.432	31,990	42,101	0.07	0.77	0.59
	GenVrbl	0.157	1.520		0.403					
	IndReas	0.129	1.653							
	AsMem*	**0.156**	**2.163**							
	TxtCom*	**0.195**	**2.340**							
	SRPM	0.011	0.117							
	AVoc*	**0.339**	**2.958**							

B. SLP Skills

N = 122	p < 0.01					
	GenVrbl	IndReas	AssMem	TxtCom	SRPM	AVoc
IndReas	0.12					
AssoMem	0.12	-0.04				
TextComp	0.07	0.10	0.03			
SRPM	**0.49**	**0.40**	0.22	**0.29**		
AVoc	**0.65**	**0.23**	0.16	**0.44**	**0.52**	
SLP Skills	**0.62**	**0.49**	0.08	**0.29**	**0.54**	**0.67**

	Variable	BETA	t (115)	R	R^2 AdjR^2	Regres sum of squares	Resid.	SLPSkills Mean	St.D.	SEE
N=122	F(6,115) = 33,587			0.798	0.637	39,417	22,493	0.08	0.72	0.44
	GeVrbl*	**0.364**	**4.402**		0.618					
	IndReas*	**0.351**	**5.591**							
	AsMem	-0.009	-0,157							
	TxtCom	0.093	1.380							
	SRPM	0.048	0.642							
	AVoc*	**0.282**	**3.135**							

* BETAs statistically significant - variables entered in the regression equation

Table 3.28. Summary of regression analyses for *SLPCompetence* and *SLPSkills*

The comparison of the correlations among the six predictors and the *SLPCompet* criterion in *Table 3.28A.* with the results of regression analysis for the sub-variable of KET-A reveals the following. Out of six predictors used in the analysis three variables: *AVoc*, *TxtCom.* and *AssocMem.* prove to be the best predictors of *SLPCompetence* factor-loaded by *Grammar* and *Vocabulary* tests of KET-A. The total variance of *SLPCompet.* criterion explained by a set of weighted predictors in A is 43.2%, (adjusted R^2=0.403). *AVoc* contributes to this effect with BETA=0.339; *TextComprehension* with BETA=0.195 and the rote memory variable *AssMem* with BETA=0.156. What is interesting is that, although *GenVrbl* is the second best to correlate with the criterion (r=0.46), and *SRPM* is the third best (r=0.43) after AVoc (r=0.60), neither of the two enters the equation. Furthermore, *AssocMemory* which has the lowest correlation with the criterion (r=0.23), turns out to be the third best predictor outstripping *InducReas*. (r=0.27), *SRPM* (r=0.43) and *GenlVrbl* (r=0.46).

In the case of *SLPSkill* the variable with the highest BETA is *GenVrbl:* r=0.364, followed by *IndReas* r=0.351 and *AVoc* r=0.282 (cf. *Table 3.28B.*). It is noteworthy that despite the highest correlation with the criterion, *AVoc* (r=0.67) is only the third best in contributing to the prediction. It is also interesting to observe that *SRPM* which correlates with the criterion at r=0.54 does not enter the regression while *IndReas* does at r=0.49. The total amount of variance in the criterion (*SLPSkill*) explained by the three predictors in *Table 3.28B.* is 63.7%. In both A and B regression models there is a considerable gain in the precision of prediction. If the models are accepted the error value as expressed by SEE will be lower than if we were to make the prediction on the basis of mean scores and standard deviation only. The prediction in the case of *SLP Skills* is slightly more precise than for *SLP Competence* as reflected in the difference between St.D. and SEE values.

Summarising, it is to be concluded that in both regressions a subset of three variables out of six may be used to efficiently predict the respective SLP criteria. One predictor, *AVoc* appears in both analyses but in reversed roles: it is the strongest predictor in *SLPComp.* but the weakest in *SLPSkills*. The remaining two predictors are four aptitude variables and they enter the regression equations in a similar, crosswise fashion: *TextComp* and *AssMem* for SLP Competence while *GenVrbl* and *IndReas* for SLP Skills. SRPM is the only variable that never enters into any analysis. The discussion of the implications of the results will be presented in *Chapter 4*.

NOTES

1. All the correlation coefficients reported are Pearson's product moment ***r***.
2. The terms 'component' and 'factor' are used interchangeably although in the literature on the subject the former is applied to the principal component (PC) method of root extraction while the latter is used with reference to those methods which rely on split variance rather than on total variance between variables i.e., principal factor (PF) methods such as principal axis factoring or Thurstone's centroid (cf. Rietvald and Hout 1993:267).
3. Many researchers (cf. Carroll 1993, Sasaki 1999) use principal component (PC) method for initial factor extraction, with unities on the diagonal of the correlation matrix, only to determine all latent roots to be extracted. This is typically achieved through the application of Kaiser-Guttman rule and Cattell's scree test. In further analysis, however, they use the principal factoring. In the present study, principal axis (a PF method) is used for initial factor extraction, as no major differences in the number of suggested factors were noticed when the PC method was used.
4. As noticed by Carroll this principle "undermines the number of factors when factors are substantially correlated, and overestimates when the average correlation among variables is low and there are a large number of variables. Indeed, even completely random data can have multiple roots greater than unity" (Carroll 1993).
5. Rotating all the factors does not make sense as the result will be exactly the same number of variables originally entered in FA.
6. Although factor loadings can be interpreted in the same way as correlation coefficients their actual significance level is higher. As Rietvald and Hout remark: "it is customary to take only those factor loadings into consideration that exceed 0.30 or 0.40" (Rietvald and Hout 1993: 274). As rule of thumb the value specified as significant for a given N is doubled. For example, the value for our sample size (N=211) would be around 0.21, so an acceptable, statistically significant factor loading is 0.42.
7. The effectiveness of the hierarchical factor analysis procedure in uncovering a nature of the relationship between a multiplicity of correlated variables was clearly demonstrated by, for example, Carroll's (1983) reanalysis of Scholz's et al. (1980) data.
8. Specifically, *principal component* (PC) is a preferable method for data reduction purposes while *principal factor* (PF) method is preferred for structure discovery.
9. On this Carroll remarks: (...) this fact has sometimes been appealed to in criticising factor analysis on some such basis as 'a factor analysis gives no more information than what you put into it' (it is hard see how the case could be otherwise) or even 'garbage in–garbage out'.

Such criticisms are not reasonable; a well-designed factorial study can yield information that is by no means immediately apparent from the input" (Carroll 1993:52).

10. For a profoundly imaginative argument against the 'indeterministic' view of casual relationships, expressed in a literary form cf. S. Lem's novel *Katar*.
11. The slight differences in ***r*** are due to the differences in the number of cases (N) for which each of the matrix was calculated (cf. *section 2.2.1.*)
12. That the inter-correlations among the four FL aptitude variables are non-significant is the result of a method induced procedure of factor analysis which always aims at producing a structure in which the factors are not correlated with each other (orthogonal). All the remaining variables are 'raw' variables.
13. An average correlation among the three factors was: r = 0.540
14. For an introduction to regression analysis cf. Hatch and Farhady 1982; for more detailed accounts see Guilford 1960, Rietvald and Hout 1993, Ferguson and Takane 1999 or Brzeziński 2002.
15. On how the best predictors are chosen in stepwise multiple regression cf. Ferguson and Takane (1999:530).
16. For descriptive statistics of the variables see *Table 3.2.*
17. The examination of the scatterplots also confirmed the assumption of linearity of the relationships between the variables as a precondition to multiple regression analysis (cf. *section 3.2.4.*).
18. cf. Guilford 1960, Góralski 1974, Niemierko 1999, Juszczyk 2001.
19. Regrettably, this is where the research referred to in *section 1.6.* stop (cf. Nieżegorodcew 1975, 1979; Lewowicki 1975; Figarski 1984; Jankojć 2001).
20. Recall that the technique of factor analysis yields factors which are very weakly correlated with each other.
21. *F* ratio tests whether the regression sum of squares (SS_{reg}) is caused by random fluctuation in the sample or whether it reflects an actual relationship between the dependent and independent variable.
22. By no means is this the case of 'concocting' or bending data. The outliers are known to affect statistical solutions to quite a degree and they either reflect the imperfections of the measuring tools or untypical behaviour attributed to random error.

CHAPTER 4

DISCUSSION OF THE RESULTS

This chapter contains the discussion of the results presented in *Chapter 3* with reference to the research questions posed in *Introduction*. The order in which this is done follows the order of the presentation of the results in *Chapter 3*. Internal structure of second language proficiency as measured by KET-A is discussed in *section 4.1. Section 4.2.*, on the other hand, is devoted to the discussion of the internal structure of FL aptitude. The discussion of the nature of the relationship between SLP, FL aptitude and intelligence is presented in *section 4.3.* The last section (*section 4.4.*) deals with assessing the usefulness of the study's aptitude instruments for predicting SLP in the sample studied. Then the implications of the findings are considered with special emphasis on their relation to the field of applied linguistics.

4.1. Structure of Second Language Proficiency

ARQ 1. *What is the structure of second language proficiency (SLP) as revealed by the scores from KET-A (Total) test?*

The need for this question seemed compelling considering a claim implicit in an FL aptitude construct to the effect that it is possible to extrapolate from a posited set of language learning potential (as measured by aptitude tests) to an actual performance and/or achievement measured by standardised language proficiency tests. Additionally, if this extrapolation from a postulated set of FL aptitudes were to go beyond a mere prediction of L2 learning success but rather inform a researcher, teacher and/or learner about the structure of the target ability and the nature of the very learning process involved in attainment of this ability then the necessity to determine and/or define second language proficiency becomes apparent. In other words, it needed to be determined, before anything else, what the nature of SLP of the subjects studied and revealed by KET-A was.

4.1.1. Discussion of ARQ 1

Factor analysis of the six variables forming the KET-A Total measure carried out in *section 3.6.1.* leads to the following observations:

1. The factor solution that lent itself best to interpretation in terms of language abilities making up L2 proficiency produced a fairly clear hierarchical structure (cf. *Table 3.12.*) with one second-order General SLP (G-SLP) factor and two first-order, specific factors (S-SLP). The

G-SLP factor (influencing two remaining, lower-order factors) is responsible for the subjects' global performance on the KET-A test. Next to this general factor there are two S-SLP factors highly loaded by *Grammar* (0.411) and *Vocabulary* (0.340) variables, and by *Listening Comprehension* (0.327), *Reading Comprehension* (0.260), *Writing* (0.298) and *Speaking* (0.162) variables respectively. The first factor, being associated with 'knowledge' variables, can be best labelled language competence factor – *SLPComp*. The other specific factor, comprising all the skill subtests of KET-A, can be best interpreted as language skills factor – *SLPSkill*. Consequently, second language proficiency structure embodied in KET-A can be said to reflect the traditional view of language proficiency as exemplified in the **skills** versus **linguistic competence** distinction. Over those two S-SLP areas there presides a strong, G-SLP factor composed off the proficiencies contributed by all six variables

2. The G-SLP factor correlates at 0.915 with the two S-SLP factors: *SLPComp* and *SLPSkills* which, in turn, intercorrelate with each other at 0.404. The strength of the relationships suggests that the specific areas of L2 proficiency are much determined by the G-SLP while being relatively independent of each other. Consequently, the data under consideration, support neither the 'unitary' nor the 'divisible' hypotheses (cf. *section 0.2.*) on the nature of second language proficiency but the view which is a compromise between the two extreme positions. In a practical dimension this would mean that it is impossible to make any judgements about a subject's L2 proficiency solely on the basis of one, general SLP test, such as Cloze or dictation, as envisaged by Oller (1979, 1981). On the other hand, an atomistic view of language proficiency as expressed in the hope that SLP consist of the sum of its component parts (the so called 'discrete point testing') cannot be maintained either.

4.1.2. Summary of the discussion of ARQ 1

The controversy between the 'unitary competence' versus the 'completely divisible competence' hypotheses on the structure of SLP originated with Oller (1978) where he argued, following Spolsky's (1973) rejection of the skills/components view of SLP, for a single trait rather than several distinct constructs as proposed earlier by Lado (1961) and Carroll (1968). The structure of second language proficiency as revealed by a hierarchical factor analysis performed on the gathered data offers evidence in favour of a multidimensional and hierarchical organisation of L2 competence.

Specifically, it is argued that SLP of 13-year-olds as measured by the KET-A test and teachers' assessment of the subjects' oral performance revealed two specific factors labelled here as *Competence* and *Skills*, over which there presides a *General Language Proficiency* factor determinant of an integrated language ability. Finally, on the basis of the analysis of the structure of SLP underlying the KET-A test it may be concluded that the test, as modelled after the 'proper' Cambridge KET test and being a 'mutilated' version of it, reflects well the original structure of the test. Consequently, the test proved a valid and relevant measure of the subjects' L2 achievement/proficiency

4.2. Structure of Second Language Aptitude

ARQ 2. *What is the nature of the abilities measured by the two aptitude tests used in the study?*

As it will be recalled from *section 2.4.2.* the focus of the aptitude part of the present study was on aptitude constructs which emphasise context-handling abilities of a learner in the L2 learning tasks. Furthermore, following Skehan (1991, 1998), memory-for-text abilities were posited as potentially determinant of L2 learning success. Additionally, the test made use of a new 'test method facet' (Bachman 1997) in the form of a 'dynamic testing paradigm' (Grigorienko et al. 2000).

With reference to Carroll's (1962 and later) standard view of aptitude the following tasks, either directly or indirectly, incorporated the abilities Carroll included in his 'four-component' model of aptitude:

1. Phonemic Coding Ability: *wingdings* and *dbrdszn*
2. Grammatical Sensitivity: *pojezierze* and *dictation*
3. Inductive Language Learning Ability: *TPJ 7* and *TPJ 8*
4. Rote Learning for Language Materials: *lubić* and *pamiętać* ("Akareke" task)

Additionally, new abilities were posited for the study:

- verbal ability (lexis in distorted versus non-distorted context): *dbrdszn/ zwolennik* and *pojezierze/sprawcy*
- memory for text (visual and aural): *ofiarą*, *handlarz*, *układ* and *przed*, *podczas*

Because only three of the tasks mentioned above were direct and 'pure' implementations of Carroll's (1961, 1962) original capacities (*lubić* and *pamiętać* for Carroll's Part 5 of MLAT - Paired-Associates and *TPJ 8* for Inductive Language Learning Ability - not included in MLAT) it was necessary to factor analyse the posited abilities in order to determine the nature and structure of the construct of FL aptitude to be used later for prediction purposes.

4.2.1. Discussion of ARQ 2.

The analysis of the data obtained from 15 measures of abilities making up the FL aptitude battery and carried out in *section 3.7.* unravelled the following factor structure of L2 learning abilities studied (cf. *Table 3.22.*):

1. The first and strongest factor to emerge may be interpreted as a ***General Verbal Skill*** factor. It is significantly loaded by four 'vocabulary' variables: *sprawcy*, *pojezierze*, *dbrdszn, zwolennik* and an auditory memory (immediate recall) variable: *dyktando*. There are two interesting points in this connection. Firstly, it is somehow surprising to find the dictation task loading the verbal factor. It was hypothesised that *dyktando* (tapping immediate recall abilities as measured by auditorily presented text of increasing length) would be factored together with other memory-for-auditory-material variables i.e., *przed/podczas*. That this was not the case suggests that the ability involved in the task is more related to a skill of integrating the meaning of heard/dictated language (semantic integration) than to mere auditory-span memory and repetition abilities. Consequently, it suggests that memory for aural text is dependent on an ability to form semantically cohesive representation on the basis of a transient signal, an ability mediated by general verbal skills. Secondly, the analysis did not confirm the hypothesis that two qualitatively different verbal processing abilities are involved in context-embedded (*sprawcy/pojezierze*) and context-disembedded (*dbrdszn/zwolennik*) language. Both groups of tasks load on the same verbal factor. (cf. however, point 4 below). The factor is additionally loaded (non-significantly, though) by two variables: *TPJ 7* (0.419) and *układ* (0.401). In both variables lexical abilities are involved: *TPJ 7* relies on the ability to analyse grammatical form in an analogy task which is lexis based; *układ*, on the other hand, calls for a visual memory for words-in-text i.e., for an ability to link a propositional meaning of fragments of a text with its spatial organisation. The total amount of variance explained by the factor is 20.3%
2. The second factor that emerged from the factor analysis can be univocally identified as ***Inductive Reasoning*** factor. It is highly loaded by *TPJ 7* (grammatical analogy task) and *TPJ 8* (a miniature artificial language). The respective loadings on the factor are: 0.569 and 0.794. The abilities of analogy formation and inductive rule acquisition, having been convincingly argued to lie at the bottom of general inductive reasoning abilities (Pellegrino and Glaser 1980; Sternberg 1982; Boyle 1987; Skehan 1991, 1998), may thus be justifiably claimed to form one inductive reasoning skill. Out of the remaining variables with non-significant loadings on the factor, *handlarz* (reading comprehension task) and *przed* (listening comprehension task) show the highest values (0.324 and 0.274 respectively). The interpretation of this fact is not entirely clear. It may be that both variables, being related to memory-for-text skills, rely in meaning integration on inductive abilities to reproduce meaningful language from insufficient context. The factor explains 9.4% of the total variance
3. The third factor is almost exclusively loaded by the two rote variables *lubić* and *pamiętać*. The variables' loadings are: 0.709 and 0.603 respectively. As such the factor is interpreted as

Paired-Associate Memory. It receives very low and mostly negligible loadings from all the remaining variables, the highest being 0.222 of *dbrdszn*. The fact that the factor is so well defined may suggest two things: there is very little overlap between the abilities tapped by the factor and other factors or, alternatively, the factor, because it is so distinct from all the other abilities implicated in the variables, emerges as a very 'robust' dimension (cf. note 2 in this chapter). The factor has the lowest amount of explained total variance: 7.3%

4. The fourth factor loaded by the three reading comprehension variables (*ofiara/handlarz/układ*) can be interpreted as ***Written Text Comprehension and Memory*** factor. It seems that the ability tapped by the three text-handling variables is related to memory-abilities for meaningful material in connected text. That a memory ability is implicated in the tasks can be seen by a very high loading of the *ofiara* variable (the first text comprehension task after the input text) and a lower contribution by *handlarz* (.480) which came after two more interfering tasks. The visual/spatial memory-for-words task *układ* which came last after a considerable delay in time loads the factor at 0.481. That memory skills are not the only abilities implicated in the factor can also be seen by high but non-significant loadings of two vocabulary variables. In this connection it is interesting to observe the differential contribution made to the factor by *sprawcy/pojezierze* (0.314 and 0.354) and *dbrdszn/zwolennik* (0.115 and 0.189). Although the loadings are non-significant in the case of both groups the difference between them is large enough to suggest that the four lexical-variables tap different underlying abilities in handling words. It seems that *sprawcy/pojezierze* differes from *dbrdszn/zwolennik* by an ability to infer lexical/grammatical meaning of a word from a phrase/sentence/paragraph context. Apparently, the two tasks share this skill with what is called for in written text comprehension tasks of *ofiara/handlarz/układ* i.e., an ability to extract and remember propositional content of a connected text
5. The variables *przed* and *podczas* do not exhibit significant loadings on any of the factor. They comprise memory-based abilities linked with handling text comprehension in an aural mode and show weak and non-significant loadings on two factors: 1 and 3. Their loadings on ***written text comprehension*** (0.260 and 0.337 respectively) are higher than their respective values on the ***general verbal skill*** factor (0.148 and 0.250). This, together with the fact that the variables did not create a separate aural text comprehension factor might suggest that there exists one memory-for-text ability which operates irrespective of the language processing channel but which did not reveal itself in the case of the listening task due to its weak psychometric properties (cf. skewed distribution in *Table 3.2.* and low communalities in *Table 3.22.*). However, an alternative explanation might point towards instructional factors as a potential reason for non-significance of the aural channel. Considering the fact that

language teaching pedagogy in almost all the schools whose learners formed subjects for the present study overemphasises the written word and neglects oral/aural skills during L2 instruction this explanation seems a feasible one

6. The variable *windings* insignificantly loading ***general verbal skills*** factor (0.356) shows negligible loadings on the remaining three factors. That the variable should load ***general verbal skills*** factor is somewhat surprising – rather it was expected to go hand in hand with *dbrdszn* and the posited ability to form a link between meaningless units of the 'signifier' (graphemes and phonemes) and meaningful units of expression ('signified') (cf. *IV* in *section 2.4.*). It may be tentatively hypothesised that the skill posited for *wingdings* was more about handling the items of the task holistically as 'lexical items' rather than processing them on the analytical, 'symbol-meaning' basis. Alternatively, weak psychometric properties of the variable i.e., its low communality and weak coverage of a posited ability by adequate tasks, might be the reason for non-emergence of a separate 'phonetic-coding' factor.

4.2.2. Summary of the discussion of ARQ 2

All in all, four aptitude factors have been isolated from the battery of tests used in the study. Components of aptitude identified in the course of factor analysis comprised the following abilities (listed from the greatest to the smallest amount of explained variance): 1. vocabulary knowledge and fluency; 2. memory for text (written); 3. language analytical abilities and 4. rote memory abilities. At least two postulated abilities did not emerge as underlying traits for the complex of language learning related skills. They were: memory for text (auditory) and a modified phonemic-coding ability (cf. *IV* and *Table 2.2.* in *section 2.4.2.*). Of possible reasons for the non-emergence of separate, strong factors capturing the abilities posited weak ability diversification and far from perfect psychometric properties of some of the tasks can be adduced.

If the abilities identified in the course of factor analysis can be further shown to be at least partially independent of other cognitive factors (verbal and non-verbal intelligence) then, the fact that several dimensions rather than one have been found underlying the posited language learning abilities is an additional proof of the non-unitary nature of second language abilities.

4.3. Language as Intelligence?[1]

RQ 1. *Is intelligence equivalent to L2 proficiency and/or to FL aptitude, or are they independent of each other?*

The problem posed in this research question was whether SLP and, underlying it, FL aptitude (understood here as a postulated group of second language learning abilities) can be claimed to be independent from a more general set of cognitively based abilities associated with intelligence. Although in his later writings Oller (1978)), with whom the controversy began, did not seriously consider the idea of equivalence of intelligence and language abilities he is still often quoted in the debate as the one who favours (or rather: favoured) the 'unitary' hypothesis of the structure of human abilities akin to Spearman's 'g'- intelligence factor. Oller himself puts this quite clearly:

> *Indeed, it seems that in an adequate theory the holistic elements must depend on the interaction of the relatively analytic components. Hence, the idea that global and particulate models were incompatible must have been quite wrong (...). It would seem that both views are needed and that they can complement each other rather than contradict one another*. (1983: 36).

Following Boyle (1987) it was speculated that if it could be shown that, as the result of factor analysis, the variables of SLP, FL aptitude and intelligence did not share membership of any of the emerging factors but, on the contrary, formed distinct and non-overlapping factors, then it could be claimed that the three ability areas enjoy 'perfect separateness' from each other – a case for 'completely divisible' hypothesis, as formulated in *section 0.2.* If, on the other hand, it could be shown that all, or some of the variables merge into one factor then this could be taken as evidence for a 'unitary' view of human abilities; a perspective which emphasises a common cognitive 'deep level' for virtually all human endeavour.

The argument implicit in this thinking was that if L2 competence was shown to be independent of intelligence (as measured by the tools used) then *ipso facto* this could be taken as a sign of independence from intelligence of language learning abilities themselves. The results presented in *section 3.8.* and discussed below argue that the 'implicational' argument was only partially correct.

