Self-assessmnent of lexical knowledge in second language vocabulary acquisition.

Calder, Maryna

Self –Assessment of Lexical Knowledge in Second Language Vocabulary Acquisition

Maryna Calder

Submitted to Swansea University

in fulfilment of the requirements for the

Degree of Doctor of Philosophy

Swansea

2013

Acknowledgements

To my supervisor Professor Paul Meara

DEAR PAUL,

Please accept my sincerest gratitude for your support and inspiration in my research work. I greatly appreciate your advice, encouragement and personal guidance. Your detailed and constructive comments showed me a new direction in developing my thesis. I always looked forward to our meetings and extensive discussions in the Chapter Arts Centre. This enabled me to develop a better understanding of the subject and influenced my work.

It has been an invaluable experience for me to carry out my research under your supervision which I hope to pass onto my students.

To my family

Thank you for your understanding and constant support.

Abstract

This thesis explores how learners of Russian as a foreign language assess and categorize their own knowledge of individual words. We start with a brief statement of our intentions and outline how this thesis develops. The first chapter reviews general research on self-assessment in second language acquisition showing contradictory opinions on the use of this approach. Some scholars report positive correlations between self-assessment and other types of assessment, whilst others appear to be sceptical about implementing self-assessment in second language acquisition. This chapter also reviews scarce research on self-assessment in L2 vocabulary acquisition pointing at the limitations of the self-assessment instruments used.

Chapter 2 discusses a self-assessment methodology (VKS) suggested by Paribakht and Wesche (1993). We point out that all the major work that has been carried out on self-assessment of L2 lexical knowledge (its breadth and depth) depended on one tool which is VKS. Chapter 2 offers a detailed analysis of the VKS approach used by Paribakht and Wesche themselves and its modified versions applied by some other researchers. We discuss a number of reasons explaining why VKS might not be the right approach in measuring L2 lexical knowledge.

Chapters 3 through 8 report the empirical research which was carried out to explore how Russian L2 learners categorize their own knowledge of words. We were interested to investigate whether learners' ways of measuring and categorizing their own knowledge of words would mirror the self-assessment categories devised by Paribakht and Wesche in terms of number and type of categories. We ran a series of 6 case studies (2 small group studies followed by 4 single subject studies). In each study, we used the same basic procedure which required participants to arrange Russian words, presented on individual cards, into groups according to how well they knew them. They were asked to describe in detail each of the categories within the classifications they created.

In studies 6 through 8, we also investigated whether exposure to targeted words via reading would influence the way learners categorise their knowledge of those words. Within this issue, we also discuss re-location of words among the categories suggested at different times of testing.

We conclude that learners' categorization of their own lexical knowledge is much more complex and varied than suggested by the levels within the VKS methodology. We also point out that learners' classification systems are very unstable with individual lexical items constantly moving between perpetually changing categories as learners' knowledge changes.

DECLARATION

This work has not previously been accepted in substance for any degree and is not currently being submitted in candidature for any degree.

Signed Maryna Calder

Date 30 October 2011

STATEMENT 1

This thesis is the result of my own investigation except where otherwise stated. Other sources are acknowledged by references. A bibliography is provided.

Signed Maryna Calder

Date 30 October 2011

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan and for the title and summary to be made available to outside organisations.

Signed Maryna Calder

Date 30 October 2011

Table of contents

CHAPTER 1

INTRODUCTION

1.1 Statement of the Intentions

In this thesis, we intend to explore self-assessment (including self-measurement and self-categorization) in L2 vocabulary acquisition and testing. Self-assessment in L2 lexical acquisition is a worthwhile area of research due to a number of reasons. First of all, it should be noted that the purpose of self-assessment testing is to make learners aware of positive and negative trends within their L2 lexical knowledge. To put it simply, using a self-assessment approach regularly L2 learners became aware of how well they know words and what gaps in their lexical knowledge they have to fill in. In other words, L2 learners self-monitor the development of their own lexical knowledge. The feedback received via self-assessment enables them to make judgements about the accuracy of their lexical performance and generally about the state of their L2 vocabulary knowledge. In this sense, self-assessment promotes vocabulary acquisition, in general and might be considered as a supplement or even alternative to an external assessment.

This suggests that self-assessment in L2 lexical acquisition is a worthwhile area of research. However, despite the apparent importance of this approach, there has not been much work carried out on self-assessment in vocabulary testing. Although there is a fair amount of research dedicated to self-assessment in second/foreign language acquisition, self-assessment of lexical knowledge does not seem to receive much attention in the field.

All these explain our interest towards the self-assessment approach in L2 vocabulary acquisition and define the topic of this research.

Initially, we set out to investigate whether learners of Russian would be able to accurately assess their own knowledge of selected Russian words. Our first empirical study (later regarded as a pilot study and reported in Appendix) aimed to establish validity and reliability of L2 learners' self-assessment in testing their vocabulary knowledge. As noted earlier we were interested to explore this issue due to a relative shortage of research into the issue of L2 lexical self-assessment and rather controversial results reported from the use of this approach.

The results obtained in the pilot study, reported in Appendix, indicated that learners of Russian effectively carried out an assessment of their own lexical knowledge and provided correct and accurate information regarding their knowledge of the words tested.

The study did not just illustrate the effectiveness of L2 lexical self-assessment but also revealed learners' ability to describe their knowledge of words. From that point we were interested to further investigate how Russian L2 learners describe and categorize their knowledge of words. This has never been explored in other research in the field. However, the importance of this data seems obvious since it might assist in understanding how lexical self-assessment is carried out. Furthermore, we believe that self-assessment used as a diagnostic instrument might provide some valuable data in regards to how L2 words are acquired and stored in the mental lexicon. In other words, the information collected via self-classification of L2 vocabulary knowledge might shed light on how this knowledge forms and develops.

This determined our decision to expand the focus of our research and explore how Russian L2 learners categorize their own knowledge of words. We planned a series of case studies to investigate this issue. Whilst exploring these questions we noticed that learners' descriptions of their knowledge of the target words change over time, with some of the target words relocating among the changing self-assessment categories. Having recorded that, we were interested to establish whether further exposure to the target lexical items, for example via reading, would lead to further changes in learners' categorization systems of their vocabulary knowledge.

We felt that the repeated longitudinal case study was the right method to investigate whether repeated exposure to particular words would lead to considerable changes in the number and types of learners' self-assessment categories. Via this method, we also intended to trace possible movements of words between the self-rate categories (the issue that surfaced earlier in our research) and determine patterns of these re-locations. We were interested to further explore this issue since we recorded some kind of links between changes in learners' self-categorization systems created at different times of testing and changes in their vocabulary knowledge. In other words, we assumed that the data collected by means of self-assessment might provide some insight into how L2 words are acquired and pro/regress in the mental lexicon.

1.2 Aims of the thesis

There are two main aims of this research. The first aim is to investigate whether the Vocabulary Knowledge Scale devised by Paribakht and Wesche can be used as an ultimate self-assessment instrument for measuring breadth and depth of word knowledge. The second aim of this study is to explore how L2 learners assess the quality of their own knowledge of words. The additional aim of this research is to analyse how knowledge of words acquired changes over time. The more general aim of this thesis is to contribute to research on L2 vocabulary acquisition and assessment.

1.3 Outline of the thesis

The thesis can be divided into three main parts: Literature Review (Chapters 1 and 2); Experimental studies (Chapters 3 through 8); Discussion and conclusion (Chapters 9 and 10).

In the first part, Chapter 1 explains our intentions in undertaking this research. It also presents the background for this thesis by summarizing the existing studies on self-assessment in L2 learning, and in vocabulary testing in particular. Chapter 2 reviews Paribakht and Wesche's studies which use the Vocabulary Knowledge Scale self-assessment methodology devised by the authors themselves. It also analyses several highly cited studies by other scholars who employ Paribakht and Wesche's self-assessment methodology in their research.

The second part of the thesis presents the experimental studies. Chapter 3 explains our choice of the case study methodology administered in this research. It also reports on the first study carried out to investigate how L2 learners assess and categorise their own knowledge of words.

Chapter 4 unveils the second experimental study which investigates how L2 learners self-measure their knowledge of multiple meaning senses of polysemous words.

Chapter 5 reports on the third experimental study which investigates in detail how an individual L2 learner measures and classifies his knowledge of 200 target words.

Chapters 6 and 7 explore in detail via the fourth and fifth experimental studies respectively, how an individual learner evaluates and categorises their knowledge of 200 words extracted

from an original Russian story. In these chapters, we also investigate how a learner's knowledge of words develops over time after vocabulary input through reading.