4.3.1. Discussion of RQ 1.

The factorial structure of the abilities analysed in *section 3.8.* revealed that:

1. SLP abilities form a strong, independent factor of their own, called here **Second Language Proficiency Factor** (Factor 2 in *Table 3.24.*) with little contribution from the other two ability domains i.e., FL aptitude and intelligence. All but one SLP variable (*Speaking*) load highly on this factor (average loading for 5 SLP variables = 0.740). Of the two FL aptitude variables the *InductiveReasoning* variable contributes slightly higher than *GeneralVerbal* to the factor indicating that language analytical skills tapped by *TPJ 7* and *8* (cf. *section 2.4.1.*) are implicated in L2 proficiency to some extent. However, the loadings for neither of the two variables reach the level of statistical significance. Of the three intelligence variables only the verbal-intelligence (*PVoc* and *AVoc*) variables could be said to contribute to the SLP Factor to a noteworthy degree. This slight degree of overlap between the factor and the verbal

intelligence variable suggests that while in both tests vocabulary is measured by almost an identical set of testing techniques the abilities referred to are nonetheless different. A slightly higher loading on the SLP Factor of *AVoc* (0.321) as compared to *PVoc* (0.299) probably reflects the 'productive' dimension that this sub-test captures together with the five SLP variables. It is to be noticed that in all the variables subjects are expected to display their command of vocabulary in an array of tasks which call more upon the retrieval than recognition skills

2. FL aptitude and intelligence exhibit some degree of overlap, which is an indication that their constructs, being not as 'robust' and monolithic as that of second language proficiency (as measured by KET-A), appeal to a similar repertoire of skills and/or potentials (cf. again *Table 3.24.*). Specifically, two areas of overlap are conspicuous. The first is marked by the appearance of a reasoning factor loaded by Raven's test of non-verbal reasoning abilities (*SRPM*) and by FL aptitude reasoning factor (*IndReas*) comprising the TPJ 7 and TPJ 8 variables. The other area covers a general verbal skill factor loaded by FL aptitude *GenVrbl* and *TxtCom* sub-tests and by two verbal intelligence measures (*TJLPVoc* and *AVoc*)
3. In this context **Inductive Reasoning Factor** (Factor 4 in *Table 3.24.*) is particularly interesting. It is highly loaded by Raven's non-verbal test of inductive reasoning (0.707) and by a language analysis component of FL aptitude (0.743). The factor also receives significant loading from FL aptitude *GenVrbl* factor (0.431) and from *SLPSpeaking* grade (0.402). Apparently, reasoning abilities, irrespective of whether they are exercised on linguistic (FL aptitude) or non-verbal (intelligence) material, make recourse to largely the same underlying cognitive processes of symbol-manipulation responsible for the very analytical operation. It is this underlying affinity of the two variables that saturate the factor under consideration
4. The other area of overlap between FL aptitude and intelligence variables is seen in connection with the TJL tests purportedly gauging verbal intelligence. The factor that emerged is composed of very heavy loadings from both TJL tests (*AVoc* = 0.770 and *PVoc* = 0.834) and from FL aptitude (*TextCom* = 0.776). It also receives low but significant loadings from FL aptitude *GenVrb* (0.482) and from *SLP Speaking* (0.403). Apparently, the type of abilities expressed by the factor emphasise lexical skills in handling vocabulary items in a variety of cognitively based tasks. Looking at the ability-structure of the four most contributory variables, the **Verbal Factor** could be identified as involving abilities having to do with: 1. contextualised (*Pojezierze* and *Sprawcy*) and de-contextualised (most TJL tasks) lexis; 2. passively recognised or actively recalled lexis; 3. language with reduced (*dbrdszn*) or excess (*zwolennik*) redundancy; 4. lexis stored visually (*układ*) or auditorily (*dyktando*) and 5. recalled immediately (*dyktando*) or with delay (*ofiara, handlarz*). With respect to the

above mentioned abilities puzzling is the fact that the *GenVrbl* loads the Verbal Factor almost twice as low as the *TextCom* variable does. Considering that the correlations (cf. *Table 3.23.*) between, on the one hand, *GenVrbl* and, on the other, *PVoc* and *AVoc* (0.55 and 0.65 respectively) are higher than the correlations between *TextComp* and *PVoc/AVoc* (0.55 and 0.45 respectively) we would expect higher correlating variables to be grouped together as associated with the same trait/ability. That this is not the case with *GenVrbl* but applies to *TxtCom* suggests that the verbal skills implicated in the latter have more in common with passive/active and contextualised/de-contextualised ability types than with the abilities that had been posited for it. It will be recalled that the three tasks making up *GenVrbl* (cf. *Table 2.2.*) were designed to measure, what Skehan (1989) calls, 'memory for text' (cf. *section 2.4.2.* and *4.2.* above) in visual mode of presentation. It seems that the 'memory for text' ability is not a viable skill for the subjects studied and that text processing is a task relying more on lexis (possibly a memory for individual key words) than on semantic integration of the text's meaning stored in short-term memory

5. The factor that accounted for the least amount of the total variance (10%) could be interpreted tentatively as **Paired-Associate Memory Factor**. It is the least clear factor of all four (cf. *Table 3.24.*) despite its being very highly (and almost exclusively) loaded by one FL aptitude variable - *AssocMem* (0.857)[2]. One SLP variable: *Speaking,* contributes its loading to the factor (0.414), while the remaining five SLP variables display very low, insignificant loadings on it (in fact two are negatively correlated with it). This fact together with the observation that the factor is very weakly loaded by the three intelligence variables but receives close-to-significant loading from FL aptitude *GenVrbl* variable again suggests that verbal-intelligence and verbal-aptitude dimensions are indeed (slightly?) different constructs.

There are two remaining points of interest in connection with the factor structure of the abilities discussed here. The first of those concerns the *SLP Speaking* variable and its involvement in all factors to, more or less, equal degree. The variable has quite satisfactory communality value (h^2= 0.483) and its loadings on the factors are: Verbal = 0.403; SLP= 0.318; Paired-Associate = 0.414 and Inductive Reasoning = 0.402. The fact that the speaking skill is not strongly associated with SLP variables (as was expected) might be due to the specificity of the demands it puts on the learners. As Carroll, in connection with his own study, remarks:

> *(this) may indicate that the requirements of the task set by this test are fairly specific and possibly to some extent unrelated to the measurement of language proficiency. (the examinee's) performance may depend not so much on his language proficiency as on his general fluency and self-confidence in spontaneous speech – whether in his native language or a foreign language* (1983:95).

Alternatively, it might be argued that the results for this variable, by virtue of being qualitatively different[3] from the remaining five, have had a 'flattening' effect on the factor loadings exhibited by *SLP Speaking*, had it not been the case that the correlations of the variable with the KET-A Total (r= 0.58) as well as with other sub-tests (x LC= 0.40; x RC= 0.87; x GR= 0.52; x VC= 0.40; x WR= 0.56) show that the correlation pattern of the variable does not differ much from the correlation patterns of the other variables either with KET-A Total or with each other.

Secondly, further analysis (cf. *section 3.8.*) revealed that a second-order factor structure (cf. *Table 3.25.*), while confirming the specific component structure of SLP in the form of four primary factors, yielded two, rather than one, higher-order factors. The examination of the second-order factor loadings would warrant their interpretation as **General Factor** and **Rote-Memory Factor**. The first of those is significantly loaded by all but one variable namely, FL aptitude rote-memory which itself loads the other higher-order factor together with *SLP Grammar*. This association of rote and grammar variable would point to the perseverance in Polish schools of traditional methods of grammar teaching relying heavily on rote-learning. The second-order factors are to be interpreted as influencing all the first-order factors. In other words, the specific four components of abilities identified in the factorial analysis are determined by the two general factors appropriately to the magnitude of the specific loadings they receive from the variables. The General Factor determines three first-order factors: Verbal, SLP and Inductive Reasoning with respective correlations: 0.736; 0.765 and 0.704, while the Rote Factor underlies its first-order counterpart and correlates with it at r = 0.404.

4.3.2. Summary of the discussion of RQ 1

Summing up, the results of the present study are generally in agreement with what earlier research had found on the relationship between intelligence and L2 proficiency (von Wittich 1962; Gardner and Lambert 1965; Oller 1979, 1983; Genesee 1976, 1987; Flahive 1980; Boyle 1987 and Sasaki 1996) and between intelligence and FL aptitude (Gardner and Lambert 1965, 1972; Wesche et al. 1982 and Sasaki 1996)[4]. Namely that while the three are related by no means are they identical.

However, a slightly different picture of the structure of the abilities studied emerges. Hierarchical analysis yielded two second-order factors: **General Cognitive** and **Rote** presiding over four separate but correlated first-order factors: **Verbal**, **SLP**, **Rote** and **Inductive Reasoning**. No separate aptitude factor was produced pointing to an overlap between aptitude and intelligence variables which turned out to be cross-related. The **General Cognitive Factor** correlates with the four first-order factors at: 0.736; 0.765; 0.042. and 0.704, while the **Rote Factor** correlates with the same four at: 0.299; 0.261; 0.404 and –0.163.

Leaving aside the precise nature of the factorial structure of human abilities sampled in the tools used, the least that can be said from examining the pattern of cross-overlap between the three domains studied in the data at hand is that:

1. FL aptitude and intelligence are not identical but related constructs
2. FL aptitude and intelligence are separate from SLP.

Consequently, neither the 'unitary' nor the 'completely divisible' view of human general mental abilities receives support from the data at hand. Rather it is the weak form of the two extreme positions, assuming both a general factor(s) and several specific factors (the 'compromise hypothesis'), that is corroborated by the data. Specifically, the study has shown that SLP as a specific domain stands out as distinct from both intelligence and aptitude. The latter two, in turn, show some degree of overlap differentially contributing to specific factors identified in the course of analysis. The fact that all the three areas are cognitively related is shown by the emergence of a second-order general factor underlying all the three domains of human faculties.

4.4. Prediction of Second Language Proficiency

RQ 2. *What combination of six independent variables used (2 intelligence and 4aptitude variables) allows for the best prediction of:*

1. *general SLP as measured by KET-A Total*
2. *the two components of SLP as expressed by* SLPCompetence *and* SLPSkills *factors*

in the sample studied?

This question addressed the problem of prognosis of second language learning success in the sample studied as measured by the end-of-year achievement measures: KET-A and teachers' spoken grades. Six variables were used in the prediction part of the study. They were:

- four FL aptitude variables factor analysed from the total of 21 variables (TPJ and PLAT):
  - a. general verbal skills (*GenVrbl*)
  - b. text comprehension i.e., memory for text (*TextComp*)
  - c. inductive reasoning (*IndReas*)
  - d. paired-associate memory i.e., rote (*AssocMem*)
- two intelligence variables:
  - a. Raven's test (non-verbal reasoning) (*SRPM*)
  - b. verbal intelligence TJL – active vocabulary (*AVoc*)

The variables' predictive potential was analysed using the stepwise multiple regression analysis. The results of the analysis, presented in *section 3.9.*, are discussed below, first with reference to RQ2a (General SLP) and then with reference to RQ2b (the *Competence* and *Skills* component of SLP).

4.4.1. Discussion of RQ 2a.

The analysis of the prediction potential of the variables studied in *section 3.9.2.* leads to the following observations:

1. Four variables, out of total six used, were found to be the best predictors of second language achievement (KET-A Total) for the sample studied. The subtests, ordered here by their unique contribution (BETA weights) to the prediction of the criterion, were: *AVoc*: 0.348; *GenVrbl*: 0.233; *IndReas*: 0.203; *TextComp.*: 0.166. The total amount of predicted variance, as indicated by the statistic R^2, is 56.5%. That is, using this particular combination of predictors we may hope that almost fifty seven percent of variability in the criterion will be attributed to our subset of four predictors. In other words, by using the prediction based on the regression model we can hope to reduce the error variance to 43.5%
2. The fact that as many as four variables are found important in accounting for the criterion suggests that the criterion itself is a multi-ability construct whose respective dimensions are linked to appropriate amount of unique variance explained by each predictor. This, in turn, implies that foreign language aptitude is a multi-dimensional construct, otherwise we would expect only one predictor to claim all the predictive power of the criterion. By the same virtue, the fact that two variables seemingly measuring the same construct: *AVoc* and *GenVrbl*, which is reflected in their moderately high intercorrelation of 0.65, are entered into the equation suggests that each contributes uniquely to the prediction. Consequently, the prediction model shows that traditional verbal intelligence measures tap qualitatively different abilities from those involving lexical material and postulated for FL aptitude construct (cf. *section 4.4.2.* for more on this issue)
3. In connection with high predictive power of *AVoc* and *GenVrbl* variables it is to be noted that Carroll in his study of two groups of adults (N_1=80; N_2=88) learning Chinese reached completely different conclusions. He found that "native-verbal ability or verbal intelligence as measured by fairly difficult English vocabulary test plays little role in foreign language learning" (Carroll 1981: 105). Other studies, in turn, attach more significance to this aspect of aptitude/proficiency continuum, cf. English vocabulary subtest in PLAB (Pimsleur 1966, 1968). Apparently what seems to be involved here is the mode of instruction predominant in training. If it heavily emphasises verbal aspect of the native language in the understanding of

the content of the instruction then it puts learners with low verbal abilities at a disadvantage (Carroll 1962, 1981). Alternatively, the situation might obtain equally likely if the instructional methods showed too much reliance on the written medium of relatively high complexity. As has been noted elsewhere in this study, this was very much the case in all the schools studied

4. Of the two variables that did not enter the analysis i.e., *SRPM* and *AssocMem*, the former had the second highest correlation with KET-A together with *GenVrbl*: r= 0.56 (cf. *Table 3.27.*) still it was not accepted into the regression analysis. At this point it is worth mentioning again that if the present researcher based his choice of the best predictors of L2 proficiency exclusively on simple correlational relationships offered by the matrix his conclusions would grossly overestimate the influence of the *SRPM* variable. Apparently, multiple regression by taking into account partial correlations between the independent variables in the analysis of their relationships with the criterion tells a researcher much more than the simple examination of the correlation matrix. The other variable (*AssocMem*) showed insignificant correlation with the criterion (r=0.15) and, as expected, did not enter the regression analysis
5. The non-inclusion of the *SRPM* variable into the prediction model implies that the abilities the Raven's test draws on i.e., non-verbal reasoning skills are not determinant of the general second language ability as measured by the KET-A test. On the other hand, the fact that the variable *IndReas* contributes uniquely to the prediction indicates that reasoning skills involving linguistic rather than non-linguistic material **do** make a difference. Similar conclusions have been reached by, for example, Carroll (1965, 1981); Gardner and Lambert (1967) and Boyle (1987). Also Chapelle and Green (1992) in their study of field in/dependence as a possible factor in L2 learning suggest that Skehan's (1998) language analytic ability (one of his information processing triad: phonetic coding ability, language analytic ability and memory) should be distinguished from general cognitive analytic abilities. They remark further that: "Such a distinction may be more informative for isolating abilities associated with success for different learners in various circumstances and for clarifying the aptitude-intelligence overlap question" (Chapelle and Green 1992:77) (cf. *section 1.3.3.2.* for Skehan's model of FL aptitude).
6. Similarly, the non-inclusion of the rote memory variable *AssocMem* into the regression equation indicates that the abilities involving mechanical type of associative memory are not predictive of general second language proficiency. Almost negligible BETA weight for the rote memory variable (0.048) suggests that it does not make any unique contribution to the total variance of the criterion. It may thus be concluded that the abilities involving the associative construct of memory despite the dynamic format of the task's structure (cf. the

discussion of the dynamic testing paradigm in *section 1.3.3.2.*), whereby the task simulates the learning process in three stages: presentation, rehearsal and testing by recognition, is either marginally involved or altogether irrelevant in L2 learning in the sample studied. On the other hand, memory for text, tested immediately and with delay (*TextComp*), proves to be a better predictor of the criterion (BETA = 0.166)

7. The study's findings on the involvement of memory are also in agreement with Skehan's research on aspects of memory in L2 learning (Skehan 1980, 1982). He investigated various facets of memory other than associative memory such as size of working memory and memory for text (auditory and visual), and concluded that tests measuring these memory constructs produce more impressive correlations with general achievement measures than mere rote memory tests (Skehan 1989). Skehan's findings, in turn, are in agreement with trends in research on memory structure in cognitive psychology which have emphasised the role of meaningful context rather than frequency of repetitions in laying down memory traits (Craik and Tulving 1975, cited in Anderson 1998). Similar conclusions have been reached by Szewczuk (1984) in his comprehensive treatment of the mechanisms of memory. In one of his experiments in which he studied spontaneous (accidental) versus intentional remembering of a list of unconnected words he found that, while there were differences between the two types of remembering, generally the remembering of decontextualised words was less efficient than the remembering of connected sentences, presented both aurally and visually. All this suggests that if memory abilities are important in predicting general SLP then it is different aspects of it that are involved. Alternatively, it might be the case that the importance of paired-associate memory is differentially linked to aspects of instructional methodologies: audio-lingual versus communicative, types of courses: intensive versus non-intensive, or types of learners: exceptional versus non-exceptional (cf. *Chapter 1*).

4.4.2. Summary of the discussion of RQ 2a

Summing up the foregoing discussion, it may be concluded that the present data support hypothesis 3 as the most feasible answer to RQ2a as formulated in *section 0.2.* The study demonstrated an insignificant role of non-verbal reasoning tests and paired-associate memory tests in the prediction of L2 general proficiency among Grade seven learners in non-intensive language course in lower secondary schools. Instead, four measures proved to be highly predictive of the overall proficiency accounting for as much as 58.6% of explained variance in the criterion test. The measures were: *AVoc* – a test of verbal intelligence; *GenVrbl* – lexis-based aptitude measure; *IndReas* – a test of linguistic analytical abilities and *TextComp* – a test of memory for text (written). The findings can thus be summarised as follows:

1. fluid intelligence is not a good predictor of L2 proficiency
2. crystallised intelligence (L1 vocabulary), verbal fluency, language analysis and memory for text tests are the most powerful predictors
3. general L2 proficiency is a multi-faceted domain as reflected in the number of variables contributing to its prediction
4. because a variety of skills/abilities make up L2 proficiency it is not feasible to expect one or two predictors only – their number reflects multidimensionality of the construct of SLP.

4.4.3. Discussion of RQ 2b.

In the second part of the prediction study multiple regression was further used in order to make separate analyses of best predictors against each of the two identified component variables of the KET-A Total test i.e., *SLPCompetence* and *SLPSkills*.

If different predictors turned out to be differentially linked with predicting different criterion variables then this fact could be interpreted in terms of different aspects of foreign language aptitude being involved in different aspects of criterion performance (Carroll 1981:101). The results of two multiple analyses of the same set of predictors against two different components of general SLP (cf. *section 3.9.3.*) are summarised below (cf. *Table 3.28.*):

- Three independent variables, out of six entered, make significant contributions to both predictions
- The only variable that does not contribute to either of the predictions is SRPM. The variable's BETA weights for *Competence* and *Skills* are minimal: 0.011 and 0.048 respectively.
- The only variable that enters both predictions is *AVoc* – a test of verbal intelligence. While the variable, with its BETA weight = 0.339, is the strongest predictor of the *Competence* dimension, it is the weakest predictor (BETA= 0.282) of the *Skills* sub-criterion.
- The remaining four variables enter the analyses pairwise in a mutually exclusive mode i.e., when *TextComp* and *AssocMem* enter *Competence* prediction *GenVrbl* and *IndReas* are left out of the equation. The reverse is true for *Skills*.
- There is a marked difference in the amount of explained variance between *SLPCompetence* (43.5%) and *SLPSkills* (63.7%).

These results led to the following observations:

1. The non-inclusion of SRPM suggests that the abilities tapped by the test are not implicated in the two more specific dimensions of second language proficiency. Non-

verbal reasoning skills seem not to be detrimental of L2 learning even if finer criteria than the general SLP are used

2. A very high BETA for *AVoc* and an insignificant BETA for *GenVrbl* in *Competence* regression analysis suggests that the two variables address different abilities, which makes them contribute differentially to the prediction of the knowledge component of SLP. Also, the fact that *AVoc*, but not *GenVrbl* contribute to the prediction of the *Competence* dimension of the general SLP can be interpreted in terms of the high demands passive knowledge of the 'what' puts on native vocabulary skills. It seems logical to assume that 'more verbal learners' profit more with respect to the declarative knowledge of language facts which are conveyed by native vocabulary in the process of L2 instruction (cf. point 3 in *section 4.4.1.*). On the other hand, a higher BETA value of *GenVrbl* (0.364) than of *AVoc* (0.282) for *Skills* prediction indicates that other lexical skills than those measured by the verbal intelligence test differentially contribute to the prediction of the component of second language proficiency (cf. *Table 2.2.* in *section 2.4.2.*). Apparently, an ability to perform successfully on listening, reading, writing and speaking tests involves more than just large lexicon as measured by verbal intelligence tests
3. Generally high predictive potential of *GenVrbl* (cf. this section and *section 4.4.1.*) could be held in favour of Sparks and Ganschow's (1991) Linguistic Coding Differences Hypothesis. In the hypothesis Sparks and Ganschow speculated that the learning of both native and foreign language is dependent on general language learning mechanisms and that problems with one component of language (eg., phonology/orthography) will have an effect on the learning of both languages (Sparks 2001). Additional empirical support for the premise that native language skills are important for learning a FL came from Dufva and Voeten's (1999) study reported in Sparks and Ganschow (2001). In their study of Finnish children (N=160) from first to third grades the authors found that native word recognition skill, reading comprehension and phonological memory forms the basis for learning an L2. It will be recalled that the *GenVrbl* variable is significantly loaded by tasks including two of the three skills mentioned: native word recognition (*zwolennik*, *dbrdszn*) and phonological memory (*dyktando*); and quite substantially, although non-significantly, loaded by reading comprehension tasks (*ofiara, handlarz, układ*).
4. The behaviour of the *AssocMem* variable in the prediction analysis of *Competence* shows a qualitatively marked difference from how the variable behaved in the prediction of *Skills* component in point 6 in *section 4.4.1.* The variable, despite its lowest correlation with the subcriterion (r= 0.23), enters the regression analysis and contributes uniquely to

the prediction of the *Competence* component of SLP. Its BETA weight (0.156) is only marginally lower than that of the second best predictor *TextComp* (0.195). This fact should not be puzzling in itself considering that memory-oriented instruction and focus on declarative knowledge are still dominant pedagogies in many a Polish schools. It is thus only natural to expect that learners' success on the lexical and grammar parts of the KET-A test should be differentially linked with the test of paired-associate memory for decontextualised language (Skehan 1989)

5. Linguistic reasoning abilities, captured by the *IndReas* variable also display an interesting prediction pattern. Namely, the variable, while not contributing to the prediction of *Competence*, enters the regression equation for the *Skills* component of general SLP. This is somewhat surprising as one would expect the 'knowledge' dimension of language ability measured by rather traditional grammar and vocabulary tests to call for analytical language abilities. This, however, is not the case and the prediction is effected by an interplay of verbal and memory aspects (paired-associate memory and memory for text). The appearance of the inductive rule acquisition variable (*IndReas*) in the prediction of the *Skills* component (BETA= 0.351) is also intriguing. In this connection a study by Esser and Kossling (1986) is particularly interesting. The authors studied a group (N=188) of L2 adult learners (average age 21.2) learning English and Russian. They investigated the relationship between SLP and aptitude. To this end the authors developed a new FL aptitude test which minimalised the interfering influence of language in instructions and which was entirely based on non-verbal tasks. The test addressed the same, well-known, four-components of aptitude posited by Carroll (1962) i.e., paired-associate learning; inductive rule acquisition; semantic integration and analogy formation but with no mediating involvement of language. It was the authors' belief that "the essential cognitive prerequisites for FL learning must be sought at a level that is 'deeper' and 'further away' from language" (Esser and Kossling 1986: 95). The results obtained from the test were correlated with the learners' achievement measures (grades and vocabulary/grammar test results) and the authors concluded that:
   - inductive rule acquisition (artificial language task) and analogy formation correlated with grammar (0.79; 0.72) but not with vocabulary (0.36; 0.25)
   - paired-associate and semantic integration correlated with vocabulary at 0.82 and 0.71 but not with grammar (0.23 and 0.38)
   - there was a significant correlation between inductive rule acquisition and the speaking grade (0.66)

The last finding by Esser and Kossling seems to point toward a markedly nonaccidental relationship between inductive reasoning and the skills component of SLP.

6. The importance of the vocabulary and linguistic analysis component of FL aptitude has also been noted by Guthke and Harnisch (1986, cited in Klauer 1993). Discussing their research Klauer states that: "The discovery of syntactic rules on the one hand, and the acquisition of words' meanings on the other, are of crucial importance to L2 acquisition at least at the beginners level" (1995:2). The results of the present study show that those abilities are even more essential when teaching young teenagers in non-intensive courses using traditional, visually and teacher-oriented instruction.
7. Once again it is to be noted that if simple correlational relationships between the independent variables and the criteria are considered as the only indication of the predictive potential of the variables then in the case of both *Competence* and *Skills* an important generalisation is missed. If limited to the correlation matrix in *Table 3.28.* the prediction would identify the same set of variables for both components: *AVoc*, *GenVrbl* and *SRPM*. This would leave the researcher in the dark as to the differential aspect of the relationship between second language competence and the abilities underlying it.

4.4.4. Summary of the discussion of RQ2b

Summing up the discussion, it may be concluded that multiple regression analysis made against two isolated components of second language proficiency as measured by the KET-A test again lends support to the 'hybrid' hypothesis 3 in response to RQ2b as formulated in *section 0.2.* An interplay of foreign aptitude and a verbal intelligence variable, to a noteworthy exclusion of a non-verbal test of fluid intelligence, were found to be related to L2 learning success. The analysis has produced two different prediction models in terms of the combination of the independent variables. In the case of *SLPCompetence*, the best results in inferring the scores on vocabulary and grammar tests are achieved when prediction is based on an active component of a test of verbal intelligence (*AVoc*) combined with memory for text aptitude subtest (*TextComp*) and associative memory for word-pairs (*AssocMem*). On the other hand, *SLPSkills* is best predicted using the following set of independent variables: *GenVrbl*, *IndReas* and *AVoc*. There is a substantial difference in the amount of variance we can hope to explain using each of the predictor sets: 43.2% in the case of *SLPCompetence* and 63.7% in the case of *SLPSkills*. The findings can be briefly summarised as follows:

1. two different combinations of predictors are differentially linked with the type of the subcriterion predicted

2. this allows for the link between the predictors and the predicted performance to be established in a non-accidental way
3. for each criterion a different best predictor can be identified: *Avoc* for *Competence*; *GenVrbl* for *Skills*
4. fluid intelligence is not a good predictor of either of the two L2 proficiency components
5. the explanation of the differential prediction may be linked to aspects of instruction, quality of teaching or individual learner differences
6. the prediction of even finer aspects of SLP can be envisaged.

NOTES

1. This is an intended reference to the title of Oller's 1981 paper.
2. The emergence of the "rote factor" in general abilities domain might be an artefact of the way the sampling of abilities included in the test was done (cf. *section 0.3.*). It is a well-known fact in factor analysis that an introduction of the so called 'marker variables' in a battery will yield a factor with invariably strong loading on the factor. As Cattell explains:

   > *When investigating domain 'A', deliberately introduce several 'irrelevant' variables you have reason to believe have a 'non-A' nature. (...this) will yield factors standing out clearly and largely orthogonal to the factor or factors in the domain you are interested in* (1971:112).

3. The results for SLP Speaking variable were obtained from teachers' assessments of pupils in the course of actual lessons and as such might be biased and unobjective.
4. cf. the discussion in *Chapter 1.*

CONCLUSION

The present study investigated the relationship between foreign language aptitude, intelligence and second language proficiency among 227 Grade 7 pupils (13-year-olds) learning a foreign language (English: N=214; German: N=13) in a non-intensive course in lower secondary schools. The main aim of the study was to find an answer to the question formulated in the title of the study i.e., whether the concept of second language aptitude is a valid factor on the basis of which the prediction of L2 learning success in the population studied can be envisaged. Four research questions followed from this general aim and were discussed in *Chapter 4*. This section summarises the main points of the discussion (*section I*), outlines the implications of the study (*section II*) and gives some suggestions for further research (*section III*).

I. Major findings

The major findings of the present research can be summarised as follows:

1. The results of multiple regression analyses point to the fact that neither intelligence nor aptitude variables alone produce best prediction models of SLP in the sample studied. In all three prediction models discussed in *sections 3.9.2.* and *3.9.3.* a combination of aptitude variables with a verbal intelligence variable yield satisfactory levels of explained variance in the criterion against which the analysis was made. Using a limited set of identified predictors we can hope to predict the following substantial proportions of the variance in the criteria variables in the population of 13-year-old Polish learners of foreign languages in lower secondary schools:
    - 58.6% of the *KET-A Total* – general second language proficiency
    - 63.7% of the *SLPSkills* – a component of general SLP
    - 43.2% of the *SLPCompetence* – a component of general SLP
2. Second language proficiency is clearly independent from both intelligence and second language aptitude. Intelligence and aptitude emerged as distinct but related constructs showing some degree of overlap in the ability domains covered (cf. *Table 3.24*). Their separateness can be seen in their differential predictive potential (cf. *sections 3.9.2.* and *3.9.3.*) of the criterion variables. Their relatedness, on the other hand, was demonstrated by the analysis of the underlying trait structure of cognitive abilities carried out in *section 3.8.*
3. The following factors seem to be central to L2 learning success:
    - size of native vocabulary (*AVoc*)

- an ability to process contextualised and decontextualised native vocabulary using all three codes of the language: phonological-orthographic (*zwolennik*, *dbrdszn*); morphological and syntactic (*sprawcy*, *pojezierze*) in both recognition and production modes (*GenVrbl*)
- an ability to generalise from actual language i.e., language inductive abilities (*IndReas*)
- an ability to understand, remember and recall text (visual and, to a lesser degree, aural) (*TextComp*)

The importance of these abilities has been shown by the emergence of strong factor loadings in relevant analysis discussed in *Chapter 4* and further, by their high BETA weights.

4. Non-verbal reasoning abilities (*SRPM*) are not crucial to L2 acquisition. However, the same abilities tested on language material contribute differentially to success in L2 learning as shown by BETA weights for *IndReas* in the prediction model of *SLPGeneral* (KET-A) and *SLPSkills* (cf. *section 3.9.2.*; *3.9.3.* and *4.4.1.; 4.4.3.*).

II. Implications of the study

There are several important implications that the present study may have for the field of applied linguistics. They can be considered from methodological, theoretical and practical perspectives.

Methodological Implications

The present research, by applying rigid procedures of scholarly investigations during pre-, while- and post-data collection stages, and by providing first-hand, hard data in the domain studied, contributes to the field of applied linguistics in its methodological implications with respect to the sample selection, testing tools, and statistical techniques used in data analysis. The complexity of the relationships between the independent variables used in the study has not been addressed so far in the context of the Polish language and the sample of the learners studied. Also with respect to the statistical techniques used in the analysis of the data gathered, the study constitutes a definite advancement in research on foreign language aptitude. Two powerful multivariate correlational techniques (factor analysis and stepwise multiple regression) were used and their advantage in the analyses seeking better understanding of causal relationships between the variables studied was demonstrated. No earlier application of the techniques in this kind of research has been attested in Poland.

The study has once again stressed the importance of theory and continuity of research in any scholarly endeavour. Accumulation of data and research in a relatively new domain of scholarly interest cannot be neglected by anyone wishing to pursue it, lest one risks stating the obvious. In the context of the present study it was forcefully demonstrated how over-reliance on intuition combined with neglect of the current state of research on FL aptitude led to misguided efforts in aptitude test design.

Within a more global perspective an important and, mainly methodological implication of this fact is that the standards and the quality of research in applied linguistics dealing with any sort of measurement rest on the validity and reliability of the measuring instruments used. When untried and unreliable scales are used whose construct validity has not been determined, little or no value can be attached to the generalisations made on the basis of the data obtained.

Theoretical implications

The results of the present study have some important theoretical implications as well. Several of those follow from the breadth of the problem dealt with. As mentioned in *Chapters 1* and *2*, very few studies abroad, and no study in Poland, have been known to investigate the interplay between aptitude, intelligence and second language proficiency. Consequently, this study which addresses the problem of the structure of FL aptitude and then, its involvement in the prediction of L2 learning success, contributes to the field of applied linguistics in Poland and, by providing evidence from an 'unresearched' language, to the field of applied linguistics in general.

Another theoretical implication relates to a new testing paradigm introduced in a few tasks of one of the study's testing tools (PLAT) and discussed in more detail in *Chapters 2* and *3*. The, so called, 'dynamic testing' was shown by Grigorenko et al. (2000) to be a promising paradigm in testing the potential rather than already formed abilities to learn. Their model of language aptitude, linking the dynamic testing paradigm (cf. *section 2.4.2.*) with acquisition processes, "affords the possibility of achieving a closer match between specific aptitudes and specific psycholinguistic processes" (Ellis 2004:533). Despite the fact that the CANAL-F Test does not perform outstandingly better as compared with the classics in aptitude testing (MLAT and PLAB), it is the first attempt at designing a test of language learning potential (with the exception of Skehan's model of aptitude which, however was not operationalised into a test) based on a new theory of L2 acquisition.

Practical Implications

The issue of special importance here is the very practical need for theoretically valid and relevant tests of second language aptitude available commercially and having very stable psychometric properties. This, in turn, calls for systematic and orchestrated efforts which would result in building up professional expertise in the field of cognitive individual differences among L2 learners. As things stand now with FL aptitude testing, it seems that this particular area of individual differences studies has become the domain of popular science and pseudo-scholarly interest where claims regarding the measurements of posited abilities are made on the basis of intuition, misguided hunches and 'wild guesses' rather than on available theories and/or research data. As long as invalid and unreliable FL 'aptitude tests' are used in educational settings, one cannot seriously speak of any practical implications of the field to second language learning/teaching contexts.

There are several important practical implications which could follow from FL aptitude research and which can be seen to directly relate to various aspects of second language education in virtually all possible contexts (formal/informal; younger learners/adults; general/ specific purposes; intensive/non-intensive and so on).

One aspect in which L2 learning might profit from aptitude research/testing is an increased understanding of individual differences that L2 learners bring into the teaching/learning situation. This is afforded by the componential structure of abilities consistently shown to differentially contribute to L2 learning success in almost any research done so far (including this one). The non-unitary nature of language learning skills allows for alternative routes to be taken by a learner who can capitalise on the stronger aspects of her/his FL aptitude and minimise the effects of her/his weaker sides.

Knowledge of learner cognitive profiles could be an invaluable help, not only for the learner her/himself, but also for an informed FL teacher. Drawing on this information teachers can make their teaching more effective in every day school practice to the extent they have never been able to do before. However, without adequate (valid and reliable) testing tools which would enable them to find out about their learners' potential to learn/acquire a second language the prospects of this happening in the near future are as poor as they have been so far.

III. Suggestions for further research

The present study, in attempting to address some of the problems connected with FL aptitude has, in effect, led to even more questions than it set out to answer. Only a few of the questions which have appeared as a result of the study will be mentioned here. While the majority of the suggestions made are applicable only to the Polish context some will have a universal appeal.

Firstly, more research is needed on Polish native language predictors of L2 proficiency in various educational settings with respect to the age of learners, instruction methodologies used and the intensity of teaching.

Secondly, differential studies on prediction of finer dimensions of second language proficiency are in order to enable the researchers to establish, on the one hand, the most effective predictors of individual language components and, on the other, the nature of the construct of FL aptitude itself.

Thirdly, the fact that Carroll's conditions for ideal prediction are hard to obtain due to early onset of L2 learning in the Polish educational system, should lead to research on FL aptitude in young learners prior to their L2 learning.

Fourthly, the research on the structure and nature of language aptitude, irrespective of the lines along which it will proceed, needs to employ more advanced and powerful statistical tools than simple correlational techniques. Specifically, multi-trait multi-method designs need to be employed to disentangle the interrelationships between the researched constructs and the methods used; confirmatory, rather than exploratory, factor analysis and structural equation modelling are needed to test prediction models against the data.

Fifthly, research on differential L2 learning success must include not only cognitive variables (FL aptitude being one of them) but also personality and affective variables. In this respect, a promising line of research is Carroll's (1962/1965) Model of School Learning. Also, qualitative approach should be used side by side with a quantitative one in order to overcome the deficiencies of the purely product-oriented study.

Finally, research should start on the adaptation onto the Polish grounds of already existing aptitude batteries such as Carroll and Sapon (1959) MLAT or Meara et al. (2001) LAT. This, in turn, should lead to the accumulation of expertise in the field and should be accompanied by parallel efforts at designing, constructing, standardising and validating national tests of FL aptitude.

BIBLIOGRAPHY

Alderson, J. C., Clapham, C. and Wall, D. 1995. *Language test construction and evaluation*. Cambridge: CUP

Anderson, J. R. 1998. *Uczenie się i pamięć. Integracja zagadnień*. Warszawa: WSiP.

Bachman, L. F. 1982. "The trait structure of cloze test scores". *TESOL Quarterly 16*. 61-70

Bachman, L. F. 1985. "Performance on cloze tests with fixed ratio and rational deletions". *TESOL Quarterly 19*. 535-556.

Bachman, L. F. 1989. "Language testing-SLA research interfaces". *Annual Review of Applied Linguistics 9*. 193-209.

Bachman, L. F. 1997. *Fundamental considerations in language testing.* Oxford: OUP.

Bachman, L. F. and Palmer, A. S. 1981. "The construct validation of the FSI oral interview". *Language Learning 31/1*. 67-86.

Bachman, L. F. and Palmer, A. S. 1982. "The construct validation of some components of communicative proficiency". *TESOL Quarterly 16*. 449-465.

Bachman, L. F. and Palmer, A. S. 1983. "The construct validity of the FSI oral nterview". In Oller, J. W. (ed.). 1983. 3-10.

Bogaards, P. 1991. *Aptitude et affectivite dans l'apprentissage des languaes etrangeres*. Paris: Les Editions Didier.

Boyle, J. P. 1987. "Intelligence, reasoning, and language proficiency". *The Modern Language Journal 71/3*. 277-288.

Brandon, E. P. 1985. "Aptitude analysed". *Educational Philosophy and Theory 17.* 13-18.

Brown, H. 1973. "Affective variables in second language acquisition". *Language Learning 23/3*. 231-244.

Brown, H. D., Yorio, C. and Crymes, R. (eds). 1977. *On TESOL '77. Teaching and learning English as a second language: trends in research and practice.* Washington, DC: TESOL.

Brown, J. D. 1988 a. "What makes a cloze item difficult?" *University of Hawaii Working Papers in ESL 7.* 17-39.

Brown, J. D. 1988 b. *Understanding research in second language learning*. Cambridge: Cambridge University Press.

Brzeziński, J. 1984. "Badanie testu psychometrycznego metodą analizy wariancji". In Brzeziński, J. (ed.). 1984. 9-45.

Brzeziński, J. 1984. *Wybrane zagadnienia z psychometrii i diagnostyki psychologicznej.* Poznań: Wydawnictwo Naukowe UAM.

Brzeziński, J. 2000. *Badania eksperymentalne w psychologii i pedagogice.* Warszawa: Wydawnictwo Naukowe Scholar.

Brzeziński, J. 2002. *Metodologia badań psychologicznych*. Warszawa: PWN.

Cambridge Key English Test 2 with answers. Examination papers from the University of Cambridge Local Examination Syndicate. Past Papers. (2001). Cambridge: CUP.

Carroll, J. B. 1958. "A factor analysis of two foreign language aptitude batteries". *Journal of General Psychology 59*. 3-19

Carroll, J. B. 1962. "The prediction of success in intensive foreign language training". In Glaser, R. (ed.). 1962. 87-136.

Carroll, J. B. 1967. *Modern Languages Aptitude Test – Elementary*. New York: The Psychological Corporation.

Carroll, J. B. 1968. "The psychology of language testing". In Davies, A. (ed.). 1968. 46-59.

Carroll, J. B. 1971. "Implication of aptitude test research and psycholinguistic theory for foreign language teaching". Paper presented at the XVIIth International Congress, International Association of Applied Psychology; Liege (ERIC Document Reproduction Service No. ED 057 071)

Carroll, J. B. 1973. "Implications of aptitude test research and psycholinguistic theory for foreign language teaching". *International Journal of Psycholinguistics 2*. 5-14

Carroll, J. B. 1974. "The aptitude-achievement distinction: The case of foreign language aptitude and proficiency". In Green, D. R. (ed.). 1974. 286-303.

Carroll, J. B. 1979. "Psychometric approaches to the study of language abilities". In Fillmore, C. J., Kempler, D., and Wang, W-S.Y. (eds). 1979. 65-88.

Carroll, J. B. 1981. "Twenty-five years of research on foreign language aptitude". In Diller, K. C. (ed.). 1981. 83-118.

Carroll, J. B. 1983. "Psychometric theory and language testing". In Oller J.W. (ed.). 1983. 74-102.

Carroll, J. B. 1987. "Review of Klein-Braley, C. and Raatz, E., (eds). 1985. C-Tests in der Praxis". In *Fremdsprachen und Hochschule, AKS-Rundbrief 13/14*. Bochum: Arbeitskreis Sprachenzentrum (AKS). *Language Testing 4.* 99-106.

Carroll, J. B. 1990 "Cognitive abilities in foreign language aptitude: then and now". In Parry, T. S. and Stansfieid, C.W. (eds). 1990. 11 –29.

Carroll, J.B. 1993. *Human cognitive abilities: a survey of factor-analytic studies*. Cambridge: CUP.

Carroll, J. B. and Sapon, S. M. 1959. *Modern Language Aptitude Test*. New York: The Psychological Corporation.

Cattell, R. B. 1940. "A culture free intelligence test". *Journal of Educational Psychology 31*. 161-180.

Cattell, R. B. 1971. *Abilities: their structure, growth and action*. Boston: Houghton Mifflin.

Cattell, R. B., and Cattell, H. E. P. 1973. *Measuring intelligence with the Culture Fair Tests.* Champaign, IL: Institute for Personality and Ability Testing.

Chapelle, J. and Claes, M. (eds). 1993. *Memory and memorisation in acquiring and learning languages.* Louvain-la-Neuve: C.L.L.

Chapelle, C. and Roberts, C. 1986. "Ambiguity tolerance and field independence as predictors of proficiency in English as a second language". *Language Learning 36/1*. 27-45.

Chapelle, C. and Green, P. 1992. "Field independence/dependence in second language acquisition research". *Language Learning 42/1*. 47-83.

Chastain, K. 1975. "Affective and ability factors in second-language acquisition". *Language Learning 25/2*. 153-161.

Cohen, L. and Manion, L. 1995. *Research methods in education*. London and New York: Routledge.

Court, J. H. and Raven, J. 1995. *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Oxford: Oxford Psychologists Press Ltd.

Cook, V. J. (ed.). 1986. *Experimental approaches to second language acquisition*. Oxford: Pergamon.

Davies, A. (ed.). 1968. *Language testing symposium: a psycholinguistic approach.* London: Oxford University Press.

Davies, A. 2003. "Three heresies of language testing research". *Language Testing 20/4*. 355-368.

Davies, A. and Elder, C. (eds.). 2004. *The handbook of applied linguistics*. Malden: Blackwell Publishing Ltd.

Dechert, H. (ed.). 1990. *Current trends in European second language acquisition research.* Clevedon, England: Multilingual Matters.

Diller, K. C. (ed.). 1981. *Individual differences and universals in language learning aptitude.* Newbury House: Rowley, Mass.

Droździał-Szelest, K. 1997. *Language learning strategies in the process of acquiring a foreign language*. Poznań: Motivex.

Ehrman, M. E. 1996. *Understanding second language learning difficulties.* Thousand Oaks, CA.: Sage.

Ehrman, M. E., and Oxford, R. L. 1995. "Cognition plus: correlates of language learning success". *Modern Language Journal 79/1*. 67-89.

Ellis, R. 1994. *The study of second language acquisition*. Oxford: OUP

Ellis, R. 2004. “Individual differences in second language learning”. In Davies, A. and Elder, C. (eds). 2004. 525-551.

Esser, U. and Kossling, B. 1986. "A general psychological approach to the diagnosis of foreign language aptitude". In Cook, V. 1986. 95-100.

Figarski, W. 1984. *Wyznaczniki powodzenia w szkolnej nauce jezyka rosyjskiego*. Warszawa: Wydawnictwa Szkolne i Pedagogiczne.

Fillmore, C. J., Kempler, D., and Wang, W-S. Y. (eds). 1979. *Individual differences in language ability and language behaviour*. New York: Academic Press.

Flahive, D. E. 1980. "Separating the *g* factor from reading comprehension". In Oller, J. W. and Perkins, K. (eds). 1980. 34-46.

Ganschow, L., Sparks, R. L., and Javorsky, J. 1998. "Foreign language learning difficulties: an historical perspective". *Journal of Learning Disabilities 31/3.* 248-258.

Gardner, R. C. 1985. *The role of attitude and motivation: social psychology and SLL*. London: Edward Arnold.

Gardner, R. C. 1990. "Attitudes, motivation and personality as predictors of success in foreign language learning". In Parry, T. S. and Stansfieid, C. W. (eds). 1990. 179-221.

Gardner, H., Kornhaber, M. L., and Wake, W. K. 2001. *Inteligencja: wielorakie perspektywy.* Warszawa: WSiP S.A.

Gardner, R. C., and Lambert, W. E. 1965. "Language aptitude, intelligence and second-language achievement". *Journal of Educational Psychology 56.* 191-199.

Gardner, R. C., and Lambert, W. E. 1972. *Attitudes and motivation in second-language learning.* Rowley, Mass.: Newbury House.

Gaas, S., Madden, C., Preston, D., and Selinker, L. (eds). 1989. *Variation in second language acquisition. Volume II: Psychological issues.* Clevedon, Avon: Multilingual Matters.

Genesee, F. 1976. "The role of intelligence in second language learning". *Language Learning 26/3.* 267-280.

Genesee, F. 1987. *Learning through two languages. Studies of immersion and bilingual education*. Rowley, Mass.: Newbury House.

Glaser, R. (ed.). 1962. *Training research and education.* Pittsburgh: University of Pittsburgh Press.

Green, D. R. (ed.). 1974. *The aptitude-achievement distinction*. Monterey, CA: CTB/McGraw-Hill.

Grigorenko, E. L., and Sternberg, R. J. 1998. "Dynamic testing". *Psychological Bulletin 124/1.* 75-111.

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". *Modern Language Journal 84/3*. 390-405.

Guilford, J. P. 1960. *Podstawowe metody statystyczne w psychologii i pedagogice.* Warszawa: PWN.

Hatch, E. and Farhady. H. 1982. *Research design and statistics for applied linguistics.* Rowley: Newbury House Publishers.

Harley, B., and Hart, D. 1997. "Language aptitude and second language proficiency in classroom learners of different starting ages". *Studies in Second Language Acquisition 19/3*. 379-400.

Harrington, M., and Sawyer, M. 1992. "L2 working memory capacity and L2 reading skill". *Studies in Second Language Acquisition 14/1*. 25-38.

Henmon, V. A. C. (ed.). 1929. *Prognosis tests in the modern foreign languages.* New York: Macmillan.

Hermann, G. 1980. "Attitudes and success in children's learning of English as a second language: the motivational versus the resultative hypothesis". *English Language Teaching Journal 34/3*. 247-254.

Heyde, A. 1977. "The relationship between self-esteem and the oral production of a second language". In Brown, H. D. et al. 1977. 86-108.

Hoffmann, B. 1962. *The tyranny of testing*. New York: Collier Books.

Hornowski, B. 1970. *Analiza psychologiczna skali J. C. Ravena*. Warszawa: PWN.

Hornowska, E. 2003. *Testy psychologiczne. Teoria i praktyka*. Warszawa: Wydawnictwo Naukowe Scholar.

Horwitz, E., Horwitz, M., and Cope, J. 1986. "Foreign language classroom anxiety". *Modern Language Journal 70/2*. 125-132.

Ioup, G. 1989. "Immigrant children who have failed to acquire native English." In Gass et al. (eds). 1989. 111-145.

Irvine, P., Atai, P., and Oller, J. W. 1974. "Cloze, dictation, and the test of English as a foreign language". *Language Learning 24/3*. 245-252.

Jafarpur, A. 1995. "Is C-testing superior to cloze?" *Language Testing 12/2.* 194-216.

Jankojć, Z. E. 2002. "Indywidualne uwarunkowania sukcesu studenta w nauce języka angielskiego". Unpublished doctoral dissertation. Uniwersytet Szczeciński.

Jaworowska, A. and Szustrowa, T. 2000. *Test Matryc Ravena w Wersji Standard. Formy: Klasyczna, Równoległa, Plus*. Warszawa: Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.

Jurkowski, A. 1997. *Test językowy LEKSYKON*. Warszawa: Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.

Kaiser, H. F. 1960. "The application of electronic computers to factor analysis". *Educational and Psychological Measurement 20/1*. 141-151.

Kaulfers, W. V. 1931. "Present state of prognosis in foreign languages". *School and Society 39/4.* 585-596.

Kaulfers, W. V. 1939. "Prognosis and its alternatives in relation to the guidance of students". *German Quarterly 12/1.* 81-84.

Klauer, K. J. 1993. "Über die Auswirkungen eines Trainings zum induktiven Denken auf zentrale Komponenten der Fremdsprachenlernfäigkeit". *Zeitschrift für Pedagogische Psychologie 7*. 1-9.

Komorowska, H. 1984. *Testy w nauczaniu języków obcych*. Warszawa: Wydawnictwa Szkolne i Pedagogiczne.

Komorowska, H. 1987. *Sukces i niepowodzenie w nauce języka obcego*. Warszawa: Wydawnictwa Szkolne i Pedagogiczne.

Krashen, S. D. 1981a. *Second language acquisition and second language learning.* Oxford: Oxford University Press.

Krashen, S. D. 1981b. "Aptitude and attitude in relation to second language acquisition and learning". In Diller, K. C. (ed.). 1981. 155-175.

Kuliniak, R. 2002. *Testy predyspozycji Językowych.* Wałbrzych: Wydawnictwo BIMART.

Lado, R. 1961. *Language testing. The construction and use of foreign language tests.* London: Longman.

Lado, R. 1964. *Language teaching. A scientific approach*. New York: McGraw Hill.

Lett, J. A., and O'Mara, F. E. 1990. "Predictors of success in an intensive foreign language learning context: Correlates of language learning at the Defence Language Institute Foreign Language Center". In Parry, T. and Stansfieid, C. W. (eds). 1990. 222-260.

Lewowicki, T. 1975. *Psychologiczne różnice indywidualne a osiągnięcia uczniów.* Warszawa: Wydawnictwa Szkolne i Pedagogiczne.

MacIntyre, P. and Charos, C. 1996. "Personality, attitudes, and affect as predictors of second language communication". *Journal of Language and Social Psychology 15/1*. 3-26.

McLaughlin, B. 1995 "Aptitude from an information-processing perspective". *Language Testing 12/3.* 370-387.

McNamara, T. 2000. *Language testing.* Oxford: Oxford University Press.

Meara, P., Milton, J. and Lorenzo-Duz, N. 2001. *Language Aptitude Test.* Newbury: Express Publishing.

Meara, P. (ed.). 1986. *Spoken language*. London: Centre for Information on Language Teaching.

Mietzel, G. 2002. *Psychologia kształcenia*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.

Milanovic, M. (ed.). 1998. *Multilingual glossary of language testing terms. Studies in language testing 6.* Cambridge: CUP.

Murakami, K. 1974. "A language aptitude test for the Japanese". *System 2*. l-47

Nelson, R. and Jakobovits, L. A. 1970. "Motivation in foreign language learning". In Tursi, J. (ed.). 1970. 125-147.

Neufeld, G. 1979. "Towards a theory of language learning aptitude". *Language Learning 29/3.* 227-241.

Nęcka, E. 2003. *Inteligencja: geneza, struktura, funkcje*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.

Niemierko, B., 1999. *Pomiar wyników kształcenia*. Warszawa: WSiP.

Niżegorodcew, A. 1975. "Przydatność tzw. testu prognostycznego jako metody pomiaru zdolnosci do uczenia się języka obcego u młodzieży licealnej". *Języki Obce w Szkole 3/4*. 299-303.

Niżegorodcew, A. 1979. "Rola specjalnych uzdolnień w nauce języków obcych w szkole". *Przegląd Glottodydaktyczny 3/1*. 31-41.

Niżegorodcew, A. 1980. "Zdolności do nauki języka obcego w świetle badań testowych". *Zeszyty naukowe Uniwersytetu Jagielońskiego. Prace psychologiczno-pedagogiczne. Zeszyt 31.* 137-156.

Oller, J. W. 1978. "How important is language proficiency to IQ and other educational tests?" In Oller, J. W. and Perkins, K. (eds). 1978. 1-16.

Oller, J. W. 1979. *Language tests at school*. London: Longman.

Oller, J. W. 1981. "Language as intelligence?". *Language Learning 31/4.* 465-492.

Oller, J. W. (ed.). 1983. *Issues in language testing research.* Rowley, Mass.: Newbury House.

Oller, J. W. and Perkins, K. (eds). 1978. *Language in education: testing the tests.* Rowley Mass.: Newbury House.

Oller, J. W. and Hinofotis, F. B. 1980. "Two mutually exclusive hypotheses about second language ability: factor analytic studies of a variety of language subtests." In Oller, J. W. and Perkins, K. (eds.). 1980. 13-23.

Oller, J. W. and Perkins, K. (eds.). 1980. *Research in language testing.* Rowley, Mass.: Newbury House.

Oller, J. W., Richards, J. C. (eds.). 1973. "*Focus on the learner: pragmatic perspectives for the language teacher."* Rowley, Mass.: Newbury House.

Osaka, M., and Osaka, N. 1992. "Language-independent working memory as measured by Japanese and English reading span tests". *Bulletin of the Psychonomic Society 30/3.* 287-289.

Osaka, M., Osaka, N., and Groner, R. 1993. "Language-independent working memory: evidence from German and French span tests". *Bulletin of the Psychonomic Society 31/2.* 117-118.