Chapter 8 reports on the repeated longitudinal study carried out to investigate (in further detail and over a longer period of time) the issues raised in the previous studies. We use self-categorization to explore how a learner's word knowledge develops over the four subsequent sessions of reading.

In the final part of the thesis, Chapter 9 discusses the findings of the experimental studies and their implications, and speculates on the issues raised in this thesis. It focuses on the ways L2 learners self-assess and classify their knowledge of words, emphasising complexity and instability of learners' own self-assessment categorizations. It speculates on what should be considered within lexical self-assessment methodologies in light of the findings of this research. It discusses the process of development of L2 lexical knowledge.

Chapter 10 summarizes the findings of this thesis.

The Appendix presents a pilot study into the self-assessment of L2 lexical knowledge. The findings of that pilot study determined the framework of our research.

1.4 Background

1.4.1 Self-assessment in L2 learning

Self-assessment in language learning is known for contradictory empirical results obtained from using this approach. Indeed, the question of validity and reliability of self-assessment in second/foreign language testing has been discussed for decades.

Generally speaking, in second language testing, all research that administered self-assessment as a measuring instrument can be divided into two groups. The first group is represented by those who report a good match between self-assessment and actual assessment results (Van Passel, 1974; Oskarsson, 1978; Von Elek, 1982; LeBlanc and Pinchaud, 1985). The other group of researchers claims that self-assessment provides inaccurate information in regards to the learners' language knowledge and skills with a tendency for over-estimating (Oscarson,

1989; Boud, 1995; Janssen-van Dieten, 1989; Peirce, Swain and Hart, 1993). It has been claimed that the accuracy of learners' self-assessment depends on their level of proficiency (Oskarsson, 1984; Ross, 1998; Heilenman, 1990), the language skill/area being measured (Evers, 1981; Raasch, 1979; Anderson, 1982; Strong-Klause, 2000), the wording of the questionnaire (LeBlanc and Painchaud, 1985), the type of self-evaluation exercise (Pierce, Swain and Hart, 1993; Strong-Klause, 2000) and learners' cultural background (Anderson, 1982; Blanche, 1988; Strong-Klause, 2000).

In second language proficiency, self-assessment has mostly been used as an evaluating instrument for the four language skills (Raasch, 1979; Heindler, 1980; Oskarsson, 1980; Fok et.al, 1981; von Elek, 1981; Rea, 1981; Anderson, 1982; LeBlanc and Painchaud, 1985; Blanche and Merino, 1989; McNamara and Deane, 1995) or one or some of them: speaking, reading and listening (Evers, 1981), speaking and reading (Palmer and Bachman, 1981), listening and speaking (Ferguson, 1978), speaking (Holec, 1979; Ferris, 1982, Blanche, 1986), reading (Brantmeier, 2006; Wan-a-rom, 2010), listening (Buck, 1992). In most of these studies, self-assessment involved answering a questionnaire on the language knowledge and/or the ability to use this knowledge. Emphasizing the importance of accurate self-assessment for learners' autonomy, Blanche and Merino (1989) provide a detailed account of research on self-evaluation of foreign language skills. It is clearly apparent from their summary of research that questionnaires were the main type of self-testing instruments used in L2 assessment: Raasch, 1979; Ferguson, 1978; Palmer and Bachman, 1981; Anderson, 1982; Fok, 1981 and others. Barrow et al. (1999) administered a self-checking survey as a questionnaire modification. The second common form of self-assessment seems to be accounted to self-rating scales/forms: Holec, 1979; Ferris, 1982; von Elek, 1981; Rea, 1981; Achara, 1980 and others. The scoring scales for questionnaires and self-rating scales are normally based on a 5 (rarely 7 or 10) point scale. In most instances, when the language skills are self-assessed, these points are awarded on the following terms: (5points) I can do it all the time; (4points) I can do this most of the time; (3) I can do this half the time; (2) I can seldom do this; (1) I can never do this (LeBlanc, 1985; Oskarsson, 1978; Lewkowicz and Moon, 1985). These terms seem quite vague. For instance, it does not appear easy to differentiate between the cases when 3 or 2 points are given. Another type of scoring scale (though still 5-point) was suggested by LeBlanc (1985) who asked learners to indicate up to what point they feel they can complete a certain activity in speaking, writing, listening and reading. This also

seems confusing since language activities might be multi structured.

Oscarson (1989) provides his summary of self- assessment techniques and materials. He notes the following:

a) progress cards and other record keeping devices. As an example, he describes personal test cards on which a learner ticks off each language activity once they are confident they can complete it;
b) questionnaires, rating scales and check-lists aiming at overall assessment of perceived ability levels. As an example, he illustrates Oskarsson's rating scale (1980) on which the learner ticks the level he thinks is appropriate to his own abilities. Another example is from Raasch (1970): a check-list on which the user indicates their ability to cope with the situation by ticking off one of the four suggested levels of abilities;
c) diaries and log books suggested by Carver and Dickinson (1981) and Scharer (1983);
d) video and audio cassettes for purposes of oral skills self-assessment undertaken by Ferris (1983) and Oskarsson (1984);
e) computer-assisted assessment which is currently becoming a more and more popular type of self-evaluation in language training and testing.

The author outlines a justification for adopting self-assessment principles in language training. A number of researchers suggest using self-assessment for placement purposes. LeBlanc and Painchaud (1985) report a series of experiments which led to the use of self-assessment as a placement test. They argue that self-assessment questionnaires covering the four basic skills can replace a proficiency test in a second language. Each of the four parts of their questionnaire (Listening, Reading, Speaking and Writing) included 10 statements graded for difficulty level. Learners were required to read each of the 40 statements and indicate their abilities on the scale ranging from 1 ("I cannot do this at all") to 5 ("I can do this all the time"). In subsequent experiments, LeBlanc and Painchaud increased the number of statements in their self-assessment questionnaire testing listening and reading skills. They also changed the original statements in such a way that they became closely related to the learners' situations as second language users. That was reported to considerably improve correlations between self-assessment and standardized proficiency tests for listening and reading. Thus, according to LeBlanc and Painchaud, the factor that influences the accuracy of language learners' self-assessment is well-formulated (with good descriptors) linguistic situations.

Generally, they report that self-assessment proved to be a very valuable tool as a placement instrument. Furthermore, the authors point out that given knowledge of the purpose of placement exams and their role in determining the level of instruction, learners do not see any reason for falsifying self-evaluation of their language abilities.

Strong-Klause (2000) also investigates the use of self-assessment for placement purposes. Her study aimed to determine which types of tasks within a self-assessment instrument should be used to best predict placement in a language course. Similar to Pierce, Swain and Hart (1993) the author found that "the most specific tasks proved to have significant predictive power with the placement exams" (p.59) She also reports that out of the four language skills being self-assessed, the speaking self-assessment appeared to be the best predictor, while the reading self-assessment turned out to be the worst. These findings are contrary to the results reported by Evers (1981) who names listening as the least accurately self-assessed skill, Anderson (1982) who claims that speaking and writing proved to be the least self-evaluated skills and Raasch (1979) who reports that receptive skills were assessed more accurately than were productive skills.

Developing the ideas raised in LeBlanc and Painchaud (1985)'s study, North (2000) elaborates "concrete task-based descriptors". The author presents a list of points which, he claims, make a descriptor work. They are: positiveness (positive formulation), definiteness (descriptors should describe concrete tasks or features), clarity (descriptors should be transparent), brevity and independence (descriptors should describe a criterion behavior). Furthermore, North involves a learner in developing a language portfolio which includes a self-assessment checklist with the self-perceived information on their language abilities.

Bachman and Palmer (1989) investigated an experimental self-rating test of communicative language abilities: grammatical competence, pragmatic competence and socio-linguistic competence. 116 non-native English speakers completed a 21-item multiple-choice self-rating test. They were asked to rate on a four-point scale (ranging from BAD to GOOD) their answers to the three question types: their ability to use each of the three traits, their difficulty in using the trait and the degree to which they were able to recognize the trait in input. The most effective, in terms of self-assessment responses, is reported to be question type 2 which is subjects' perceived difficulty with various aspects of the language. As an overall

conclusion, the authors claim that self-ratings can be reliable and valid measures of communicative language abilities.