Oxford, R. and Ehrman, M. 1989. "Psychological type and adult language learning strategies: a pilot study". *Journal of Psychological Type 16/1*. 22-32.

Ozga, J. and Tabakowska, E. 1973. "Egzamin wstępny z języka angielskiego dla kandydatów na filologię angielską Uniwersytetu Jagielońskiego w Krakowie". *Języki Obce w Szkole 1.* 46-57.

Parry, T. and Stansfield, C. W. (eds.). 1990. *Language aptitude reconsidered.* Englewood Cliffs, New Jersey: Prentice Hall.

Parry, T. S., and Child, J. R. 1990. "Preliminary investigation of the relationship between *VORD, MLAT,* and language proficiency". In Parry, T. and Stansfieid, C. W. (eds). 1990. 30-66.

Pedhazur, E. J. 1982. *Multiple regression in behavioral research: explanation and prediction.* New York: Holt, Rinehart and Winston.

Pearsall, J. (ed.). 1998. *The new Oxford dictionary of English*. Oxford: Clarendon Press.

Pellegrino, J. W. and Glaser, R. "Components of inductive reasoning". In Snow, R. E. et. al. 1995. 126-42.

Petersen, C. R., and Al-Haik, A. R. 1976. "The development of the Defence Language Aptitude Battery (DLAB)". Educational and Psychological Measurement 6. 369-380.

Pimsleur, P. 1962. "Predicting achievement in foreign language learning". *International Journal of American Linguistics 2/3.* 238-249.

Pimsleur, P. 1966. *The Pimsleur Language Aptitude Battery and Manual.* New York: Harcourt Brace Jovanovich.

Pimsleur, P. 1968. "Language aptitude testing". In Davies, A. (ed.). 1968. 98-106.

Politzer, R.L and Weiss, L. 1969. "An experiment in improving achievement in foreign language learning through learning of selected skills associated with language aptitude". Stanford, Cal.: Stanford University, (ERIC Document Reproduction Service, ED 046261).

Prejbisz, A. 1973. "Warunki wpływające na wyniki nauczania języków obcych w szkole." *Języki Obce w Szkole 1.* 1-8.

Purpura, J. E. 1999. *Learner strategy use and performance on language tests: A structural equation modelling approach.* Cambridge: CUP

Raatz, U. and Klein-Braley, C. 1981. "The C-test – a modification of the cloze procedure". In

Culhane, T., Klein-Braley, C. and Stevenson, D. K., (eds.). *Practice and problems in language testing. University of Essex Occasional Papers 26*. Colchester: University of Essex.

Raven, J., Raven, J.C., Court, J. H. 1988. *Raven manual: section 3. Standard Progressive Matrices*. Oxford: Oxford Psychologists Press.

Reber, A. S. 1985. *The Penguin dictionary of psychology.* London: Penguin Books.

Reber, A. S. 1989. "Implicit learning and tacit knowledge". *Journal of Experimental Psychology: General 188/3.* 219-235.

Reber, A. S. 1993. *Implicit learning and tacit knowledge.* Oxford, England: Clarendon Press.

Reynolds, B. 1999 "Phonetic coding in language aptitude: Different roles for different languages". In Yamashiro, A.D. (ed.). http://www.tuj.ac.jp/tesol/press/papers0014/renyolds.html

Rietveld, T. and Hout van, R. 1993. *Statistical techniques for the study of language and language behaviour*. Berlin: Mouton de Gruyter.

Robinson, P. 1995 "Aptitude, awareness and the fundamental similarity of implicit and explicit second language learning". In Schmidt, R. (ed.). 1995. 303-358.

Robinson, P. 1997. "Individual differences and the fundamental similarity of implicit and explicit adult second language learning". *Language Learning 47.* 45-99.

Robinson, P. and Cornwell, S. (eds.) 2000. *Individual differences in foreign language learning: Effects of aptitude, intelligence and motivation*. Tokyo: Aoyama Gakuin University.

Rubin, J. 1975. "What the 'good language learner' can teach us?" *TESOL Quarterly 9/1.* 41-51.

Rysiewicz, J. 2003. *A few remarks on language aptitude testing*. An unpublished paper presented at 12th Annual Polish Association of Schools of English Conference in Miedzeszyn. April 2003.

Sarig, G. 1987 "Comments on Skehan: The role of foreign language aptitude in a model of school learning". *Language Testing 4/1*. 51 –54.

Sasaki, M. 1996. *Second language proficiency, foreign language aptitude and intelligence. Quantitative and qualitative analyses*. New York: Peter Lang.

Schmid, J. and Leiman, J. M. 1957. "The development of hierarchical factor solutions." *Psychometrika 22/1.* 53-61.

Schmidt, R. (ed.). 1995. *Attention and awareness in foreign language learning*. University of Hawaii at Manoa, Second Language Teaching and Curriculum Center, Teaching Report No 9; University of Hawaii Press.

Seashore, R. H. 1951. "Work and motor performance". In Stevens, S. S. (ed.). 1951. 1342-1365.

Seliger, H. W. and Shohamy, E. 1990. *Second language research methods.* Oxford: OUP.

Service, E. 1992. "Phonology, working memory and foreign language learning". *Quarterly Journal of Experimental Psychology 45/1*. 21-50.

Service, E. 1993 "Phonological and semantic aspects of memory for foreign language". In Chapelle, J. and Claes, M. (eds.). 1993. 167-198.

Service, E. and Kohonen, V. 1995. "Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition?". *Applied Psycholinguistics 16/2*. 155-172.

Silva, J. M., and White, L. A. 1993. "Relation of cognitive aptitudes to success in foreign language training". *Military Psychology 5/1.* 79-93.

Skehan, P. 1980. "Memory, language aptitude, and second language performance". *Polyglot 2*: Fiche 3.

Skehan, P. 1986a."The role of foreign language aptitude in a model of school learning". *Language Testing 3/2*. 188-221.

Skehan, P. 1986b. "Where does language aptitude come from?" In Meara, P. (ed.). 1986. 34-58.

Skehan, P. 1986c. "Cluster analysis and the identification of learner types". In Cook, V. J. (ed.). 1986. 47-75.

Skehan, P. 1989. *Individual differences in language learning*. Edward Arnold; London

Skehan, P. 1990. "The relationship between native and foreign language learning ability: Educational and linguistic factors". In Dechert, H. (ed.). 1990. 83-106.

Skehan, P. 1991. "Individual differences in second language learning". *Studies in Second Language Acquisition 13/3.* 275-98.

Skehan, P. 1992. "Foreign language learning ability: cognitive or linguistic?" in *Thames Valley University Working Papers in English Language Teaching. Volume 2*. 151-191.

Skehan, P. 1998. *A cognitive approach to language learning.* Oxford: OUP.

Skehan, P. 1999. "Aptitude". Paper presented at the conference on Individual Differences in Foreign Language Learning: Effects of Aptitude, Intelligence and Motivation. Tokyo, Japan.

Snow, R. E., Federico, P. A., and Montaigne, W. 1995. *Aptitude, learning and instruction: cognitive process analysis.* Hillsdale, NJ: Erlbaum.

Sparks, R. 2001. "Foreign language learning problems of students classified as Learning disabled and non-learning disabled: is there a difference?". *Topics in Language Disorders 21/1.* 38-54.

Sparks, R. and Ganschow, L. 1991a. "Foreign language learning difficulties: affective or native language aptitude differences?". *Modern Language Journal 75/1.* 3-16.

Sparks, R. and Ganschow, L. 1991b. "Individual differences in second language learning".

Studies in Second Language Acquisition 13/3. 275-298.

Sparks, R. and Ganschow, L. 1993. "The effects of a multisensory structured language approach on the native language and foreign language aptitude skills of at-risk learners: A follow-up and replication study". *Annals of Dyslexia 43/2*. 194-216.

Sparks, R. and Ganschow, L. 2001. "Aptitude for learning a foreign language*". Annual Review of Applied Linguistics 21/1.* 90-111.

Sparks, R. L., Ganschow, L., and Patton, J. 1995. "Prediction of performance in first-year foreign language courses: Connections between native language and foreign language learning". *Journal of Educational Psychology 87/4.* 638-655.

Sparks, R. L., Ganschow, L., Fluharty, K., and Littie, S. 1995. "An exploratory study on the effects of Latin on the native language skills and foreign language aptitude of students with and without disabilities". *The Classical Journal 91/2.* 165-184.

Spearman, C. 1927. *The abilities of man*. London: Macmillan.

Spoerl, D. T. 1939. "A study of some of the possible factors involved in foreign language learning". *Modern Language Journal 23/4*. 428-431.

Spolsky, B. 1973. "What does it mean to know a language: or how do you get someone to perform his competence?" In Oller, J. W., Richards, J. C. (eds.). 1973. 164-176.

Spolsky, B. 1989. *Conditions for second language learning*. Oxford: OUP.

Spolsky, B. 1995a. *Measured words.* Oxford: OUP.

Spolsky, B. 1995b. "Prognostication and language aptitude testing 1925-62". *Language Testing 12/4*. 321-340.

Stasiak, H. 1979. "Badania nad efektywnością intensywnego nauczania laboratoryjnego języków obcych w warunkach szkoły wyższej". *Przegląd Glottodydaktyczny 3.* 119-130.

Sternberg, R. J. 1982. *Handbook of human intelligence*. Cambridge: CUP.

Sternberg, R J. 1985. *Beyond IQ; A triarchic theory of human abilities.* New York: Cambridge University Press.

Sternberg, R. J. 1988. *The triarchic mind. A new theory of human intelligence.* New York: Viking.

Sternberg, R. J. 1997. *Successful intelligence.* New York: Plume.

Sternberg, R. J. 1998. "Abilities are forms of developing expertise". *Educational Researcher 2/17.* 11-20.

Stevens, S. S. (ed.). 1951. *Handbook of experimental psychology*. New York: John Wiley and Sons.

Strelau, J. 2002. *Psychologia różnic indywidualnych*. Warszawa: Wydawnictwo Naukowe Scholar.

Strong, M. 1984. "Integrative motivation: cause or result of successful second language acquisition?" *Language Learning 34/1.* 1-13.

Symmonds, P. M. 1930. "Foreign language prognosis test". *Teachers College Record 31/4*. 540-546.

Szewczuk, W. 1984. *Psychologia zapamiętywania*. Warszawa: PWN.

Tabakowska, E. and Ozga, J. 1975. "Przydatność testu prognostycznego przy rekrutacji kandydatów na filologie angielską". *Języki Obce w Szkole 1*. 35-37.

Taylor, W. L. 1953. "Cloze procedure: a new tool for measuring readability". *Journalism Quarterly 30/4*. 414-438.

Thurstone, L. L. and Thurstone, T. G. 1946-65. *Primary mental abilities*. Chicago: Science Research Associates.

Terman, L. M. 1970. *Concept Mastery Test.* New York: Psychological Corporation.

Tursi, J. (ed.). 1970. *Foreign languages and the new student: reports of the workings committees.* New York: Northeast Conference on the Teaching of Foreign Languages.

Upshur, J. A. and Homburg, T. J. 1983. "Some relations among language tests at successive ability levels". In Oller, J. W. (ed.). 1983. 3-10.

Vollmer, H. J. and Sang, F. 1983. "Competing hypotheses about second language ability: a plea for caution". In Oller J.W. (ed.). 1983. 29-79.

Wells, G. 1985. *Language development in the pre-school years*. Cambridge: CUP.

Wesche, M .B. 1981. "Language aptitude measures in streaming, matching students with methods, and diagnosis of learner problems". In Diller, K. C. (ed.). 1981. 119-154.

Wesche, M. B., Edwards, H. and Welss, W. 1982. "Foreign language aptitude and intelligence". *Applied Psycholinguistics 3/1.* 127-140.

Witkin, H. A. and Godenough, D. R. 1981. "Cognitive styles - essence and origins – field dependence and field independence". *Psychological Issues Monograph 51/1*. 36-72.

Wittich, B., von., 1962. "Prediction of success in foreign language study*". Modern Language Journal 46/3.* 208-212.

Yamashiro, A. D. (ed.). 1999. "Individual differences in the Japanese EFL context". *Working Papers in Applied Linguistics 14.* http://www.tuj.ac.jp/tesol/press/papers00 14/renyolds.html

Zeidner, M. 1986. "Are English language aptitude tests biased towards culturally different minority groups? Some Israeli findings". *Language Testing 3/1*. 80-98.

Zeidner, M. 1987 "A comparison of ethnic, sex and age biases in the predictive validity of English language aptitude tests: Some Israeli data". *Language Testing 4/1*. 55-71.

APPENDIX 1

Correlation matrix for all the variables (independent and dependent).

Missing data removed pairwise.*

(Section 1.5.1.)

	GnVrbl	IndRea	AsMem	TxtCom	TJL	TPJ	PLAT	PVoc	AVoc	SRPM	KET-A
IndReas	,0965										
	N=146										
AssoMem	,0934	,0269									
	N=146	N=146									
TextCom	,0811	,1014	,0744								
	N=146	N=146	N=146								
TJL	**,6448**	**,2528**	,1227	**,5221**							
	N=141	N=141	N=141	N=141							
TPJ	**,5376**	**,7009**	,1443	**,3626**	**,6480**						
	N=146	N=146	N=146	N=146	N=206						
PLAT	**,8282**	**,3069**	**,3665**	**,4985**	**,8036**	**,7183**					
	N=146	N=146	N=146	N=146	N=147	N=146					
PVoc	**,5644**	**,2440**	,0742	**,5488**	**,9482**	**,6181**	**,7446**				
	N=141	N=141	N=141	N=141	N=216	N=206	N=147				
AVoc	**,6600**	**,2345**	,1605	**,4426**	**,9385**	**,6005**	**,7837**	**,7806**			
	N=140	N=140	N=140	N=140	N=215	N=205	N=146	N=215			
SRPM	**,4422**	**,3859**	**,1695**	**,2105**	**,5190**	**,5451**	**,5356**	**,5104**	**,4679**		
	N=140	N=140	N=140	N=140	N=212	N=203	N=146	N=212	N=211		
KET-A	**,4546**	**,3840**	,1522	**,3565**	**,5968**	**,6193**	**,6073**	**,5528**	**,5695**	**,4759**	
	N=143	N=143	N=143	N=143	N=203	N=203	N=149	N=203	N=202	N=200	
L2 Aptitu**	**,8122**	**,3995**	**,3580**	**,4855**	**,7274**	**,7330**	**,9943**	**,6791**	**,7031**	**,5289**	**,5797**
	N=146	N=146	N=146	N=146	N=144	N=148	N=147	N=144	N=143	N=143	N=146

* *because pairwise rather than casewise deletion of missing data is used for the matrix some of the indices might differ from those quoted elswhere in the study due to different sample size (N)*

** *this is a 'combined aptitude measure' formed from collapsing the results of two TPJ tasks (TPJ 7 & 8) and PLAT (cf. section 3.7.4.2.).*

APPENDIX 2

Jurkowski's Test Językowy *Leksykon* (TJL)
(Section 2.3.2.)

I. Uzupełnianie zdań

W każdym z poniższych zdań brakuje jednego wyrazu w miejscu oznaczonym taką oto kreską:_______. Na każdej kresce napisz odpowiedni wyraz wybrany ze słowniczka pomocniczego. Chodzi o to, aby zdania miały sens.

Przykład: „Nocna wichura złamała dwa stare drzewa ____ w parku".

1. Rozproszeni po ciężkiej bitwie żołnierze zaczęli się samorzutnie ______ i tworzyć zwarte oddziały zdolne do walki.
2. Bardzo chciał pojechać na tę wycieczkę, jednak przedtem musiał uzyskać ________ od swoich rodziców.
3. Zbudził go dziwny hałas, więc zaczął ________ z napiętą uwagą.
4. Na łączce pod lasem umieszczono _________, aby zwierzęta mogły uzupełniać swoje pożywienie solami mineralnymi.
5. Aby zrozumieć wykład, trzeba uważnie _________.
6. Występ tego śpiewaka spodobał się wielu ludziom, lecz z pewnością nikogo nie _________.
7. Upolowanego zająca należy wpierw_______ , żeby nadawał się do celów kulinarnych.

Słowniczek pomocniczy.

absolutorium	nadsłuchiwać	oskubać	słomiany	zachwycać
drzewo	ligawka	polecenie	słuchać	
fistuła	lizawka	pozwolenie	uciszać	
gałąź	nefryt	radioaktywny	uważać	
grupować	oskórować	rozbrajać	zachwaścić	

II. Wyrazy o podobnych znaczeniach

Są różne słowa, które znaczą to samo - na przykład: „ziemniak" i „kartofel". Spośród sześciu wyrazów w każdym rządku poniżej trzeba wybrać i podkreślić ten wyraz, który znaczy to samo, co wyraz napisany dużymi literami nad danym rządkiem.

„ZIEMNIAK - karton, skrobia, zmiennik, smaczny, <u>kartofel,</u> roślina"

W podanym przykładzie należało podkreślić wyraz „kartofel". A teraz już samodzielnie rozwiązujemy kolejne pozycje zadaniowe:

1. DUBELTÓWKA - śrut, strzelba, dwururka, dwukropek, dubler, polować
2. BUNT - buńczuk, rębnia, wydarzenie, zgiełk, groźny, rebelia
3. ROMANSOWAĆ - spotykać, flirtować, uczuciowość, filtrować, śpiewać, lansować
4. ZNAK - symulant, dźwięk, firma, sygnał, widzieć, znalazca
5. LIBACJA - ukrywać, bibelot, karnawał, biba, liberia, harmider

III. Jaką to ma nazwę?

Będziemy mieli do czynienia z informacjami o różnych rzeczach, których nazw nie podano. Nazwy musimy sami ustalić posługując się słowniczkiem pomocniczym do tego zadania. Na kreskach pod kolejnymi zdaniami należy napisać wybrane ze słowniczka wyrazy, które najlepiej odpowiadają treści tych zdań.

Przykład: „Miejsce, gdzie zwykle kupuje się lekarstwa". ***apteka***

1. Wystaje z ziemi po ścięciu niewielkiego drzewa.
2. Gotowana kulka z ciasta; zwykle ma śliwkowe nadzienie.
3. Ma to na sobie każda owca przed strzyżeniem.
4. Postępowanie nieuczciwe, oparte na różnych wybiegach.
5. Lasek, w którym rosną małe sosenki.
6. Leży pod dziką gruszą.
7. Trzeba to okazać wartownikowi, aby wejść na teren strzeżony.

Słowniczek pomocniczy

apteka	farmacja	krętactwo	pieniek	runo
attyka	gajówka	krokiet	piernik	tupet
chojniak	knebel	osęk	przepiórka	ulęgałka
chojrak	knedel	oszczypek	przepustka	ulicówka
dziczka	kręgarstwo	pieczątka	run	

IV. Określenia

Oto zestaw określeń przymiotnikowych:

alkoholowy (lub alkoholowa) gąsienicowy (lub gąsienicowa)
jazzowy (lub jazzowa) meteorologiczny (lub meteorologiczna)
piaszczysty (lub piaszczysta) wielkopiecowy (lub wielkopiecowa)
zamaszysty (lub zamaszysta)

Te określenia należy wpisywać w miejsca oznaczone poziomą kreską obok wyrazów-nazw różnych rzeczy. Oczywiście przy danym wyrazie napiszemy tylko to określenie, które można do danego wyrazu sensownie zastosować, np.: „droga piaszczysta". Niektóre określenie trzeba wykorzystać kilkakrotnie. Gdy w zestawie brak odpowiedniego określenia dla jakiegoś wyrazu, wtedy na kresce postawimy znak zapytania. Część pracy została już wykonana.

alkowa	________________	nalewka	________________
arbuz	________________	orkiestra	________________
droga	*piaszczysta*_______	pogoda	________________
fermentacja	________________	pojazd	________________
gaz	________________	pieczeń	________________
Jaz	________________	prognoza	________________
komunikat	________________	surówka	________________
koncert	________________	teren	*piaszczysty*______
krok	________________	wyciąg	________________
muzyka	________________		

V. Jeśli...

Poniżej znajduje się kilka zdań zaczynających się od słowa „jeśli". W każdym z nich opuszczono jeden wyraz. Przeczytaj uważnie zdania i na kreskach napisz potrzebne wyrazy, aby całość miała sens. Wpisywane wyrazy trzeba koniecznie wybrać ze słowniczka pomocniczego. Można czasem zmienić formę wyrazu ze słowniczka, żeby zdanie było poprawne.

Przykład: „Jeśli będzie ładna pogoda, pojedziemy na *wycieczkę".*

1. Jeśli nam wystygła zupa, można ją po prostu _____________ .
2. Jeśli możemy czymś kierować na odległość, mówmy wtedy, iż jest to tak zwane _____________ kierowanie.
3. Jeśli pewien zakład wytwarzał dawniej 20 ton marmolady rocznie, a obecnie 40 ton - to oznacza, iż _____________ on wielkość swej produkcji.
4. Jeśli czegoś nie rozumiemy, można poprosić o _____________.
5. Jeśli jakieś genialne dzieła lub idee weszły na zawsze do dorobku światowej kultury, można o nich powiedzieć, iż zasłużyły sobie na _____________.
6. Jeśli od jakiegoś wyrazu oddzielimy końcówkę, to zostanie nam część _____________ tego wyrazu.
7. Jeśli poszerza się zakres swobód obywatelskich i zwiększa udział ludności w sprawowaniu władzy, wówczas mówimy o zachodzeniu procesu _____________ życia społecznego.

Słowniczek pomocniczy:

artykuł	nieśmiertelność	tematowy
centralny	odgrzać	tułaczka
demobilizacja	pochwała	wycieczka
demokratyzacja	podwoić	wyjaśnienie
gramatyczny	praktykant	zdalny
kompetentny	przypomnienie	zestalić
liścień	rozlać	zważyć

VI. Dziesięć zdań

Poniżej znajduje się dziesięć zdań jakby wziętych z różnych opowiadań. W każdym z nich opuszczono jeden wyraz, a lukę oznaczono poziomą kreską. Przeczytaj kolejne zdania, zastanów się i napisz na każdej kresce taki wyraz, który byłby najbardziej odpowiedni w danym zdaniu.

1. Piotruś zdjął w końcu upragnioną książeczkę z górnej półki, ale musiał _____________ na stołku - inaczej by jej nie dosięgną!.
2. Siedząca obok dama rozłożyła trzymany w ręce _____________ i poruszała nim miarowo, aby się ochłodzić.
3. Los uśmiechnął się wreszcie do Małgosi - to tak, jakby zjawił się _____________ z bajki i swą cudowną pałeczką przegnał troski.
4. Strzelajcie celnie - powiedział dowódca - gdyż kończy się nam amunicja i nie wolno _____________ ani jednego naboju.
5. Prócz skromności cechowała go również _____________ na piękno.
6. Nie mogąc zdobyć solidnie zbudowanego zamku, nieprzyjaciel podciągnął od jego mury ciężkie machiny _____________ .
7. Słońce było już _____________ na niebie, gdy się nareszcie obudził.
8. Od pewnego czasu nie wiodło się Hani dobrze w szkole, a dzisiejsza dwójka z fizyki _____________ miary jej niepowodzeń.
9. Nadajnik psuł się często i po każdej jego naprawie trzeba było od nowa _____________ łączność z dowództwem.
10. Obywatel górzystego kraju europejskiego słynącego z wyrobu serów i precyzyjnych zegarków - to z pewnością _____________

VII. To samo, ale trochę inaczej

Bywają różne słowa o tak podobnych znaczeniach, iż często używane są zamiennie. Tego właśnie dotyczy zadanie.Niżej podano kilka par zdań. Pierwsze zdanie każdej pary jest kompletne, w drugim zaś znajduje się luka oznaczona poziomą kreską. Na kresce należy napisać taki wyraz, aby oba zdania informowały o tym samym, przy czym wpisany wyraz nie może być powtórzeniem żadnego z wyrazów występujących w zdaniu pierwszym. Dla ułatwienia na każdej kresce napisano już pierwszą literę poszukiwanego wyrazu.

Przykład: a. Pociąg stanął, bo lokomotywa się zepsuła.
b. Pociąg stanął, bo *parowóz* się zepsuł.

Podobnie trzeba postąpić z kolejnymi pozycjami tego zadania:

1.

a. Od pewnego czasu wzrasta spożycie olejów roślinnych.
b. Od pewnego czasu wzrasta k_____________ olejów roślinnych.

2.

a. W ogródku Kasi zakwitły dziś pierwsze mieczyki.
b. W ogródku Kasi zakwitły dziś pierwsze g__________

3.

a. Wszyscy bardzo go lubili, więc gdy wrócił po latach, spotkał się z gorącym przyjęciem.
b. Wszyscy bardzo go lubili, więc gdy wrócił po latach, spotkał się z s_____________ przyjęciem.

4.

a. Pochodzi z zamożnego rodu, lecz teraz cała jego fortuna ledwie starczy na pokrycie długów. b. Pochodzi z zamożnego rodu, lecz teraz cała jego rn__________ ledwie starczy na pokrycie długów.

5.

a. Zjawił się dziwny człowieczek o wręcz lilipucim wzroście.
b. Zjawił się dziwny człowieczek o wręcz k____________ wzroście.

VIII. Jak w encyklopedii ...

Na poziomych kreskach trzeba będzie napisać te słowa, których dotyczą podane obok wyjaśnienia. Pierwsze litery odpowiednich wyrazów już napisano.

Przykład: *armata*, rodzaj działa o długiej lufie, płaskim torze pocisku i dużej donośności.

Napisz pięć wyrazów o podanych niżej znaczeniach:

l.

b_____________, płaskie i podłużne naczynie metalowe służące zwykle do pieczenia mięs.

2.

i_____________, gad morski żyjący dziesiątki milionów lat temu (w erze mezozoicznej), którego szczątki kopalne znajdowane są w ziemi; rybojaszczur.