Janssen-van Dieten (1989) compares a test of Dutch as a second language and a parallel version of that test in self-assessment format. The examples of self-assessment items are as follows: Do you think you can complete this personal questionnaire (name, place of residence, nationality, etc.)? Look at the parking sign. Do you think you can complete this sentence correctly? This self-assessment test resembles Von Elek's self-assessment test of Swedish as a second language (1985). Similar to Von Elek's test, Janssen-van Dieten's self-assessment test offers three optional answers: YES (I can), I am not quite sure and NO. A scoring method applied by Von Elek scored the option "I am not sure" as either "yes" or "no" depending on the criterion score on that item. A scoring method used by Janssen-van Dieten awarded the "yes" choice with two points and "I am not sure" –with one point. However, Janssen-van Dieten states that he applied Von Elek's scoring method afterwards in order to compare the correlation coefficients between the two studies. Unlike Von Elek's conclusions, Janssen-van Dieten reports that "both sets of correlation coefficients were too low to call the self-assessments valid predictors of criterion scores" (p.38)

Peirce, Swain and Hart (1993) investigate validity and reliability of self-assessment in French immersion programs. A detailed self-assessment questionnaire and French proficiency tests were administered to the immersion students. The subjects were asked to assess their own French language skills against the two benchmarks: "francophone peers" and difficulties of specific everyday tasks conducted in French. Within Francophone peer benchmark, subjects were asked to indicate how they would compare their 4 main language skills in French to those of a native speaker. The options on the self-assessment scale were: about the same (3points), somewhat worse (2points) and much worse (1point). For difficulty with specific tasks benchmark, participants were required to indicate their abilities on the scale ranging from 1 ("Not at all") to 5 ("Without any difficulty"). The results are reported to have revealed weak correlation between self-assessments and objective measures of foreign language proficiency.

Ferguson (1978) examines three tests of self-assessment of oral skills. Check-lists contained the following types of questions: Can you ask and tell the time of day, date…? Can you order

a simple meal? Are there grammatical features of the language which you try to avoid? etc. Learners were required to tick a box if their answer to the question was positive. The results of the self-assessment tests were reported to be well correlated with those of the listening comprehension test (r=0.87), but less well with the speaking assessment (r=0.39).

Ross (1998) reviews research on self-assessment in language testing. The combined meta-analysis conducted by Ross summarizes the relationship between 60 correlations reported in the research literature. Ross analyses studies in which four second language skill areas were correlated with the self-assessment scales (Bachman and Palmer, 1981; Buck, 1992; Blanche and Merino, 1989 and others). Ross also reports the results of an empirical analysis of validity of a self-assessment instrument. 236 learners of English as a foreign language completed self-assessment of their English language skills. The results were correlated with those provided on the achievement test. Contrastive multiple regression analyses revealed differential validities for self-assessment compared to teacher assessment. Ross concludes that the accuracy of learners' self-assessment is influenced by the degree of learners' experience with the language skill self-assessed.

Brantmeier (2006) reviews studies on self-assessment of L2 learning abilities. She concludes that generally these studies (Oscarson, 1978; Krausert (1991); Hargan (1994), Birckbichler et al. (1993); Deville and Deville (1999); Brantmeier (2005) and others) support the use of self-assessment "as an indicator of second language abilities" (p.19). Brantmeier also discusses the results of her research project in which she tests the reliability of self-assessment of L2 reading skills by using a questionnaire. 71 advanced L2 learners of Spanish read a short story and completed three different comprehension assessment tasks. Self-assessment of L2 reading ability was measured before and after the reading via a 5-point scale. The Pre-reading self-assessment questionnaire was taken from Birckbichler et al. (1993) and Deville and Deville (1999), with the typical options: a) not very well at all; b) not very well; c) OK; d) well; e) very well. The post-reading self-assessment questionnaire was taken from Schraw et al. (1995): "I found the passage I just read easy to understand", with the 5 options to choose from: a) I strongly disagree; b) I disagree; c) I somewhat agree; d) I agree; e) I strongly agree, and from Tobias and Everson (1998): "How much of the passage did you understand?" with the following 5 options: a) I did not understand very much at all; b) I did not understand very much; c) I understood some of it; d) I understood most of it; e) I understood all of it.

The author reports that advanced learners failed to accurately estimate their L2 reading abilities which contradict earlier findings discussed in the paper. She states in conclusion that the results of her study highlight the need for "a more contextualized and criterion-referenced self-assessment instrument".