3.

m____________, podręczny zbiór tekstów na różne okoliczności religijne; książka do nabożeństwa, mszalik.

4.

p_____________, roślina warzywna o jadalnych owocach; również przyprawa wytwarzana z jej owoców w postaci czerwonoceglastego proszku o smaku ostrym lub łagodnym (charakterystyczna szczególnie dla kuchni węgierskiej).

5.

t_____________, ogół narządów wewnętrznych znajdujących się w klatce piersiowej, jamie brzusznej i miednicy; wnętrzności.

IX. Zadanie z ułatwieniami

W zdania przedstawione niżej trzeba będzie wpisać po jednym wyrazie. Aby to ułatwić, zastosowano dwa rodzaje podpowiedzi: znana będzie liczba liter w szukanym wyrazie oraz dwie litery tego wyrazu (pierwsza i jedna ze środkowych). Wyrazy należy pisać na liniach przerywanych, gdzie każda mała kreseczka oznacza miejsce na jedną literę, a dwie litery znajdują się już na swoich pozycjach. Gdyby np. chodziło o wyraz „fasola", to miejsce na ten wyraz wyglądałoby następująco: _f _ _ o _ _ . A oto zdania do uzupełnienia:

1. Dyżurna telefonistka zwykle a _ _ _ _ u _ _ rozmowę międzymiastową, zanim połączy nas z rozmówcą.
2. Nie ma sensu ciągle p _ _ _ c _ _ _ wahadło; ten stary zegar trzeba po prostu oddać do naprawy.
3. Włosy miała spięte pamiątkową z _ _ i _ _ _ , którą otrzymała od babci.
4. Nie należy r _ _ _ ł _ _ _ _ _ nie sprawdzonych informacji.
5. W nocy musiał być przymrozek, bo dachy domów p _ _ _ _ l _ _ _ od szronu.
6. Z rozrzewnieniem wspominał czasy dzieciństwa, gdy pracowita jak mrówka matula i poczciwy o _ _ _ _ l _ _ opiekowali się nim.
7. Badania archeologiczne wykazały, iż ludność żyjąca tu przed wiekami reprezentowała bardzo niski poziom rozwoju cywilizacyjnego; szczególnie znalezione narzędzia cechował p _ _ _ _ _ _ w _ _ _ techniczny.
8. Potwierdziło się p _ _ _ _ u _ _ _ _ _ _ _ _ Janka - rzeczywiście był błąd w obliczeniach.
9. Krwinki białe należące do granulocytów ze względu na ich pewne właściwości chemiczne dzielimy na trzy grupy: obojętnochłonne, zasadochłonne oraz k _ _ _ o _ _ _ _ _ _ _.
10. Prawa dotyczące współoddziaływania zmiennych pól elektrycznych i magnetycznych - to prawa e _ _ _ _ _ _ _ y _ _ _ _ _ _ _ _.

APPENDIX 3

Test Predyspozycji Językowych
(Section 2.4.1.)

Witaj!
Przystępujesz do badania predyspozycji językowych. Przed Tobą zestaw 40 zadań. Przeczytaj uważnie polecenia. Wracaj do treści zadań tyle razy, ile będzie Ci to potrzebne. Wybierz jedną odpowiedź spośród 4-8 podanych. Pamiętaj o zakodowaniu odpowiedzi na karcie odpowiedzi. Pracuj spokojnie. Na rozwiązonie zadań masz 60 minut.
Życzymy powodzenia!

ĆWICZENIE 1
W każdym z poniższych tekstów brak jest fragmentu zdania. Znajdź właściwy fragment odpowiednio wybierając odpowiedź a, b, c lub d.
1. Podczas wizyty w Paryżu w 1894 roku Antoine Lumiere, właściciel fabryki płyt fotograficznych w Lyonie, wybrał się na pokaz filmowy. Film wyświetlany w wynalezionym trzy lata wcześniej kinetoskopie mogła oglądać tylko jedna osoba ,..
Na ekranie o powierzchni kilkunastu centymetrów kwadratowych rozgrywały się króciutkie scenki, a taśmę przesuwał elektryczny silniczek, uruchamiany przez widza za pomocą monety.
A. która zaglądając przez okular do wnętrza niewielkiej skrzynki
B. zaglądając przez okular do wnętrza niewielkiej skrzynki
C. podczas zaglądania przez okular do wnętrza niewielkiej skrzynki
D. która zaglądała przez okular do wnętrza niewielkiej skrzynki
2. Ankieta to narzędzie urzędowej maszyny wtłaczajacej człowieka w anonimowy tłum osobników w przedziale wieku od 12 do 1 7 lat; .... uczniów szkół państwowych czy prywatnych. Wypełniamy ją przy okazji wizyt w urzędach, biurach, przychodniach, a także przy wyborze szkoły. Jej celem jest zdobycie informacji o człowieku.
A. na samodzielnym utrzymaniu rodziny lub krewnych
B. na utrzymaniu samodzielnej rodziny lub nie
C. utrzymujących samodzielnie rodzinę lub siebie
D. utrzymujących się samodzielnie lub będących na utrzymaniu rodziny
3. Różnice między bogatą szlachtą (magnaterią) a zwykłymi szlachciurami pogłębiają się. Magnaci wstydzą się swych ubogich krewniaków. Rozbu-dowują swe posiadłości —zamki przerabiają na pałace, wyposażają je w najnowsze osiągnięcia techniki, ...
A. przyozdabiają czym się da
B. które przyozdabiają, jak im się podoba
C. przyozdobione według własnego gustu
D. których nie da się przyozdobić
4. NUDA? NIE MAM NATO CZASU! Właśnie przeczytałem artykuł o nudzie. Nie chcę się chwalić, ale ja... nie mam na to czasu! Sześć razy w tygodniu chodzę do szkoły muzycznej, odrabiam zadania domowe z normalnej szkoły, ale także z harmonii czy z kształcenia słuchu. Mam bardzo wymagającą panią .... Zresztą prawie cała moja klasa ciężko „haruje". A nuda? Myślę, że wszystko zależy od pomysłowości konkretnej osoby.
A. nauczycielkę, która mnie wkurza prawie codziennie
B. z biblioteki, w której spędzam dużo czasu
C. od fortepianu, więc muszę ćwiczyć codziennie po dwie, trzy godziny
D. od fortepianu, na którym trzeba ćwiczyć każdego dnia

ĆWICZENIE 2
W każdym zdaniu podkreślone słowo można zastąpić innym słowem, które ma to samo znaczenie. Wybierz odpowiednie słowo, zakreśla-jąc odpowiedź a, b, c lub d.

5. Pasowanie na rycerza to uroczystość przyjęcia młodego mężczyzny do grona rycerzy. Nazwa wywodzi się od symbolicznego gestu przypięcia ozdobnego pasa. Pasowania dokonywał władca lub znany rycerz.

A. zachowania C. ruchu
B. machania D. pokazania

6. Elektroniczne listy Darka nadal sprawiają, że czuję się kimś wyjątkowym, niepowtarzalnym, potrzebnym. Skłaniają mnie do refleksji, do zadawania sobie pytań, które wcześniej wydawały mi się błahe, nieistotne.

A. rewanżu C. powstrzymania się
B. zastanowienia D. odwagi

7. Historia milczy na temat honorariów pierwszych bohaterów kina. Aktorstwo filmowe dopiero za kilka lat miało stać się jednym z najlepiej opłacanych i najbardziej pożądanych zawodów świata.

A. zarobków C. zagrożeń
B. zaszczytów D. zarabiania

8. Dziewicza Amazonia, porośnięta gigantycznym lasem deszczowym i bedąca największym rezerwuarem słodkiej wody świata, stało się celem wyprawy dwóch niemieckich dziennikarzy telewizyjnych.

A. zapleczem C. obszarem
B. bogactwem D. zbiornikiem

9. Yukon Quest to jeden z najtrudniejszych i zarazem najbardziej prestiżowych wyścigów psich zaprzęgów. Jego szlak wiedzie przez śnieżne bezdroża Alaski, a z dziką naturą, zmęczeniem i rywalami zmaga się zarówno człowiek, jak i towarzyszące mu psy.

A. poważanych C. poważnych
B. nieoczekiwanych D. zaskakujących

ĆWICZENIE 3
Do słów w tabelce dobierz słowa z ramki tak, aby parami tworzyły słowa o tym samym znaczeniu. Nie wszystkie słowa z ramki nadają się do wykorzystania w tym ćwiczeniu.

11. dyskusja
12. rywalizować
13. sporządzić
14. wyjątkowy

A. niecodzienny D. niewybredny
B. rozmowa E. przygotować
C. pospolity F. współzawodniczyć

ĆWICZENIE 4
Do słów w tabelce dobierz słowa z ramki tak, aby parami tworzyły znaczenia przeciwne. Nie wszystkie słowa z ramki nadają się do wykorzystania w tym ćwiczeniu.

15. zainspirować	
16. abstrakcyjny	
17. szkalować	

18. ujarzmiony	

A. rzeczywisty D. niezawisły
B. zniechęcić E. doceniać
C. obmawiać F. pokonany

ĆWICZENIE 5

W każdej linijce jest jedno słowo, które nie pasuje do pozostałych. Znajdź to słowo. Zaznacz wybraną przez siebie odpowiedź, tak jak na podanym przykładzie. (Odpowiedź zaznacz na karcie odpowiedzi.)

19. A. USCHNIE C. BÓJ SIĘ
B. SIĘGNĄŁEM D. SKARŻYPYTA
20. A. MIOTŁA C. SANKI
B. NOŻYCZKI D. DRZWI
21. A. LEPSZY C. DŹWIĘCZNIE
B. DZIWNIEJSZY D. BARDZIEJ DOCIEKLIWY
22. A. BLADO C. CZWORO
B. SZYBKO D. NIEZNACZNIE

ĆWICZENIE 6

Wybierz analogiczne formy od słów według przykładów.

23. czesać — uczesałem
biec — ....
A. biegłem C. biegałem
B. pobiegłem D. pobiegałem
24. pisać – piszących
nałożyć -- ....
A. nałożonych C. nakładanych
B. nałożących D. nakładający
25. karcić – skarciwszy
grzmotnąć - ....
A. grzmotnąłszy C. zgrzmociwszy
B. grzmotnąwszy D. grzmociwszy
26. kaganiec – kagańców
kajakarz - ....
A. kajaków C. kajakarzów
B. kajakowców D. kajakarzy
27. chwytać – został uchwycony
sadzić
A. został posadzony C. został usadzony
B. został osądzony D. został podsadzony
28. źli - gorsi
nieświadomi - ....
A. nieuświadomieni C. najbardziej nieuświadomieni
B. nieświadomsi D. bardziej nieświadomi

ĆWICZENIE 7

Porównaj formy utworzone ze słów j. łacińskiego z ich polskim znaczeniem,
laudare — chwalić
laudabam — chwaliłem laudabit — pochwali
laudatur — jest chwalony laudate — chwalcie

laudaret — pochwaliłby
a następnie dobierz spod tabeli odpowiednie formy słów łacińskich, które są tłumaczeniem słów polskich.

29. opowiadać - ....
30. jest opowiadany - ....
31. opowie - ....
32. opowiadałem - ....
33. opowiedziałby - ....
34. opowiedzcie - ....

A. narraret	E. narrare
B. narrat	F. narrate
C. narratur	G. narrabam
D. narrantur	H. narrabit

ĆWICZENIE 8
Przeanalizuj dokładnie poniższe zdania (napisane w języku fikcyjnym) oraz ich polskie tłumaczenia.

lana — dziewczynka	brivan — malować	kres — obrazek
vesso — pokój	ter — duży	

Lana briva kresok em vesso.	Dziewczynka maluje obrazek w pokoju.
Lanas brivas kresosk em vessos.	Dziewczynki malują obrazki w pokojach.
Lana brivada kresok ter.	Dziewczynka namalowała duży obrazek.
Lana brivadra vessok.	Dziewczynka namaluje pokój.
Lana brivabat em kres.	Dziewczynka jest namalowana na obrazku.

Teraz z wykorzystaniem kolejnej grupy słów z tego języka:

chłopiec — pok	sadzić — melivan	drzewo — desso
ogród — sana	maty — ker	

Dobierz do podanych polskich zdań ich odpowiedniki w tym języku.

35. Chłopiec posadził małe drzewo.
36. Drzewo jest posadzone w ogrodzie.
37. Chłopiec sadzi drzewo w ogrodzie.

A. Pok melivada ker dessok.	C. Desso melivabat em sana.
B. Pok meliva dessok em sana.	D. Pok melivada dessok ker.

Teraz dobierz poprawne końcówki w tłumaczeniu zdania polskiego na ten język.
Chłopcy sadzą drzewa w ogrodach Pokos meliv......38..... dess....39.... san...40...

38.	A. – a	B. –as	C. –ada	D. –dra
39.	A. –o	B. –ok.	C. –os	D. –osk
40.	A. –as	B. –ask	C. –a	D. -ak

APPENDIX 4

Progressive Language Aptitude Test - PLAT

(Section 2.4.2.)

ZADANIE 1

W każdej ponumerowanej linijce poniżej jest ukryte jedno słowo języka polskiego. Znajdź to słowo i zakreśl je. ***Pracuj szybko!*** *Na wykonanie tego zadania masz* ***5 minut****. Po tym czasie gzaminator poprosi Cię o odłożenie kartki z zadaniem.* **UWAGA: słowa muszą być dłuższe niż 4 litery!**

Przykład:

0	O	D	W	E	D	***Z***	***W***	***O***	***L***	***E***	***N***	***N***	***I***	***K***	L	A	F	O	N	Y
1	P	L	A	Z	A	Ż	E	N	O	W	A	N	I	E	R	O	B	A	J	U
2	A	W	A	R	O	B	R	O	N	K	J	A	Ł	M	U	Ż	N	A	K	O
3	P	E	D	A	N	U	Ł	U	D	A	L	E	K	R	O	M	K	U	S	K
4	Z	A	D	E	M	I	S	Z	Y	D	E	R	C	Z	Y	S	T	K	R̨	E
5	S	P	R	O	R	Ę	D	O	W	N	I	K̨	L	O	D	O	Ł	Ę	Ż	Y
6	Z	A	W̨	Y	L	A	N	S	O	W	A	Ć	M	O	D	L	U	T	R	A
7	B	U	Ń	C	Z	U	C	Z	N	Y	H	U	F	N	O	M	I	P	N	I
8	P	R	Z	E	T	R	A	D	Y	Z	A	N	O	S	T	A	L	G	I	A
9	S	T	R	Z	E	M	O	D	U	L	A	R	N	Y	L	O	M	R	Z	O
10	S	W	A	D	L	I	A	S	C	E	T	A	C	Z	Y	T	G	U	R	A
11	O	K	O	N	F	I	D	E	N	T	J	A	N	T	R	Z	E	B	R	O
12	M	O	T	K̨	R	Y	M	I	N	O	G	E	N	N	Y	Z	A	S	Z	P
13	P	R	Z	Ó	T	R	Z	E	W	I	A	D	O	M	K̨	R	O	S	Ł	I
14	P	L	O	T	C	I	Ż	B	A	N	A	K	R	O	Ś	Ó	D	R	A	B
15	B	R	U	Z	A	K	I	M	P	O	N	O	W	Ą	Ć	P	O	K	Ǫ	J
16	O	P	N	A	D	W	Ą	T	L	I	Ć	M	O	Ż	L	I	T	O	Ś	T
17	Z	D	O	H	A	R	M	I	D	E	R	K	O	L	A	C	Z	F	O	P
18	B	Ą	C	Z	A	K	C	J	O	P	T	U	Ł	A	C	Z̨	K	A	F	K
19	P	Ǫ	S	M	A	L	E	R	G	I	A	R	T	R	A	Ń	S	P	Z	O
20	M	Ó	R	T	O	P	A	C	T	W	O	R	D	O	R	A	C	Z	K	E

ZADANIE 2

Usłyszysz krótką historyjkę składającą się z 6 zdań. Po każdym zdaniu będzie przerwa, w której zapiszesz usłyszane zdanie. Każde zdanie usłyszysz tylko ***jeden raz****.. Słuchaj z uwagą czytanych zdań. Długość zdań będzie się stopniowo zwiększać.*

ZADANIE 3

Poniżej masz 5 zdań do tekstu, który przed chwilą usłyszałeś. Zdecyduj, które z podanych zakończeń A, B czy C, odpowiada temu co pamiętasz z tekstu. Zakreśl kółkiem właściwą odpowiedź.

1. Przed wynalezieniem aparatu do rejestracji czynności bioelektrycznej mózgu badania nad snem:
   A. były w ogóle nie możliwe
   B. były możliwe tylko w ograniczonym zakresie
   C. nie były w ogóle podejmowane przez badaczy
2. Badanie elektroencefalografem jest dziś:
   A. powszechnie stosowane na całym świecie
   B. stosowane tylko w klinikach dla psychicznie i nerwowo chorych
   C. prawie w ogóle zapomniane
3. Podczas pierwszej rejestracji czynności bioelektrycznej mózgu:
   A. uzyskano bardzo słaby zapis fal alfa
   B. uzyskano bardzo wyraźny zapis fal alfa
   C. nie uzyskano żadnego zapisu fal

ZADANIE 4

W wyrazach od 1 do 16 usunięto wszystkie litery-samogłoski (i, y, e, ę, a, o, ą, u, ó)*, ale zachowano kolejność liter-spółgłosek w wyrazie. Wstaw litery-samogłoski tak aby powstały wyrazy języka polskiego. Jeżeli powstanie więcej niż jeden wraz, podaj tylko jeden.Patrz przykład.*

Przykład: 0. **prztłmczć – przetłumaczyć** 00. **ntrnt - internet**

1. dbrdszn	9. bstwć
2. przdptpw	10. klgczn
3. sttstczn	11. wnklw
4. dpwdnk	12. przwljwn
5. kmpn	13. frmmnt
6. ndpwdzln	14. rzklktn
7. ncjtw	15. grnstj
8. grlw	16. tprtrt

ZADANIE 5

Jeszcze raz przypomnij sobie tekst o badaniach nad snem. Zdecyduj, które ze zdań poniżej są prawdziwe a które fałszywe. Zakreśl kółkiem TAK albo NIE w arkuszu odpowiedzi.

1. Podczas snu trudno jest mierzyć oddech, tętno i temperaturę ciała.
2. Najwcześniejsze badania głębokości snu polegały na tym, że wyrywano osobę badaną ze snu.
3. Już w wieku XIX fizjolog Kohlschutter ustalił, że faza snu tuż po zaśnięciu jest najpłytsza.
4. Rytm fal 'alfa' jest cykliczny i pojawia się 10 razy na sekundę.
5. Fale 'alfa' występują w stanie silnego pobudzenia emocjonalnego.
6. Odkrycie Bergera zostało od razu entuzjastycznie przyjęte przez świat nauki.

ZADANIE 6

*Przeczytaj bardzo uważnie i w skupieniu poniższy tekst, który jest fragmentem strony z gazety. Zwróć uwagę na **treść i układ strony**. Poświęć temu zadaniu co najmniej 5 min.*

(full text in Appendix 4A)

ZADANIE 7

*Odpowiedz teraz na pytania do tekstu, który przed chwilą przeczytałeś. **Nie możesz wracać do tekstu z Zadania 6.** Polegaj tylko na swej pamięci.*

1. Ofiara zamachu w Holandii to: A. Michał Bajorek C. Giscard d'Estaing
B. Pim Fortuyan D. Jospen
2. Rada Czterech zajmuje się: A. sprawą mordu politycznego w Holandii
B. wykrytym przypadkiem BSE w Małopolsce
C. oceną ekonomii kraju
D. organizowaniem prareferendum dot. UE
3. Fortuyan był popularny bo: A. popierał go nowy premier Francji
B. chciał wystąpienia Holandii z Unii Europejskiej
C. głosił że Holandia powinna być dla Holendrów
D. chciał obowiązkowych szczepień dla krów
4. Michał Bajorek sprzedał 2 krowy bo: A. nie mógł mieć od nich cieląt
B. miały 'dziury w mózgu"
C. bo handlarz dawał dobrą cenę
D. nie miał ich czym karmić
5. Państwo Bajorkowie: A. mieszkają 10 km od Tarnowa
B. sprzedali krowy bo nie mieli dla nich 'papierów'
C. po sprzedaży krów zostali z pustą oborą
D. są bardzo zdziwieni, że Krasula ma BSE
6. W Kujonie Polskim: A. w piątek opinie specjalistów o gospodarce
B. dziś rady dotyczące matury z polskiego
C. w niedzielę 'czat' z Fortuyn'em
D. jutro wyniki sondażu o UE z 114 miast Polski

ZADANIE 8

Poniżej masz listę czasowników w afrykańskim języku Akareke *i ich polskie znaczenie. Zapamiętaj tę listę. **Nic nie notuj!** Polegaj jedynie na swej pamięci!! Poświęć temu zadaniu nie więcej niż 2 minuty.*

Akareke		***Polski***
minum	-	**pić**
tuka	-	**lubić**
apal	-	**wiedzieć**
duduk	-	**siedzieć**
tidur	-	**spać**
omong	-	**mówić**
ingat	-	**pamiętać**
wafat	-	**umierać**
tertawa	-	**śmiać się**
datang	-	**przychodzić**

(adapted from Meara et al. 2001. "Language Aptitude Test".)

ZADANIE 9

Przypomnij sobie listę słów w języku Akareke, których przed chwilą się uczyłeś. Wpisz poniżej brakujące polskie znaczenia słów. **Jeszcze raz przeczytaj i zapamiętaj tę listę.**

Akareke		***Polski***
minum	-	**pić**
tuka	-	....
apal	-	**wiedzieć**
duduk	-	....
tidur	-	**spać**
omong	-	....
ingat	-	**pamiętać**
wafat	-	....
tertawa	-	**śmiać się**
datang	-	....

(adapted from Meara et al. 2001. "Language Aptitude Test".)

ZADANIE 10

Przypomnij sobie wcześniej przeczytany tekst z gazety. W arkuszu odpowiedzi zakreśl kółkiem TAK lub NIE.

1. Handlarz kupił krowy od Bajorków by je sprzedać na ubój.
2. Partia Fortuyana miała szansę na 35% głosów w wyborach parlamentarnych.
3. Bajorek karmił Krasulę kupnymi paszami i mączką.
4. Osobnik podejrzany o dokonanie zamachu usiłował uciec samochodem.
5. Nowy premier Francji współpracował kiedyś z prezydentem d'Estaingem.

ZADANIE 11

Po raz ostatni przypomnij sobie listę słów w języku Akareke i ich polskie odpowiedniki. Lista czasowników polskich jest pomieszana. Przyporządkuj polskie znaczenia słowom w Akareke.

Akareke	***Polski***
(1) minum	A. lubić
(2) tuka	B. przychodzić
(3) apal	C. umierać
(4) duduk	D. siedzieć
(5) tidur	E. śmiać się
(6) omong	F. wiedzieć
(7) ingat	G. mówić
(8) wafat	H. spać
(9) tertawa	I. pić
(10) datang	J. pamiętać

(adapted from Meara et al. 2001. "Language Aptitude Test".)

ZADANIE 12

W tekście poniżej z większości wyrazów usunięto drugą ich połowę w taki sposób, że zostały wolne miejsca na brakujące litery. Przeczytaj uważnie cały tekst i uzupełnij brakujące litery tak jak to zrobiono w przykładzie.

Przykład: **Cza_ _ rozb_ _ _ _ _ t_ la_ _ niew_ _ _ prus_ _ _ _.**
Czasy rozbiorów to lata niewoli pruskiej.

Pojezierze Leszczyńskie

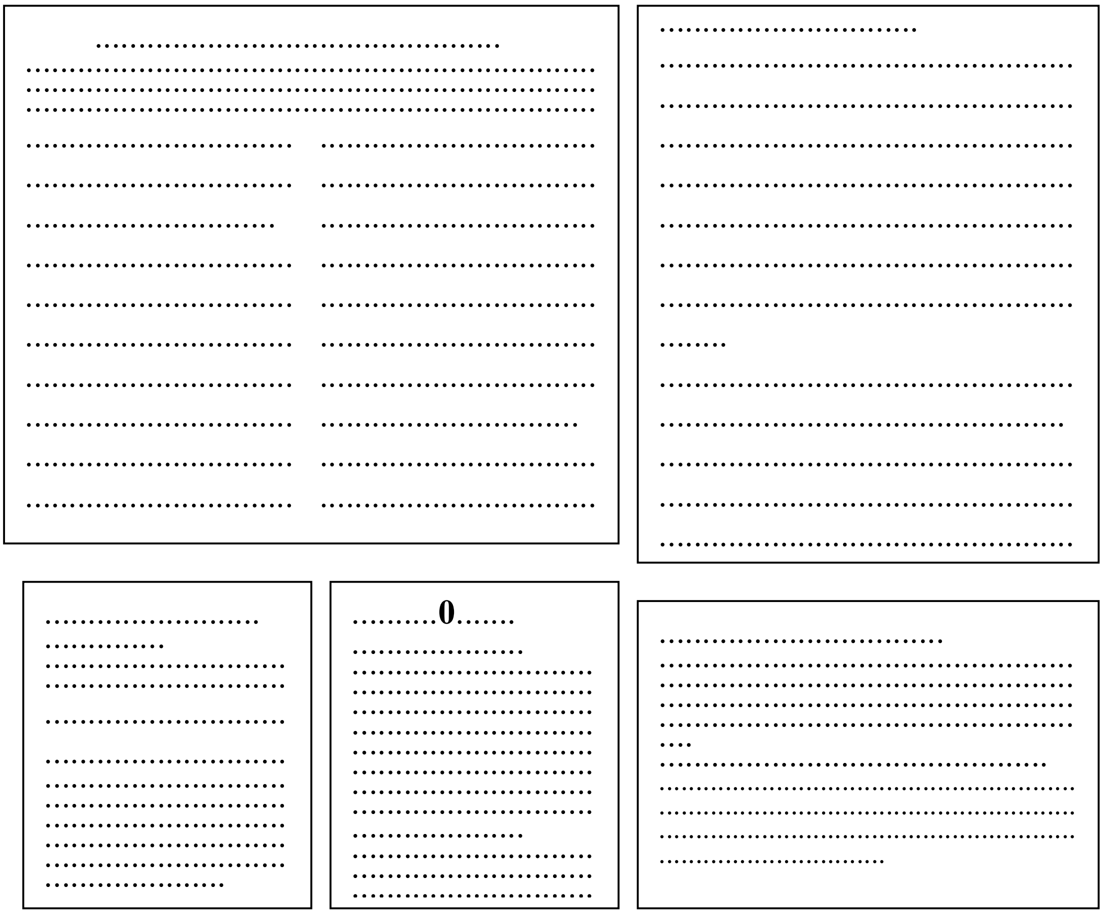
Pojezierze Leszczyńskie to obszar leżący w południowo-wschodniej części Wielkopolski. W wodach ukryt _ _ _ (1) pomię _ _ _ (2) morenow _ _ _ (3) wzgórz _ _ _ (4) rynnowych jezior odbijają się sta _ _(5) sos_ _ _ _ (6) bo_ _ (7), a także malow_ _ _ _ _ (8), choć znisz_ _ _ _ _ (9) wiatraki-koźlaki, przyda _ _ _ (10) tutejsz _ _ _ (11) krajob_ _ _ _ _ _ (12) niezw_ _ _ _ _ _ (13) ur _ _ _ (14).

Ten piękny teren dzielił lo_ _ (15) całej Wielkopolski. Licz _ _ (16) znalezi _ _ _ (17) archeol_ _ _ _ _ _ _ (18) w post_ _ _(19) zachowa _ _ _ _(20) do dz_ _(21) grodz _ _ _(22) stożkowatych świad _ _ _ (23) o osadnic _ _ _ _ (24) sięgaj _ _ _ _ (25) VII – IX wieku. W okresie rozbicia dzielnic _ _ _ _ _ (26) obec _ _ (27) zie _ _ _ (28) wschowska nale_ _ _ _ (29) do książąt głogowskich. W 1343 roku przyw_ _ _ _ _ (30) ją Polsce Kazimierz Wielki. Z kolei wysunięte najbardziej na zachód miasto Sława najpierw wchodziło w skład Śląska, a potem Prus. *(adapted from 'Spotkania z zabytkami" 8/2001)*

ZADANIE 13

Poniżej masz układ graficzny tekstu gazetowego, który czytałeś w zadaniu 6. Na dole strony masz 10 ponumerowanych zdań wyjętych z tego tekstu. Twoim zadaniem jest wpisanie cyfr od 1 do 10 w przybliżone miejsca, w których zdania te wystąpiły w tekście.

Nie pisz całych zdań – wpisz tylko numery zdań. Patrz - przykład.



Przykład: **0. Dzień w skrócie**

1. Referendum Unii o Unii.
2. Francuskie tempo.
3. Zabili populistę.
4. Miał broń i próbował się ukryć na stacji benzynowej.
5. Pełny tekst analizy dotyczącej I kwartału – na s.25.
6. Raport Rady Czterech nr 1
7. – opowiada o pierwszej polskiej krowie chorej na BSE jej właściciel.
8. Jego partia miała szansę na co najmniej15% głosów w wyborach parlamentarnych planowanych na 15 maja.
9. Jutro: czat maturalny na www.gazeta.pl
10. Płacił z góry, cenę dawał dobrą.

ZADANIE 14

Poniżej masz podane jakim literom języka polskiego odpowiadają litery w języku warackim. 'Przetłumacz' na język waracki słowa od 1 do 5 (w miarę dokładnie zapisując litery warackie); a na język polski słowa od 6 do 10 zapisane w języku warackim.

Polski		Waracki	Polski		Waracki
i	=	∈	b	=	≡
y	=	®	d	=	Δ
e	=	O	l	=	℘
a	=	T	r	=	τ
o	=	↓	m	=	ϖ
u	=	Φ_	n	=	×
p	=	ℵ	s	=	©
k	=	÷	z	=	Σ

1.) smyk	6.) ℘T℘÷T
2.) plama	7.) O÷τT×
3.) brudny	8.) ℵ℘↵ϖ≡T
4.) zimno	9.) ↵ΔτT
5.) monsun	10.) ≡℘↓×Δ®×÷∈

ZADANIE 15

W wolnych miejscach w tekście brakuje po ***jednym*** *wyrazie. Dobierz te wyrazy tak by pasowały znaczeniem do całości tekstu i wpisz je w arkusz odpowiedzi.*

UWAGA: By wykonać to zadanie najpierw przeczytaj uważnie cały tekst

Sprawcy piątkowego napadu na komisariat **...(1)....** i bank w Gołczewie w Zachodniopomorskim są nadal na **...(2)....** . Trzech zatrzymanych w piątek pod Dziwnowem sprawców **....(3)....** o napad, policja **....(4)....** ponieważ nie udowodniono im winy. Policja skrupulatnie **....(5)....** wszystkie informacje , które do niej docierają. Zamaskowani **....(6)....** wtargnęli do komisariatu, gdzie skrępowali jedynego pełniącego wtedy **....(7)...** policjanta oraz przesłuchiwanego przez **....(8)....** mężczyznę, a potem napadli na sąsiedni **....(9)....** spółdzielczy i **....(10)....** z niego około 30 tys. złotych. *(from a test by Z. Dzięgielewska)*

ZADANIE 16

Opierając się na poprzednim zadaniu zdecyduj, które 'wyrazy' języka warackiego (A,B,C, D, E czy F) odpowiadają polskim słowom podanym poniżej.

Nie wracaj do poprzedniego zadania!

1.) karny	-	A. ÷TτϖT	C. ÷Tτ×®
		B. ↓Σ×®ϖ	D. ΔΦ_τℵ↓
2.) mama	-	A. ΣTΣT	C. ©∈©∈
		B. ℘Φ_℘Φ_	D. ϖTϖT
3.) korona	-	A. ÷↓τ↓×T	C. ÷↓÷↓Δ∈
		B. ×OτO×T	D. ℵ®Δ®ΣT
4.) ulubiony	-	A. Φ_℘Φ_≡∈O÷T	C. Φ_℘Φ_≡∈↵×®
		B. ®×®≡∈↵×®	D. ↓℘↓Δ∈Φ_τO
5.) malarz	-	A. ϖT℘TτΣ	C. ×↓ΔTτΣ
		B. ℘TϖTτΣ	D. ϖT℘T©ϖ

KONIEC TESTU

DZIĘKUJĘ.

APPENDIX 4A

Texts for PLAT tasks 2, 3, 5 and 6

Text for listening comprehension – tasks 3 and 5

Badania snu.

Do niedawna większość uczonych była zdania, że sen jest mało wdzięcznym obiektem badań naukowych. Nic w tym zresztą dziwnego, gdyż w przeciwieństwie do stanu czuwania, który łatwo jest obserwować i mierzyć, i o którym *sam* badany może udzielić naukowcom informacji, sen jest zjawiskiem niełatwym do śledzenia. Można, co prawda, mierzyć bez trudu oddech, tętno, ciepłotę ciała czy też notować zmiany pozycji podczas snu obserwacje te dotyczą jednak zjawisk ubocznych i nie wyjaśniają istoty snu. Jeśli przedmiotem obserwacji jest głębokość snu, trzeba osobę badaną budzić a przynajmniej zakłócać jej sen. W ten to właśnie sposób w 19 wieku przeprowadzono najwcześniejsze badania nad przebiegiem snu, w rezultacie których fizjolog Kohlschutter ustalił, że pierwszy sen jest najgłębszy.

Nowoczesne badania nad snem datują się dopiero od chwili, gdy można było rejestrować czynność bioelektryczną mózgu. Wynalazek ten spowodował prawdziwy przełom. Obecnie na całym świecie w niezliczonych pracowniach i szpitalach bada się czynność bioelektryczną mózgu. Zapis czynności tej nazywamy elektroencefalogramem, w skrócie: EEG.

Rejestracji czynności bioelektrycznej mózgu z powierzchni głowy dokonał po raz pierwszy Hans Berger, było to w latach dwudziestych ubiegłego stulecia stulecia. Berger był wówczas znanym i cenionym w środowisku lekarskim kierownikiem kliniki dla nerwowo i psychicznie chorych w Jenie, w Niemczech. Wieczory, wolne od pracy w oddziale, spędzał na próbach zarejestrowania bioprądów mózgowych. Posługiwał się w tym celu bardzo prymitywną według współczesnych pojęć aparaturą a mianowicie zestawem galwanometrów. Jako elektrod używał płytek srebrnych, przyklejanych do skóry. Rejestrowana w ten sposób czynność bioelektryczna niewiele się wybijała ponad poziom szumów lichego aparatu. Mimo to udało mu się utrwalić zapis regularnych fal o częstotliwości około 10 cykli na sekundę, występujących w stanie odprężenia psychicznego. Fale te znane są dziś powszechnie pod nazwą rytmu alfa.

Z początku prace Bergera przyjęto z lekceważeniem lub niedowierzaniem. Dopiero gdy Adrian i Matthews, fizjolodzy cieszący się wielkim autorytetem, potwierdzili jego odkrycie w 1934 roku, świat nauki odpłacił Bergerowi należnym uznaniem.

(2' 40") *(adapted from Borbely, A. 1990. "Tajemnice snu" Warszawa:PWN)*

Text for dictation task 2

Dyktando

1. Kierownik ogrodu zoologicznego okazał się karierowiczem.
2. Zwierzęta, których nie cierpiał, traktował tylko jako szczebel do wybicia się.
3. Nie dbał także o należytą rolę swojej placówki w wychowaniu młodzieży.
4. Żyrafa w jego ogrodzie miała krótką szyję a borsuk nie posiadał nawet swojej nory.
5. Niedociągnięcia te nie powinny mieć miejsca, tym bardziej, że ogród bywał często odwiedzany przez wycieczki szkolne.
6. Toteż nic dziwnego, że rozczarowani mieszkańcy zaczęli pisać skargi do rady miasta.

(adapted from Mrożek S. "Słoń")

Text for reading comprehension task 6

ZABILI POPULISTĘ

Pim Fortuyn, przywódca rodzącej się holenderskiej populistycznej prawicy, padł wczoraj ofiarą pierwszego politycznego zamachu w najnowszej historii Holandii.

Od początku marca, kiedy zdobył w lokalnych wyborach w Rotterdamie 35% głosów, notowania Fortuyana rosły. Jego partia miała szanse na co najmniej 15% głosów w wyborach parlamentarnych planowanych na 15 maja. On sam zyskiwał poklask ostrymi wezwaniami do zaprowadzenia porządku na ulicach i walki z napływem imigrantów.
Fortuyn został zastrzelony przez nieznanego sprawcę około godz. 18, gdy chciał wsiąść do swojej limuzyny po skończonym wywiadzie radiowym. Zabójca oddał co najmniej sześć strzałów. W pobliżu miejsca zamachu policja aresztowała osobnika podejrzanego o dokonanie zamachu. Miał broń i próbował się ukryć na stacji benzynowej. Wiadomo tylko, że to biały Holender i że nie chciał rozmawiać z policją

MATURA W GAZECIE
kujon polski

DZIŚ: Jak pisać maturę z polskiego? Patrz w samym środku "Gazety"
Jutro: czat maturalny na www.gazeta.pl

W PIĄTEK: drukujemy wszystkie tematy z polskiego i opinie specjalistów, co należało napisać

W SOBOTĘ: rozwiązania zadań z matematyki

DZIEŃ W SKRÓCIE

Francuskie tempo

Już dzień po wyborach Francja ma nowego premiera. Następcą Jospena został Jean-Pierre Raffarin były współpracownik pre-zydenta Giscarda d'Estainga, zwolennik zakorzenie-nia Francji w UE.

Prareferendum Uni o Uni

Dziś w 114 miastach Polski sondażowe głosowanie w sprawie przystąpienia Pol-ski do UE. Organizatorem jest Unia Wolności

Pierwsza BSE w Polsce

Krasula, córka Sekwany

Urodziła się latem 93 albo 94 roku. Pasła się na łące, żarła trawę, siano, buraki i kiszonkę. Nic więcej. Żadnych "kupnych" pasz ani mączek – opowiada o pierwszej polskiej krowie chorej na BSE jej właściciel.

Michał Bajorek, gospodarz z Rzepiennika Strzyżewskiego 30 km od Tarnowa, wciąż nie potrafi uwierzyć, że *Krasula* miała dziury w mózgu: "Nic jej nie było. Sprzedaliśmy ją tylko dlatego, że nie dało się jej już zacielić."

W oborze Bajorków stoi jeszcze 11-mie-sięczna jałówka, córka *Krasuli*, i dwumie-sięczny cielak kupiony od sąsiada.

Bajorek: "Handlarz wziął *Krasulę* i trzyletnią *Lalkę*, która też jakoś nie chciała się zacielić. Płacił z góry, cenę dawał dobrą."

Policja przypuszcza, że handlarz zdobył dla krów fałszywe dokumenty, bo nie miały 'papierów', a potem sprzedał rzeźni w Moch-naczce Wyżnej. Tam u *Krasuli* wykryto BSE.

Raport Rady Czterech nr 1

Gospodarko, jak się masz

W jakim stanie jest gospodarka? Co jest wart program gospodarczy rządu? Kto ma rację w sporze o wysokosć stóp procentowych? – odpowiadają dziś ekonomiści z Rady Czterech.

Rada jest inicjatywą "Gazety". Tworzą ją ekonomiści o wielkim dorobku, znani z niezależności sądów. Pełny tekst analizy dotyczącej I kwartału – na s. 25.

APPENDIX 4B

Tasks removed from PLAT after pilot

ZADANIE I.

Wyobraź sobie, że na wyrobie pochodzenia zagranicznego znalazłeś ważne ostrzeżenie:

ZABAWKA NIE NADAJE SIĘ DLA DZIECI PONIŻEJ TRZECH LAT.

W każdym z punktów od 1 do 4 znajdź zdanie, które zawiera to samo ostrzeżenie dotyczące dzieci. Zakreśl A, B, C lub D w arkuszu odpowiedzi.

1. A. Si raccomanda di conservarlo orizzontalmente. (włoski)
   B. Hracka neni vhodna pro deti do trech let. (czeski)
   C. La veo aqui todos los dias. (hiszpański)
   D. Wir nehmen den grauen Anzug nicht. (niemiecki

2. A. Skall inte ges till barn under tre ar. (szwedzki)
   B. La tinta de olor es lavable. (hiszpański)
   C. Mijn nieuwe woning is op de derde etage. (holenderski)
   D. Questa casa e sporca e dobbiamo pulirla. (włoski)

3. A. Markers fur weisse Tafel geruchlos. (niemiecki)
   B. Mercadores para pizarra blanca. (hiszpański)
   C. De koffie is koud geworden. (holenderski)
   D. Gioccatolo non andato a minori di tre anni. (włoski)

4. A. Small parts might be swallowed. (angielski)
   B. Ellos se lavan por la manana. (hiszpański)
   C. Igracka ne prikladna za djecu miadu od tre godine. (chorwacki)
   D. Ho preso una bottiglia d'acqua. (włoski)

ZADANIE II

W języku łotewskim występują rzeczowniki o w rodzaju męskim i żeńskim. Przymiotniki zgadzają się z rzeczownikiem co do rodzaju, liczby i przypadku. Przeanalizuj poniższe wyrażenia w języku łotewskim (podane w **mianowniku** i **dopełniaczu)** oraz ich polskie tłumaczenia i wykonaj polecenia 1 i 2.

silts ezers	- ciepłe jezioro	liela kalna	- wielkiej góry
liels ezers	- wielkie jezioro	liels kalns	- wielka góra
sarkana ezera	- czerwonego jeziora	melna putna	- czarnego ptaka
lielas juras	- wielkiego morza	sarkans putns	- czerwony ptak
silta jura	- ciepłe morze	sarkanas lapas	- czerwonego liścia
melns lauva	- czarny lew	liela krasns	- wielki piec
liela lauvas	- wielkiego lwa	sarkana pils	- czerwony zamek
melna kafija	- czarna kawa	melnas kafijas	- czarnej kawy

1. Które wyrazy łotewskie są rzeczownikami a które przymiotnikami?
   A. pierwszy wyraz to rzeczownik a drugi przymiotnik
   B. pierwszy wyraz to przymiotnik a drugi rzeczownik
2. Podziel rzeczowniki łotewskie na dwie grupy tak by w każdej znajdowały się tylko rzeczowniki tego samego rodzaju. Nie musisz wskazywać, który z rodzajów jest męski a który żeński - wystarczy,że wypiszesz poniżej wszystkie rzeczowniki należące do tej samej grupy.

Grupa I.. Grupa II...

ZADANIE III

Poniżej podano zestawy prostych zdań w trzech nie znanych ci językach (A, B, C). Przy każdym zdaniu podano jego tłumaczenie na j. polski. Wykorzystując wszystkie dostępne ci informacje przetłumacz wyróznione zdania od 1 do 10:

Przykład:	a. Jeg har et kokken	Mam kuchnię.
	b. Kokkent er sma.	Kuchnia jest mała.
	c. Jeg har et rum.	Mam pokój.
	d. *Rumet er sma.*	Pokój jest mały.

A.

Mi gader le.	Koń widzi ojca.
Mi gader la.	Koń widział ojca.
de	(on) niesie
1.	Koń niósł ojca.
Gad mier le.	Ojciec widzi konia.
Dez mier le.	Człowiek widzi konia.
2.	Koń widział człowieka.

B.

Ki poi sademane?	Czy to jest chłopiec?
Ki pooni sademane?	Czy to jest królik?
dag	pies
3.	Czy to jest pies?
Mane dag vuf.	Pies szczeka.
Manes dag vufvuf.	Psy szczekaja.
bin	biegnie
4.	Chłopiec biega.

C.

Ta hao.	On jest zdów.
Ta hao ma?	Czy on jest zdrów?
Ni hao.	Jesteś zdrów.
5.	Czy jesteś zdrów?
Ta bu hao.	On nie jest zdrów.
Ta bu hao ma?	Czy on nie jest zdrów?
6.	Czy ty nie jesteś zdrów?
Ta gao.	On jest wysoki
Ta hen gao.	On jest bardzo wysoki.
Ta bu hen gao.	On nie jest bardzo wysoki.
7.	Czy nie jesteś bardzo wysoki?
Ta shi ren.	On jest lekarzem.
Zhe shi shemma?	Co to jest?
Ta yao shemma?	Czego on chce?
8.	Kim on jest?
Ta yao zhe.	On to chce.
9.	On chce lekarza.
10.	Nie jest bardzo zdrowy.

(*z* J. Ozga & E. Tabakowska "Egzamin wstępny z ....)

ZADANIE IV

Przeczytaj bajkę w języku czeskim i wykonaj umieszczone pod nią zadania. Pomimo tego, że nie znasz czeskiego jesteś w stanie zrozumieć tę historię. Uwaga! W czeskim 'h' czyta się jak 'g'.

Martina byste museli znat. Nikdy nebyl z niczim spokojen. Když slunicko na nebi hralo, tu Martin nażikal a prala si, aby jen preslo. A když zase padal dest a lilo jako z konve, tu Martin znova nażikal, jak je to na svete vσ&echno spatne zażizeno, że slunce nesviti. Nikdy mu nebylo nic vhod, nic mu nebylo po chuti, nikdo se Martinovi nicim nezavdecil.

Jednou odpocival Martin na kraji lesa pod vysokym, statnym dubem. A to se vi, że i ted' nespokojene reptal: "Co to tam vlastne roste na dube? Żaludy, malićke żaludy! Je to opravdu k smichu. Na takovem strome by mely rust dyne, velikanske dyne, a ne se plazit po czerne zemi. Eh, divne je to na svete zażizeno, **moc** divne."

V tom nażikani a reptani Martin usnul. A když nejlepe spal, prihnal se vitr, zatrżasl **vetvemi** a kolem Martina spadlymi żaludy jen zabubnovalo. A jeden żalud si namiżil zrovna na Martinuv nos. O je, ted' jste meli videt! Martin prudce vyskoćil, jako by ho **vcela** bodla, zlobne se rozhliżel, trel si pszitom naklepnuty nos a uvidel – żalud!

"Aby te všichni ćerti vzali!" Zalateril Martin. "Jeste je dobre, że na dube rostou jenom nimrave żaludy. Zle mohlo byt s mym nosem, kdyby misto żaludu rostly na dube dyne."

Tak poprve se stalo, że byl Martin spokojen. **Opravdu**, nekedy **staći** male, malinke klepnuti, aby przisel ćlovek k rozumu.

I. Odpowiedz na pytania dotyczące tekstu.

1. Pierwszy akapit opisuje charakter Martina. Napisz krótko jaki był Martin.
2. Gdzie odpoczywał Martin?
3. Z czego się Martin naśmiewał?
4. Co się wydarzyło gdy Martin spał?
5. Dokończ to co Martin powiedział do siebie po przebudzeniu.
   "Dobrze, że na..

II. Jak w języku polskim będą brzmiały podkreślone fragmenty z tekstu?

Przykład: jak je to na svete všechno špatne zażizeno

poprawnie: *jak to na świecie wszystko jest źle zorganizowane*

niepoprawnie: *jak jest to na świecie wszystko szpetnie zarzezane*

III. Co najprawdopodobniej w języku polskim znaczą wyrazy wytłuszczone w tekście?

Przykład: velikanske

poprawnie: *ogromne, wielkie, (wielgachne)*

niepoprawnie: *welikańskie*

11. hralo:	13. vetvemi:	15. opravdu: ....
12. moc:	14. vcela:	16. staći: ....

ZADANIE V

Z tekstu poniżej usunięto ***dwa słowa*** *języka polskiego. Jeden wyraz zastąpiono 8 razy nie istniejącym wyrazem "potrasto". Podaj jakie to będzie słowo w języku polskim i napisz je* ***we właściwej formie.***

Drugie usunięte słowo ma numer (2) w tekście – podaj również co to jest za wyraz.

By wykonać to zadanie najpierw przeczytaj uważnie cały tekst.

Obrzędy i zwyczaje związane z (1) **potrasto** były kultywowane we wszystkich epokach i pod każdą szerokością geograficzną. W wielu kulturach na przykład przez pierwsze lata życia nie (2)______ dzieciom (3) **potrasto**. Małym Słowianom skracano dziewczęce pukle dopiero gdy w wieku kilku lat przechodzili spod skrzydeł matki pod opiekę ojca. Pieczętowano to uroczystym obrzędem postrzyżyn.

(4) **Potrasto** przekazują ważne informacje zmysłowe. Ma to głębokie uzasadnienie. Na linii (5) **potrasto** znajdują się liczne receptory dotyku, poza tym na (6) **potrasto** gromadzą się substancje zapachowe, w tym feromony produkowane przez skórę. Siła oddziaływania kobiecych (7) **potrasto** jest tak wielka, że niektóre religie zakazują ich publicznego pokazywania. Muzułmanki zakrywają je, by uchronić się przed spojrzeniami mężczyzn a Żydówkom po ślubie niegdyś golono głowy. Podobną funkcję pełniły dawniej w Polsce oczepiny - obrzęd, podczas którego (8) **potrasto** świeżo upieczonej mężatki znikały na zawsze pod czepcem lub chustą. Także dziś na podstawie sposobu ułożenia (9) **potrasto** można wiele powiedzieć na temat etnicznego i społecznego pochodzenia człowieka, a także o jego wierzeniach religijnych.

ZADANIE VI

Ponownie przypomnij sobie listę słów w Akareke i ich polskie odpowiedniki.

Uzupełnij niekompletną listę poniżej. Jeszcze raz przeczytaj i zapamiętaj tę listę.

Akareke		***Polski***
minum	-	....
tuka	-	lubić
apal	-	....
duduk	-	siedzieć
tidur	-	....
omong	-	mówić
ingat	-	....
wafat	-	umierać
tertawa	-	....
datang	-	przychodzić

APPENDIX 5

SLP Test KET-A VERSION A

(section 2.5.)

I. Wysłuchasz krótkiego tekstu w 2 częściach. Zakreśl lub wpisz odpowiedzi w miejscach podanych. Każdą część usłyszysz dwa razy.

Część 1: Anita Spencer jest instruktorem 'kursów przetrwania' i omawia z grupą uczestników trasę jaka ich czeka:

route – trasa ; **report at checkpoint A** – zgłosić się na punkcie kontrolnym

1. Can they follow any route that they like?	YES / NO
2. When must they report at checkpoint A?	....
3. When must they report at checkpoint B?	....
4. When can they put their tents up?	....
5. Can they finish the exercise after 10 o'clock tommorow morning?	YES / NO

Część 2: Anita przypomina jeszcze raz co wolno a czego nie wolno uczestnikom:

6. They can't travel by car, bus or train .	TRUE / FALSE
7. They must take some money with them.	TRUE / FALSE
8. They can't sleep in the open air.	TRUE / FALSE
9. They can't separate - they must go in pairs.	TRUE / FALSE

II. Uzupełnij rozmowę między Sarą a Davidem. Wybierz 5 zdań z podanej listy tak by pasowały do dialogu. Wpisz litery od A do H w miejsca od 10 do 14. UWAGA !: Jest więcej zdań niż potrzebujesz!

DIALOG

Sarah: I'm tired. I went to see a late film last night.
David: 10 ....
Sarah: That one with Brad Pitt, the American actor.
David: 11 ....
Sarah: I think so. You liked his last film, didn't you?
David: 12 ....
Sarah: The new one in the city centre.
David: 13 ....
Sarah: Yes, £6, but it's much better than the old one.
David: 14 ....
Sarah: Well, I hope you like it.

ZDANIA DO WYBORU

A. Perhaps I'll go next weekend.
B. Are the tickets expensive there?
C. Oh, what did you see?
D. Fine thanks, and you?
E. Do you think I'd like it?
F. Is the city centre far?
G. Yes, it was really good. Which cinema did you go to?
H. You should go to bed early.