Rivers (2001) and Tseng et al. (2006) explored self-assessment from a psychological perspective. Rivers analyzed self-directed language learning behaviours of L3 learners. The study was based on the survey data collected from the students. All learners were reported to accurately assess their progress, learning styles, strategy preferences, and conflicts with teaching styles. Learners were found to demonstrate "a high tendency towards learner autonomy, requesting …changes to every aspect of the course" (p.287)

Dickinson (1987) argues that self-assessment in Language learning is both "possible and desirable". She emphasizes the importance of self-assessment within self-monitoring in language training since it enables learners to make judgments about the accuracy of their performance.

Summarizing research on self-assessment in second language proficiency, it should be noted that despite controversial results received from its administering, most researchers agree that self-assessment is an important means for the development of learners' autonomy. Described by many scholars as a powerful learner-directed assessment tool, self-assessment is claimed to "highten learner awareness of personal strength and weaknesses, and to promote language acquisition" (Ekbatani, 2000). In regards to implementing self-assessment into the language acquisition process as an alternative or supplement to the objective testing, most scholars emphasize the need for improvement of self-assessment forms/tools in terms of the questions/statements proposed in questionnaires/self-rating scales/checklists as well as changes in scoring scales accompanying them.

Conclusion. As seen from the survey of research in self-assessment, most of the studies focused on the issues of learners' self-evaluation of their language skills: reading, writing, speaking and listening. It is readily apparent that these main language skills are based on the knowledge of words. Therefore, it would be interesting to see how self-assessment of vocabulary knowledge is carried out.

1.4.2 Self-assessment in L2 vocabulary testing.

1.4.2.1 Review of different approaches and types of lexical self-assessment

Generally speaking, there has been a relative shortage of research on the issue of L2 vocabulary self-assessment. In particular, a relatively small amount of research has been conducted to evaluate the effectiveness of self-assessment instruments used in testing learners' vocabulary knowledge.

As noted by Oskarsson (1980), the most common (traditional) form of self-assessment of learners' word knowledge consists of ordinary "paper-and-pencil" or their computer version vocabulary (often multiple-choice) tests scored by learners themselves with the help of an answer key and self-scoring answer sheets.

Most research implementing self-assessment as a measurement instrument for testing L2 learners' knowledge of words administered a checklist approach. Within this approach, learners are required to check or indicate whether they know each of the words on the word list. Read (1988) emphasizes the simplicity of the checklist approach. He writes: "...it strips away irrelevant task demands that may make it difficult for...readers...to show what they know." (p.23). Another advantage of the checklist approach is considered to be its capability to cover a wide range of words. Non-words are often included into the word list to control the problem of over-estimating. Nation (1990) notes the checklist with non-words proved to be efficient and useful in testing the vocabulary size.

Barrow et al. (1999) administer a self-checking type familiarity survey of words in an attempt to establish which English words are unfamiliar to the Japanese learners. Their study also aimed to investigate the validity and practicality of the familiarity survey as a measurement of vocabulary knowledge. They asked 1283 Japanese students of English to indicate whether they knew the word on the vocabulary list. The results showed that learners had a tendency to over-estimate their own vocabulary knowledge when using a self-checking list of words.

Laufer and Yano (2001) investigate how accurately learners can assess their understanding of words and which factors affect this accuracy. 106 learners of English as a foreign language from China, Japan and Israel took part in the research. The study consisted of three stages.

At stage 1, subjects were required to read a text and complete the self-assessment form. They were asked to rate their understanding of each of the 20 target words on a scale from 0 to 2 (0- don't understand; 1- understand approximately; 2-fully understand). At stage 2, the target word understanding was checked via translation/explanation test. Stage 3 repeated stage 1. Subjects' scores from stage 2 were compared against the self-assessment scores obtained at stages 1 and 3. The authors report over-estimation in learners' comprehension of the target words. They also found that this over-evaluation was related to the level of subjects' lexical knowledge as well as their cultural backgrounds.