III. Przeczytaj tekst i zaznacz kółkiem prawidłową odpowiedź.

ESTHER'S STORY

When Esther left school at the age of sixteen, her aunt Flory gave her £500 for her birthday. Most of Esther's friends decided to go to College, but Esther used her aunt's money to Start her own business. She bought fruit, sugar and some glass jars and began making her own jam. She sold the jam for £1 a jar to her friends and she soon doubled her aunt's £500.

At first, her parents didn't want Esther to spend her time making jam and they thought that she should study instead. They hoped that one day she would be a teacher or a doctor. But Esther didn't listen to them. She just kept on making jam. After a few months, she started selling it to the local market. Then she started making orange juice. She sold this to a school where one ofher friends worked.

After two years, her business was very large and her parents were very pleased with her. She made all kinds of food which she sold to shops and supermarkets. She was so busy that she had to get some people to work for her.

15. Why did Aunt Flory give Esther some money?
A. Esther asked for it
B. Esther's friends needed it.
C. It was a present.

16. After she left school, Esther
A. went to College.
B. started her own business.
C. worked for her aunt.

17. Why did Esther make jam?
A. She liked to eat it.

B. She had a lot of fruit.
C. She wanted to make money.

18. When Esther left school, her parents wanted her to
A. go to College.
B. be a business woman.
C. work in a market.

19. Esther sold orange juice to
A. the local market.
B. a school.
C. her friends.

20. Esther's parents were happy because she
A. cooked for them.
B. was busy.
C. had a good business.

21. After two years, Esther
A. paid people to help her.
B. worked in a supermarket.
C. opened a shop.

IV. Z tekstu poniżej usunięto 9 słów i zastąpiono je numerami od 22 do 30. Przeczytaj ten tekst i wybierz odpowiednie słowo zakreślając A, B lub C.

A HISTORY OF AIR TRAVEL

In 1783, two French brothers built **....22....** first balloon to take people into the air. One hundred and twenty years later, in 1903, the Wright brothers built the first plane with an engine and flew in it. This was **....23....** the United States. Then, in 1918, the US Post Office **....24....** the first airmail Service. Aeroplanes changed a **....25....** in the next thirty years. Then, in the 1950s, aeroplanes became much **....26....** because they had jet engines.

In 1976, Concorde was built in the UK and France. It is the fastest passenger plane in the world and it **....27....** fly at 2500 kilometres an hour, so the journey **....28....** London to New York is only four hours. Today, millions of people travel **....29....** aeroplane, and it is difficult to think of a world without **....30....** .

22	A	the	B	a	C	one
23	A	in	B	at	C	through
24	A	begun	B	begin	C	began
25	A	lot	B	many	C	few
26	A	fast	B	faster	C	fastest
27	A	must	B	should	C	can
28	A	between	B	from	C	of
29	A	with	B	on	C	by
30	A	them	B	their	C	they

V. Przeczytaj dwa krótkie listy. Wstaw po jednym brakującym wyrazie w wolne miejsca od 31 do 40 umieszczone na dole strony.

List A Dear Sir,

I lost **(Przykład:** *my***)** bag on a train **.....31.....** week. I was on the 8.30 a.m. train to Cambridge **.....32.....** l0th May. It is **.....33.....** large blue bag and my name **.....34.....** written on the outside. **.....35....** you found this bag?

Yours faithfully,

Mary Johnson

List B Dear Ms Johnson,

I **.....36.....** pleased to tell **.....37.....** that we have your bag here. If you come **.....38.....** this office, we **.....39.....** give it to you. The office **.....40.....** at 9 a.m. and closes at 6.30 p.m. every day.

Yours sincerely,

J Wilson

List A: 31............; 32............; 33............; 34.............; 35............

List B: 36............; 37............; 38............; 39.............; 40............

VI. Dopasuj definicje ponumerowane od 41 do 50 do słów od A do N. UWAGA! Jest więcej słów niż definicji !

DEFINICJE	**SŁOWA**
41 You pack your clothes in this.	**A** gloves
42 You stand in this when you wash yourself.	**B** boring
43 This is where you can choose your holiday.	**C** hotel
44 You often do this when you watch a comedy film on TV.	**D** scisors
	E suitcase
45 In winter you put those on your hands.	**F** nurse
46 You sleep inside this on a camp-site.	**G** travel agency
47 Works in a hospital.	**H** knife
48 You use them to cut paper.	**I** laugh
49 You can get from one floor to another in this.	**J** airport
50 Opposite of **"interesting"**.	**K** lift
	L tent
	M shower
	N exciting

41. ...... ; 42. ...... ; 43. ...... ; 44. ...... ; 45. ...... ; 46. ...... ; 47. ...... ; 48. ...... ; 49. ...... ; 50. ......

VII. Dostałeś/aś pocztówkę od swojego przyjaciela z Anglii. W 20 – 25 słowach odpisz mu na jego pytania.

POSTCARD	
I am going to visit your country next-month. Can I see you? Please tell me about the weather there. What clothes should I bring? *Alex*	______________________ ______________________ ______________________ ______________________ ____

Twoja pocztówka do Alexa:

POSTCARD	
....	______________________ ______________________ ______________________ ______________________ ____

Koniec Testu

DZIĘKUJĘ ☺

VERSION B

I. **As in Version A**
II. **As in Version A**
III. **As in Version A**

IV. Z tekstu poniżej usunięto 9 słów i zastąpiono je numerami od 22 do 30. Przeczytaj ten tekst i wybierz odpowiednie słowo zakreślając A, B lub C.

A HISTORY OF AIR TRAVEL

In 1783, two French brothers built **....22....** first balloon to take people into the air. One hundred and twenty years later, in 1903, the Wright brothers built the first plane with an engine and flew in it. This was **....23....** the United States. Then, in 1918, the US Post Office began the first airmail Service. Aeroplanes changed a **....24....** in the next thirty years. Then, in the 1950s, aeroplanes became much **....25....** because they had jet engines.

In 1976, Concorde was built in the UK and France. It is the fastest passenger plane in the world and it **....26....** fly at 2500 kilometres an hour, so the journey **....27....** London to New York is only four hours. Today, millions of people travel **....28....** aeroplane, and it is difficult to think of a world without **....29.... .**

22	A	the	B	a	C	one
23	A	in	B	at	C	through
24	A	lot	B	many	C	few
25	A	fast	B	faster	C	fastest
26	A	must	B	should	C	can
27	A	between	B	from	C	of
28	A	with	B	on	C	by
29	A	them	B	their	C	they

V. Przeczytaj dwa krótkie listy. Wstaw po jednym brakującym wyrazie w wolne miejsca od 30 do 37 umieszczone na dole strony.

List A Dear Sir,

I lost **(Przykład:** *my***)** bag on a train **.....30.....** week. I was on the 8.30 a.m. train to Cambridge **.....31.....** l0th May. It is **.....32.....** large blue bag and my name is written on the outside. Have **.....33.....** found this bag?

Yours faithfully,

Mary Johnson **List A: 30............; 31............; 32............; 33............. .**

List B Dear Ms Johnson,

I **.....34.....** pleased to tell **.....35.....** that we have your bag here. If you come **.....36.....** this office, we give it to you. The office **.....37.....** at 9 a.m. and closes at 6.30 p.m. every day.

Yours sincerely,

J Wilson **List B: 34............; 35............; 36............; 37............. .**

VI. Dopasuj definicje ponumerowane od 38 do 45 do słów od A do M. UWAGA! Jest więcej słów niż definicji !

DEFINICJE

38 You pack your clothes in this.
39 You stand in this when you wash yourself.
40 You often do this when you watch a comedy film on TV.
41 In winter you put those on your hands.
42 You sleep inside this on a camp-site.
43 Works in a hospital.
44 You use them to cut paper.
45 You can get from one floor to another in this.

SŁOWA

A gloves
B airport
C hotel
D scisors
E nurse
F suitcase
G travel agency
H knife
I laugh
J lift
K tent
L shower
M exciting

38. ...... ; 39. ...... ; 40. ...... ; 41. ...... ; 42. ...... ; 43. ...... ; 44. ...... ; 45. ......

VII. As in Version A

VERSION C

I. Usłyszysz mężczyznę opowiadającego o tym co mu się kiedyś przydarzyło w dzieciństwie. Po wysłuchaniu zdecyduj czy zdania od 1 do 8 są prawdziwe czy fałszywe. Nagranie trwa 1 min. i usłyszysz je tylko jeden raz.

1. The boy was six years old. TRUE / FALSE
2. He was with his mother. TRUE / FALSE
3. He saw an old woman in the cellar. TRUE / FALSE
4. The woman had black hair. TRUE / FALSE
5. She had a coloured apron. TRUE / FALSE
6. The boy's mother saw her too. TRUE / FALSE
7. He stayed and talked to her. TRUE / FALSE
8. His parents told the neighbours about it. TRUE / FALSE

II. Usłyszysz jak John i Clara, którzy nie widzieli się od bardzo dawna, rozmawiają na przyjęciu. Zdecyduj które fakty z pierwszej kolumny odnoszą się do John'a a które do Clary? <u>Uwaga! – są 3 zdania, które nie pasują do żadnego z nich; jedno zdanie odnosi się i do John'a i do Clary.</u> Usłyszysz tekst 2 razy. Nagranie trwa 3 min.

	Who...	John	Clara
9	**... has put on weight?**		
10	**... works in the same place?**		
11	**... has had several jobs?**		
12	**... got married?**		
13	**... makes a lot of money?**		
14	**... has got children?**		
15	**... was given a higher position at work?**		
16	**... wanted to work at school?**		
17	**... has stopped smoking?**		
18	**... owns a house?**		
19	**... lives alone?**		
20	**... has become famous?**		
21	**... moved to London?**		
22	**... has little time to enjoy home?**		

III. Przeczytaj tekst i odpowiedz na pytania od 23 do 27 poniżej zakreślając A, B, C lub D.

Winter Driving

Winter is dangerous because it's so difficult to know what is going to happen and accidents take place so easily. Fog can be waiting to meet you over the top of the hill. Ice might be hiding beneath the melting snow, waiting to send you off the road. The car coming towards you may suddenly slide across the road.

Rule Number One for driving on icy road.s is to drive smoothly. Uneven movements can make a car suddenly very difficult to control. So every time you either turn the wheel, touch the brakes or increase your speed, you must be as gentle and as slow as possible. Imagine you are driving with a full cup of hot coffee on the seal next to you. Drive so that you wouldn't spill it.

Rule Number Two is to pay attention to what might happen. The more ice there is, the further down the road you have to look. Test how long il takes to stop by gently biaking. Remember that yuu may be driving more quickly than you think. In general, allow double your normal stopping distance when the road is wet, three times this distance on snow. and even more on ice. Try to stay in control of your car at all times and you will avoid trouble.

23. What is the writer trying to do in this text?
   A. complain about bad winter driving
   B. give information about winter weather
   C. warn people against driving in winter
   D. advise people about safe driving in winter
24. Why would somebody read this text?
   A. to find out about the weather
   B. for information on driving lessons
   C. to learn about better driving
   D. to decide when to travel
25. What does the writer think?
   A. People should avoid driving in the snow.
   B. Drivers should expect problems in winter.
   C. People drive too fast in winter.
   D. Winter drivers should use their brakes less.
26. Why does the writer talk about a cup of coffee?
   A. to explain the importance of smooth movements
   B. because he thinks refreshments are important for drivers
   C. because he wants drivers to be more relaxed
   D. to show how it can be spilled
27. Which traffic sign shows the main idea of the text?

DRIVE CAREFULLY ICE ON ROAD	*REDUCE SPEED NOW* ***FOG AHEAD***	**SLOW DOWN ACCIDENT AHEAD**	DRIVE CAREFULLY **ROAD REPAIRS AHEAD**

IV. Przeczytaj cały tekst i zdecyduj czy zdania poniżej są prawdziwe czy fałszywe.

A JOURNEY ALONG THE BEAUTIFUL DOURO RIVER

<u>7 nights from</u>

£695.00

VOYAGES JULES VERNE operate a 'hotel ship' along the Douro river in Portugal. The MV *Lady Ivy May* can take 160 guests in double cabins, all of which face outside and have a private shower and WC. On board the ship, which has air-conditioning, you will find a sun-deck, lounge, bar, dining room, shop, and library facilities.

Itinerary

Day 1 Depart in the early evening from London Heathrow to Oporto. Your guide will meet you on arrival and take you to the *Lady Ivy May,* where you will spend the night.

Day 2 After a morning's sightseeing in Oporto, you will return to the ship and depart for Entre-os-Rios. This part of the journey up the river takes four hours.

Day 3 You will continue travelling up the river. In the early evening, the ship stops at Peso da Regua, where port wine is produced. At dinner you will be able to try the delicious food and drink from this area.

Day 4 After breakfast you will travel south by bus to the ancient town of Lamego and visit the cathedral, a few churches and a museum. The museum was formerly a palace and now has an excellent collection of paintings, tapestries and sculptures. You will then return to the ship and sail on to Tua.

Day 5 At this point the Douro becomes very narrow. Depending on the depth of the river at the time, you may be able to continue by a smaller boat to the Spanish frontier at Barcad'Alva. The return journey to Tua is by coach and there is much to see along the way.

Day 6 In the morning you will drive to Sao Joao da Pesqueira for one of the most wonderful views in the whole of the Douro valley. You will return to the ship for lunch and then join the Douro river valley railway for a beautiful ride through the countryside to Regua, where the *Lady Ivy May* will be waiting for you.

Day 7 There will be a trip to Vila Real before returning to the ship for lunch and then setting off down the river to Oporto. You will sleep on board the *Lady Ivy May.*

Day 8 You will arrive in Oporto in plenty of time for independent sightseeing and last-minute shopping, before you catch the flight home to London Heathrow**.**

Departure dates and prices (all prices are per person)

June 5,12,19,26	£770
July 3,10,17,24,31	£695
August 7,14,21,28	£695
September 4,11,18,25	£820

Prices include:
return flights, 7 nights' accommodation on board the *Lady Ivy May* with all meals, excursions and guides.

Not Included:
travel insurance, tips.

VOYAGES JULES VERNE
TEL: 0171 723 5066

28. Each cabin on the *Lady Ivy May* is for two people. TRUE / FALSE
29. You can borrow books on board the ship. TRUE / FALSE
30. On arrival at Oporto, guests find their own way to the ship. TRUE / FALSE
31. You spend a day looking round Oporto. TRUE / FALSE
32. It takes a day to travel from Oporto to Entre-os-Rios. TRUE / FALSE
33. The museum at Lamego used to be a palace. TRUE / FALSE
34. It is sometimes possible for the *Lady Ivy May* to sail to the Spanish border. TRUE / FALSE
35. The trip includes some travel by train. TRUE / FALSE
36. You arrive back in Oporto on the day before your return flight. TRUE / FALSE
37. *Voyages Jules Verne* arrange your travel insurance. TRUE / FALSE

V. Przeczytaj tekst i zdecydój, który z czterech podanych wyrazów pasuje w miejsca od 38 do 48.

Zakreśl A, B, C lub D lub wpisz ją w miejsca wykropkowane.

Przykład: 0 = B

KEEPING YOUR CAR SAFE

Apart from your home, your car is probably your most **(0) ....B....** possession. It's also your most vulnerable. Car and cassette and radio thefts make up over a quarter of all recorded crime. Together they **(38) ....** everyone a lot of money. It takes up police time to deal with these offences, taking criminals through the **(39) ....** is expensive and, of course, motorists end up paying higher insurance premiums.

Over 460,000 cars are reported **(40) ....** in Britain each year and many are never found again. Many of those which are found have been **(41) ....** by thieves. A **(42) ....** car is also far more likely to be involved in an accident than the same car driven by its **(43) ....;** car thieves are often young and sometimes drunk.

Yet car crime can be cut drastically if motorists follow a few simple **(44) ....** to keep thieves out of their cars in the first place. Most car thieves are unskilled criminals and many are under twenty. So make your own car a less inviting target, to **(45) ....** thieves from trying.

Lock your car every time you leave it even for a **(46) ....** minutes. Astonishingly one car in every five is left unlocked. Don't leave valuables on display inside or briefcases or coats which a thief might think **(47)..........** money. If you can't take valuables with you, lock them out of **(48).........** in the boot of your car.

	A	rich	B	valuable	C	worthy	D	wealthy
38	A	cost	B	spend	C	pay	D	charge
39	A	judgements	B	judges	C	courts	D	juries
40	A	disappeared	B	lost	C	gone	D	missing
41	A	injured	B	hurt	C	damaged	D	harmed
42	A	new	B	stolen	C	taken	D	borrowed
43	A	owner	B	holder	C	master	D	landlord
44	A	announcements	B	warnings	C	laws	D	rules
45	A	put down	B	put off	C	get out	D	get up
46	A	little	B	bit	C	few	D	five
47	A	have	B	contain	C	enclose	D	include
48	A	eye	B	hand	C	reach	D	sight

VI. Przeczytaj tekst i z podanych obok wyrazów utwórz wyrazy pochodne tak by pasowały w miejsca ponumerowane od 49 do 55. Wpisz utworzone wyrazy w miejsca wykropkowane.

PUPPET SHOWS

Puppets are dolls representing **(0)** ..... or modern characters in stories. **TRADITION** traditional
They are a popular form of **(49)** ..... for both children and adults. **ENTERTAIN** ....
Some puppets seen in Europe today were **(50)** ..... created in Italy **ORIGIN** ....
in the 1 500s. A puppet show was an inexpensive way to enjoy yourself
so people could go to see the same **(51)** ..... again and again. **PERFORM** ....
So what is the continuing **(52)** ..... of the puppet show? Used **ATTRACT** ....
cleverly by artists, puppets have a living magic. However beautiful,
funny or **(53)** .... to look at, they are really simple tools used by **FASCINATE** ....
artists to express their ideas, **(54)** ..... and character. Of course they **PERSON** ....
must have great technical **(55)** ..... , but the fun and charm come from **ABLE** ....
the way the artists establish comunication with the audience.

VII. Dokończ zdania tak aby znaczyły mniej więcej to samo co zdania podane. Patrz przykład.

Przykład: **When Ann has to walk to the supermarket she finds it tiring.** **get**

Ann gets tired when she has to walk to the supermarket.

56. I haven't smoked for three years. gave
 I .... three years ago.
57. My advice for you is to buy a computer. ought
 You ....a computer.
58. Psychology interests me a lot. am
 I ....psychology.
59. You can't smoke in here. allowed
 You .... smoke in here.
60. Our driving laws and theirs are not the same. different
 Our driving laws ....theirs.
61. It is really not necessary for you to speak to him now. have
 You .... speak to him now.
62. He hates travelling by bus. stand
 He .... by bus.
63. She is often driven to school by her boyfriend. lift
 Her boyfriend often .... her .... to school.
64. Mandy is excited about going to Greece on holiday. forward
 Mandy is .... to going to Grece on holiday.
65. Maria didn't listen carefully to the news broadcast. attention
 Maria didn't ....to the news broadcast.

VIII. Napisz list do kolegi/koleżanki z Anglii, w którym opisujesz swoje wrażenia z pobytu weekendowego z kilkoma przyjaciółmi. W liście umieść następujące informacje:

a. gdzie spędziłeś ten weekend,
b. co robiłeś w czasie pobytu
c. dlaczego tak Ci się bardzo podobało.

Napisz nie więcej niż 100 słów.

Dear ....,

I had a really wonderful day last week. ....

....

VERSION D (German)

I. Reporter przeprowadza sondaż uliczny na temat możliwości spędzania czasu wolnego przez młodzież. Wysłuchasz dwóch wypowiedzi. Zakreśl czy zdania są zgodne z treścią wywiadu , czy nie.

Wypowiedź 1 : Frau Gudrun Reich, 73, Rentnerin

1. Die alte Dame wohnt im Stadtzentrum... RICHTIG / FALSCH
2. Die Jugendlichen verhalten sich laut in den öffentlichen Plätzen.... RICHTIG / FALSCH
3. Die älteren Leute haben Angst vor den Jugendlichen....................... RICHTIG / FALSCH
4. In der Nähe gibt es kein Jugendzentrum.. RICHTIG / FALSCH

Wypowiedź 2 : Jens Möller, 16, Schüler

5. Es gibt viele Spielplätze für kleine Kinder..................................... RICHTIG / FALSCH
6. Es gibt auch viele Freizeitangebote für Jugendliche....................... RICHTIG / FALSCH
7. Sie dürfen sogar im Park nicht ruhig den Fußball spielen............... RICHTIG / FALSCH
8. Leute regen sich auf, wenn sie Skateboard fahren.......................... RICHTIG / FALSCH
9. Die Jugendlichen brauchen einen Platz , wo sie andere nicht stören.. RICHTIG / FALSCH

II. Uzupełnij dialog wybierając odpowiednie zdania od A do G. Wstaw je w miejsca od 10 do 14. Jest więcej zdań niż potrzebujesz.

A: Jugendherberge Anabella, Guten Tag
B: ....
A: Ja, natürlich. Wann kommen Sie?
B: ....
A: Wie lange wollen Sie bleiben?
B: ....
A: 20 Euro.
B: ....
A: Nur das Fruhstuck
B: ....
A: Bis zum 15. August.

A. Wir bringen Schlafsäcke mit.!
B. Was ist im Preis inbegriffen
C. Am 20. August.
D Bis wann soll ich die Reservierung bestätigen?
E Bis zum 28 August. Was kostet eine Übernachtung?
F. Bei uns bekommen Sie warmes und kaltes Essen.
G. Guten Tag! Kann ich bei Ihnen ein Zimmer reservieren ?

III. Przeczytaj tekst i zaznacz kółkiem prawidłową odpowiedź.

MACHT SCHULE KRANK?

Schon in der Früh geht der Stress los: die Schüler müssen aufstehen und frühstücken.
Sie sind durch das viele Lernen unausgeschlafen. Zuerst müssen sie den halben Tag in der Schule sein, dann müssen sie den ganzen Nachmittag lernen, und abends sitzen sie noch vor dem Fernseher. Dann ist der Tag vorbei. Sie haben viele Tests und die Lehrer fragen sie ab. Oft ist der Schulstress schwer zu ertragen. Heute haben die Schüler zu viele Termine und zu viele Aufgaben. Nachmittags haben sie auch keine Freizeit, denn sie nehmen auch an verschiedenen Kursen teil . Auf dem Schulhof dürfen die Schüler nicht herumstehen, sondern sie sollen sich bewegen, aber nicht rennen. Viele Schüler fahren zur Schule mit dem Bus und einem schlechten Gewissen, weil sie von ihren Schulkameraden die Hausaufgaben abschreiben. Sogar in den Ferien müssen sie das Material wiederholen, die Pflichtlektüre lesen und zusätzliche Matheaufgaben lösen. Viele Schüler halten den Stress nicht aus. Immer mehr Ärzte warnen vor Drogen. Das Schlimmste an der Sache ist ,dass die Drogensüchtigen immer jünger werden.

15. Die Schüler müssen viel lernen, deshalb
A. stehen sie früh auf.
B. frühstücken sie nicht
C. schlafen sie wenig

16.Die meisten Schüler verbringen den halben Tag in der Schule und dann
A. sehen sie den ganzen Nachmittag fern .
B. lernen sie am Nachmittag zu Hause.
C. haben sie frei.

17. Sie haben keine Freizeit, weil
A. sie zu viele Termine haben
B. sie die Zeit nicht organisieren können
C. sie im Stress leben

18. Die Schüler verbringen die Pausen
A. auf dem Schulhof.
B. in den Klassenräumen.
C. Mit dem Schulleiter

.

19. Viele Schüler haben ein schlechtes Gewissen, weil sie
A. mit dem Bus zur Schule fahren.
B. die Hausaufgaben nicht allein machen.
C. sich gut für den Unterricht vorbereiten.

20. In den Ferien sollen die Schüler
A. sich erholen.
B. weiter lernen.
C. sich viel bewegen.

21. Die Ärzte warnen vor
A. Drogen
B. Lehrern
C. Eltern

IV. Z tekstu poniżej usunięto 9 słów i zastąpiono je numerami od 22 do 30. Przeczytaj ten tekst i wybierz odpowiednie słowo zakreślając A, B lub C.

Liebe Katrin!

Unsere Klasse plante seit langem eine Klassenfahrt **22** Grüne. Wir wollten unbedingt mit **23** fahren und **24** Jugendherberge wohnen. Das machten wir auch . Zwei Jungen **25** ihre Gitarren mit und wir sangen im Zug.

26 Stimmung war fantastisch .Dann wanderten wir zwei Stunden **27** Wald. Todmüde kamen wir **28** Ziel an. Unsere Jugendherberge war bequem und die Schlafräume gemütlich. Wir spielten oft Ball , machten Wanderungen und lernten auch viel .Alle waren zufrieden. Lange noch bleibt diese Klassenfahrt ein Gesprächsthema in unserer Klasse.

29, schreib **mir** mal!

Ich grüße **30** Deine Renate

22.	A	in den	B	ins	C	in die
23.	A	dem Zug	B	den Zug	C	der Zug
24.	A	in einem	B	zu einer	C	in einer
25.	A	nehmen	B	genommen	C	nahmen
26.	A	der	B	die	C	das
27.	A	durch die	B	durch den	C	in den
28.	A	ans	B	ins	C	am
29.	A	Dass, du Zeit hast	B	Ob du Zeit hast	C	Wenn, du Zeit hast,
30.	A	dir	B	dich	C	ihr

V. Przeczytaj krótki tekst. Wstaw po jednym brakującym wyrazie w miejsca od 31 do 40.

31 Wochenende bin ich **32** der ganzen Familie **33** Poznań gefahren. Zuerst sind wir **34** Museum gegangen. Dann haben wir ein Geschenk **35** unsere Oma **36** Geburtstag gekauft. Nachher sind wir noch **37** Cafe gegangen und wir haben Kaffee **38**. Am späten Nachmittag sind wir nach Hause zurückgekehrt. Ich habe noch schnell Ordnung in meinem Zimmer gemacht. Am Abend bin ich mit **39** Hund spazieren gegangen. **40** 23 Uhr bin ich ins Bett gegangen.

VI. Dopasuj słowa od A do N do zdań! Jest więcej słów niż potrzebujesz.

41. Herr Rohrer ist im Restaurant. Der Kellner bringt ihm eine ....

42. Ich möchte mir die Hände waschen und brauche eine ....
43. Ich bin müde .Ich gehe gleich ....
44. Wenn ich Ferien auf einem Campingplatz verbringe, schlafe ich in einem ....
45. Ich habe heute nichts gegessen, deshalb habe ich großen ....
46. Ich hänge den Mantel in den ....
47. Die Verkäuferin arbeitet im ....
48. Wenn ich .... habe, nehme ich eine Tablette ein.
49. Ich gehe heute zu einer Party. Ich ziehe mein neues................an.
50.Es regnet und ich nehme meinen ....mit

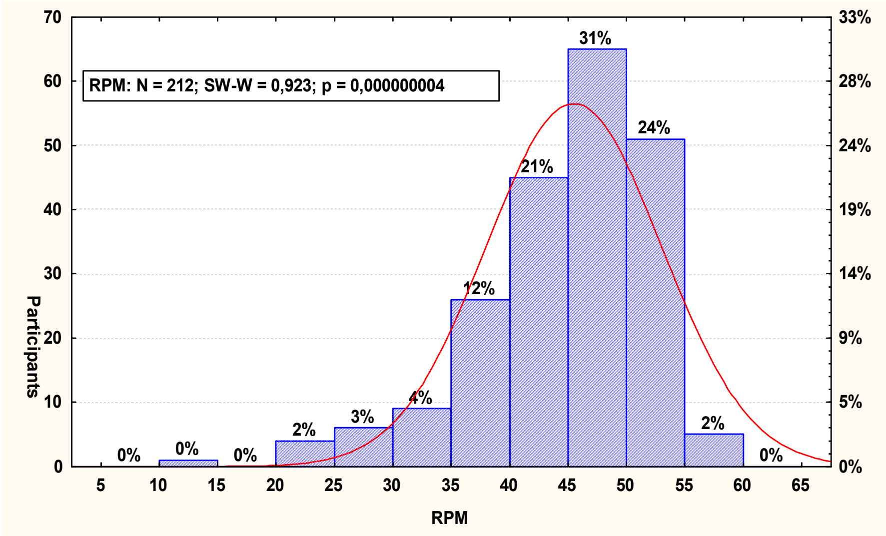
A. Durst B. Schrank C. Regenschirm D. Speisekarte E. Kleid F. Kopfschmerzen G. Seife H fernsehen. I. Kleid J. Zelt K. schlafen L. Rathaus M. Hunger N. Warenhaus

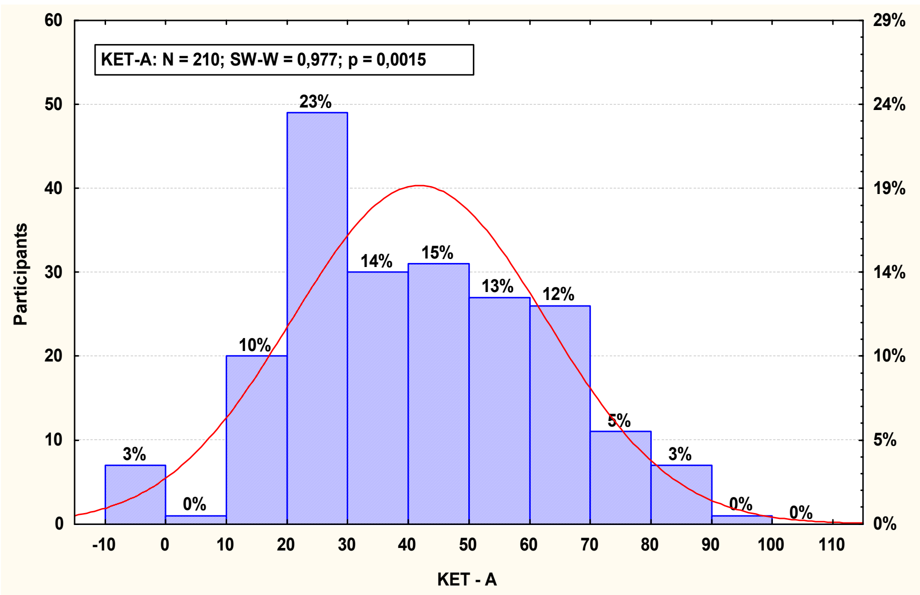
VII. Jesteś na wakacjach w Hiszpanii. Napisz kartkę do rodziców (20 – 25 słów). W kartce umieść następujące informacje:

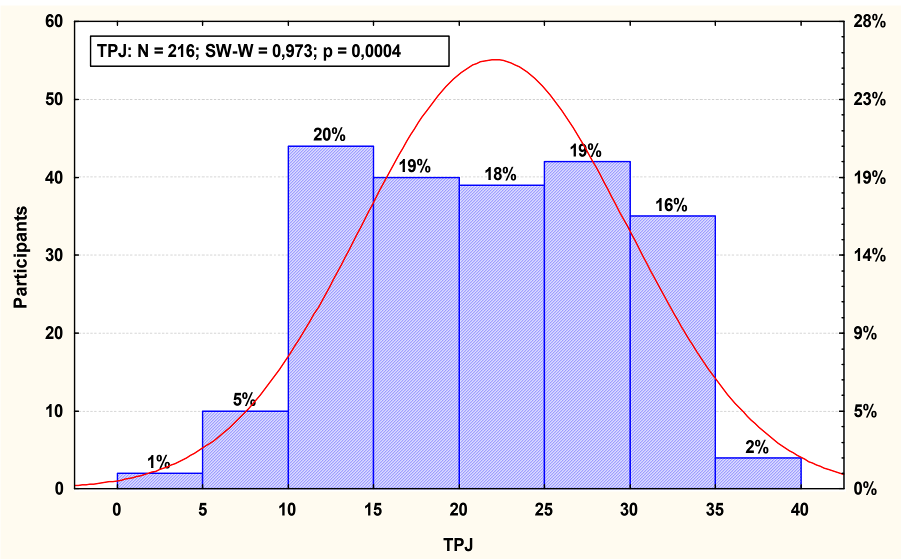
1. jak spędzasz tam czas;
2. jaka jest pogoda
3. jak smakuje ci jedzenie.

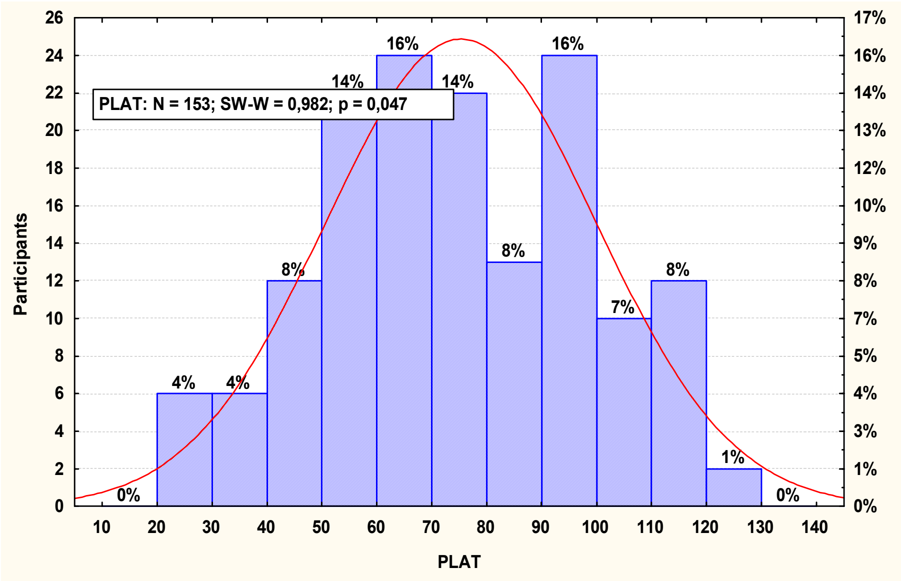
APPENDIX 6

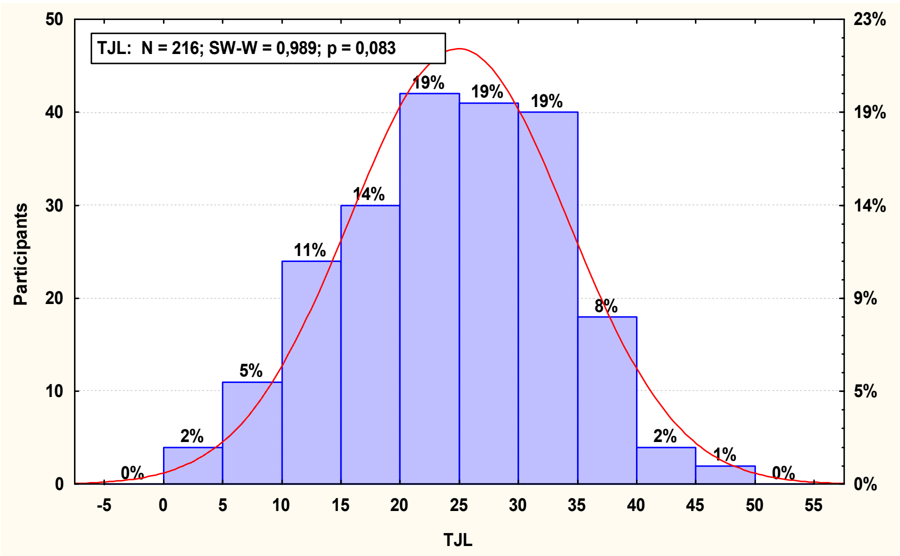
Histograms of five test measures. (section 3.2.3.)











StatSoft, Inc. (2003). STATISTICA (data analysis software system), version 6. www.statsoft.com.

APPENDIX 7

Reproduced correlations for PVoc, AVoc & RPM (A,B,C,D,E)
2-factor solution in Table 3.8.

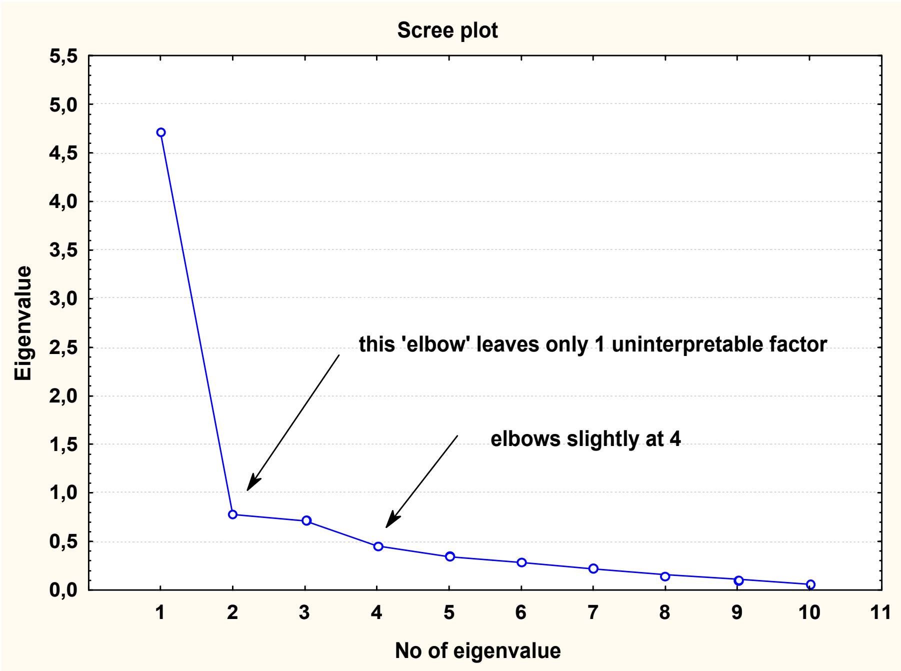
(section 3.5.)

Variables	Extraction: principal axis (PF); N=211						
	PVoc	AVoc	RPM A	RPM B	RPM C	RPM D	RPM E
PVoc	0,78						
AVoc	0,77	0,77					
RPM A	0,39	0,35	0,40				
RPM B	0,36	0,31	0,43	0,48			
RPM C	0,43	0,38	0,50	0,54	0,62		
RPM D	0,40	0,35	0,47	0,52	0,59	0,57	
RPM E	0,41	0,37	0,41	0,45	0,51	0,49	0,43

APPENDIX 8

Cattell's 'scree' test for PLAT

(section 3.6.3.)



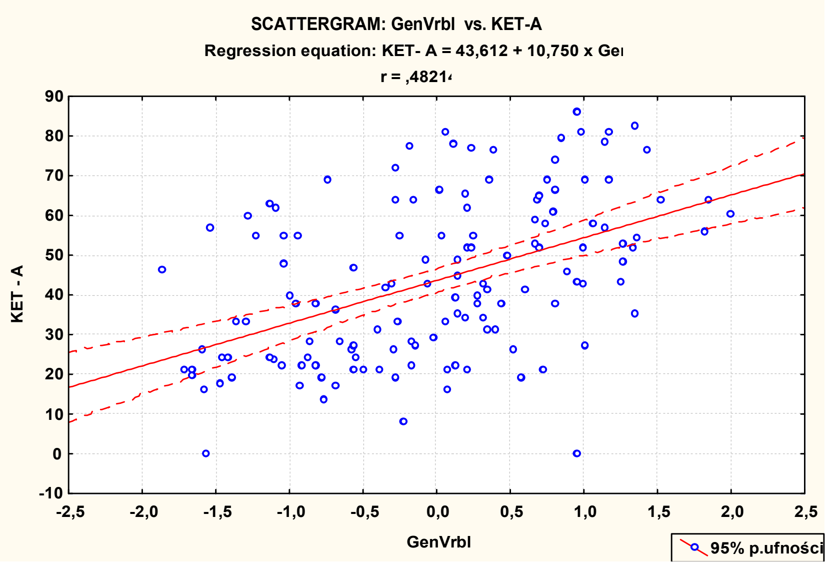
Two and four-factor rotated solutions for PLAT

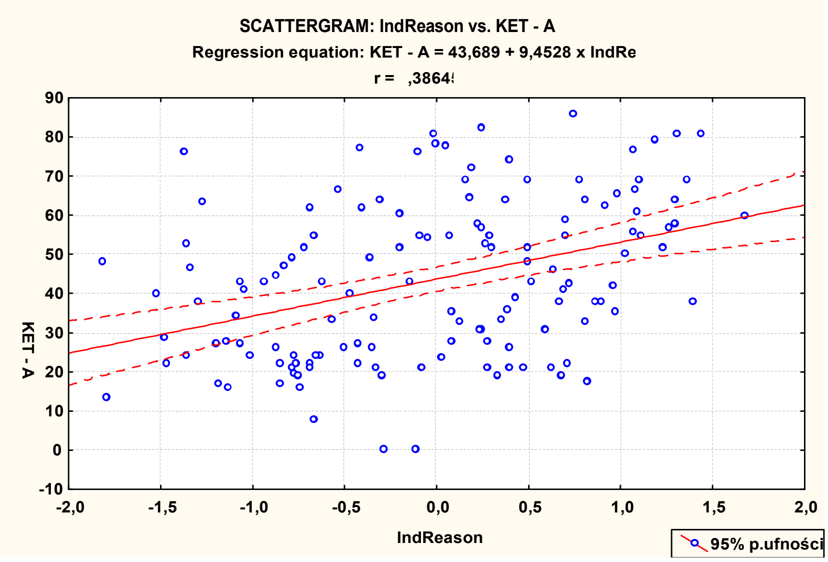
Variables	Extraction: principal axis (PF); N=153 Factor loadings >.500 in bold 2-factor solution		4-factor solution			
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 3	Factor 4
ofiara	**0,621**	0,277	0,230	0,191	**0,888**	0,091
handlarz	**0,528**	-0,027	0,302	-0,114	0,473	0,102
układ	**0,614**	0,234	0,433	0,151	**0,511**	-0,027
przed	0,291	0,100	0,154	0,072	0,192	**0,576**
podczas	0,391	0,101	0,239	0,054	0,298	0,259
lubić	0,176	**0,734**	0,174	0,730	0,196	0,016
pamiętać	0,300	0,634	0,308	0,571	0,199	0,068
dyktando	0,654	0,096	0,614	0,024	0,234	0,197
windglings	0,403	0,150	0,399	0,099	0,182	-0,076
sprawcy	0,580	0,102	0,503	0,032	0,308	-0,004
pojezierze	0,758	0,222	0,692	0,133	0,374	0,001
dbrdszn	0,715	0,299	0,803	0,234	0,162	0,112
zwolennik	0,701	0,223	0,732	0,145	0,229	0,032
expl. variance	**30.0%**	**10.2%**	**23.0%**	**8.0%**	**14.4%**	**3.7%**

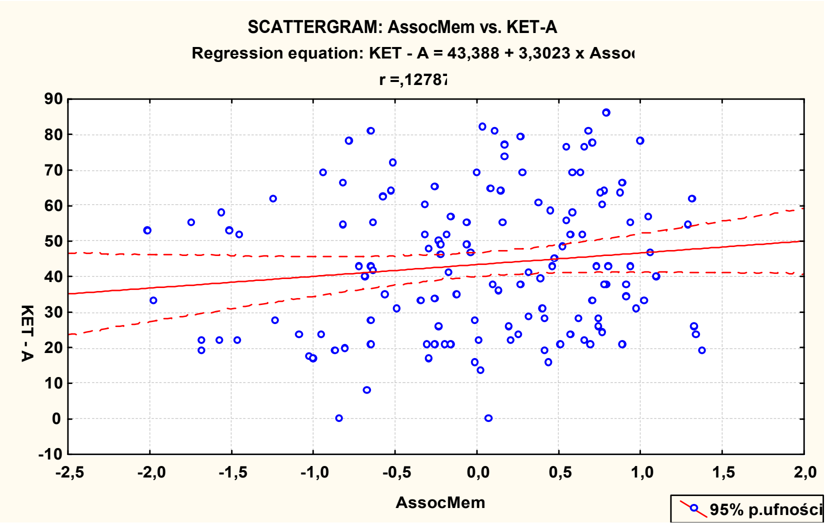
StatSoft, Inc. (2003). STATISTICA (data analysis software system), version 6. www.statsoft.com.

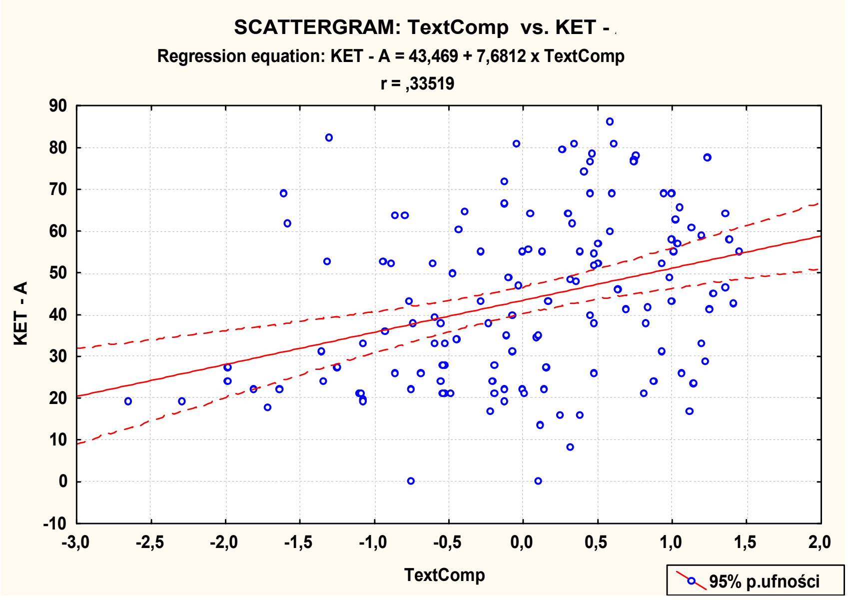
APPENDIX 9

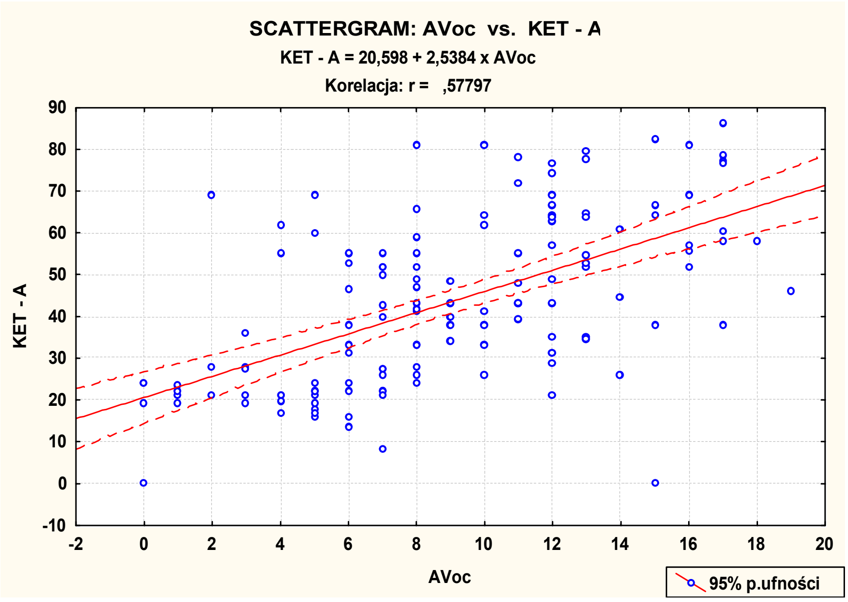
Scattergrams for six predictors and the criterion
(section 3.9.1.)

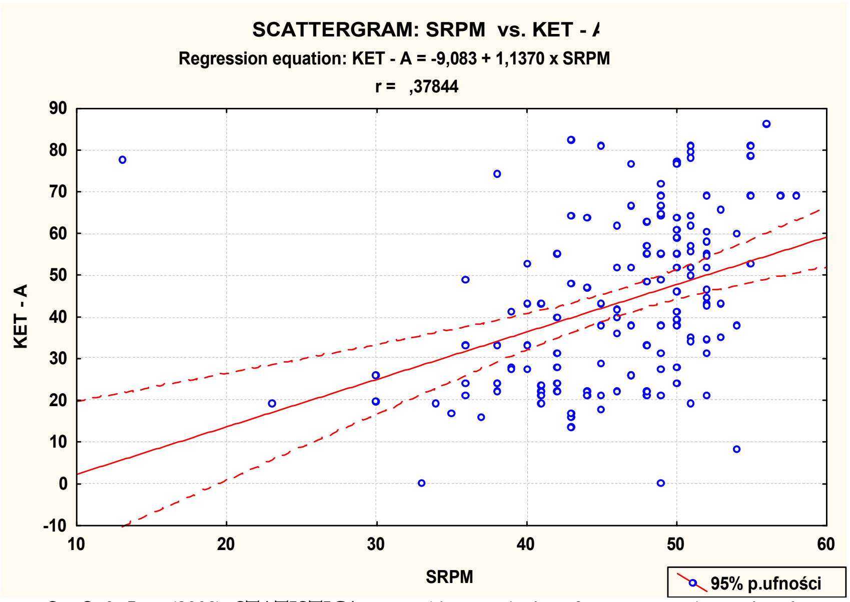












StatSoft, Inc. (2003). STATISTICA (data analysis software system), version 6. www.statsoft.com.

APPENDIX 10

Results of multiple regression analysis for L2 aptitude variables
(Chapter 3.9.2.)

Variable	Correlation matrix for L2 predictors and the criterion					
	GenVerbal	IndReason	AssocMem	TextCompr	SRPM	AVOC
IndReason	0,12					
AssocMem	0,12	-0,03				
TextCompr	0,10	0,12	0,00			
SRPM	0,50	0,40	0,25	0,29		
AVOC	0,65	0,27	0,14	0,46	0,55	
KET - A	0,56	0,39	0,15	0,41	0,56	0,70

Summary of multiple regress.	R=0,766; R^2=0,586; Adjusted R^2=0,565 F (6,118)=27,860; p<0,0001; SEE = 12,194					
N = 125	BETA	St. Error BETA	B	St. Error B	t (124)	p level
Constant			15,304	11,294	1,355	0,178
GenVerbl	**0,233**	**0,086**	**4,691**	**1,738**	**2,699**	**0,008**
InducReaso	**0,203**	**0,067**	**4,469**	**1,470**	**3,040**	**0,003**
AssocMem	0,048	0,062	1,127	1,451	0,777	0,439
TextCompr	**0,166**	**0,071**	**3,474**	**1,484**	**2,340**	**0,021**
SRPM	0,107	0,081	0,317	0,241	1,317	0,190
AVOC	**0,348**	**0,096**	**1,385**	**0,383**	**3,620**	**0,000**

	Variables in the equation						
N = 125	Beta in	Partial Cor.	Semiprtl Cor.	Toleran.	R square	t(124)	p level
GenVerbl	**0,233**	**0,241**	**0,160**	**0,470**	**0,530**	**2,699**	**0,008**
InducReaso	**0,203**	**0,270**	**0,180**	**0,787**	**0,213**	**3,040**	**0,003**
AssocMem	0,048	0,071	0,046	0,908	0,092	0,777	0,439
TextCompr	**0,166**	**0,211**	**0,139**	**0,695**	**0,305**	**2,340**	**0,021**
SRPM	0,107	0,120	0,078	0,534	0,466	1,317	0,190
AVOC	**0,348**	**0,316**	**0,214**	**0,380**	**0,620**	**3,620**	**0,000**

Analysis of variance					
	Sum of squares	df	Mean square	F	p level
Regression	**24857,6**	**6**	**4142,9**	**27,860**	**0,000**
Residual	17547,2	118	148,7		
Total	42404,8				