Wan-a-rom (2010) reports the results of the case study investigating how five Thai learners of English self-assessed their own word knowledge in a graded reading text. In the trial session, the checklist was administered to 96 students who were supposed to assess their knowledge of the 94 content words. It had three options for learners to choose from: Yes, Not sure and No. The checklist was reported to provide the materials used in the main study. In the main study, students read a text and carried out text-based self-assessment of certain words in the text. They were asked to identify the unknown words in the text they had read. Observation, interviews and a translation test were used to investigate how self-assessment was performed by each of the five subjects. Wan-a-rom reports that all five learners acknowledged problems with the words they were not confident about. He notes that in most instances, those words were not identified as unknown which led to overestimating. Indeed, the overall results were reported to indicate that all the five learners overestimated their word knowledge when they carried out text-based self-assessment. Despite this, Wan-a-rom concludes that "the study provided evidence to support the value of self-assessment as an easy procedure to direct learners to an appropriate reading level" (p. 323)

Discussing a checklist format, Paribakht and Wesche (1996) point at its positive (easily administered, computerized, automated and self-scoring) and negative sides (unreliability, overestimation and lack of ability to test various aspects of word knowledge). Paribakht and Wesche (1996) also criticize descriptive scales used in second language acquisition "to guide rater judgments of vocabulary features". They point at their ambiguous terminology and underlying concepts.

Meara and Buxton (1987) suggested a yes/no vocabulary test for L2 learners. In this study,

a multiple-choice test of vocabulary and a yes/no test were administered to 100 subjects. The yes/no test was based on 100 items with 40 non-words. The authors report that the yes/no test proved to be reliable in predicting those who passed the Cambridge First Certificate Examination. In conclusion, they state that the main advantage of yes/no tests is its ability to test a significant proportion of the words a learner is expected to know, which "reduces the risk of arbitrary sampling"(p.150).

The idea of Yes/No tests was further developed by Meara and Jones (1990) and Meara and Milton (2003). Meara and Jones (1990) devised the Eurocentre's Vocabulary Size Test which enables a learner's knowledge of the most frequent 10000 words to be estimated. Meara and Milton (2003) designed X_Lex tests which provide an estimate of the words a learner knows out of the most frequent 5000 lemmatised words in each language. Each test includes 20 false words which help adjust a learner's score for guessing and over-estimating.

In the studies reviewed in this section, self-assessment was used to measure the quantitative characteristics of learners' L2 lexical knowledge or, in other words, to measure the size or breadth or range of a learner's vocabulary. The qualitative features or depth of a learner's word knowledge were not measured. According to Barrow et al. (1999), "measuring depth of (word) knowledge would require the use of more sophisticated recognition and elicitation (self-assessment) devices..." (p. 225).

Read (1989) emphasizes the importance of measuring depth as well as breadth of word knowledge. He carried out research into self-assessment of word knowledge via checklists and interviews with learners, followed by specific questions about each of the target words. Sixty target words of low frequency were offered to subjects in written checklist form. They were asked to indicate whether they knew the word, did not know or were not sure. In the interviews, subjects were required to "a) pronounce each word, b) explain what it meant, c) identify the fields of study in which it belonged, d) name other words with which each was associated and e) indicate their knowledge of other forms of the word". Paribakht and Wesche (1996) point out that though Read's research provides valuable information for the development of vocabulary depth measures, no testing instrument was suggested.

Generally speaking, apart from the Vocabulary Knowledge Scale (VKS) suggested by

Paribakht and Wesche in 1993 (and developed over the period 1993- 1997) and its slight modifications made by a number of other researchers over time (Wolter, 2001; Rott, 2005; Folse, 2006; Zareva, 2007) there are no self-assessment instruments in second/foreign language acquisition which were claimed to measure depth of learners' vocabulary knowledge. Paribakht and Wesche argue that their self-assessment tool can be used for measuring both breadth and depth of knowledge of words being acquired. If this is the case, Paribakht and Wesche's Vocabulary Knowledge Scale might be regarded as a pioneering or revolutionary approach in second language vocabulary self-assessment. In fact, Paribakht and Wesche (1996; p.29) argue that their VKS "goes beyond the instruments currently available for naturalistic studies of vocabulary acquisition and instruction".

In order to assess this claim, we need to understand the concept of depth in L2 vocabulary acquisition and assessment. We also need to be certain about the purpose and aims of self-assessment in L2 vocabulary testing. These rather important issues are considered in the sections that follow.

1.4.2.2 The purpose of self-assessment in L2 vocabulary testing.

Here we are discussing the purpose and aims of self-assessment in vocabulary testing. What is a lexical self-assessment test intended to achieve? We argue that the idea of lexical self-assessment is to assess the whole of knowledge about a word, i.e. to measure both breadth and depth.

As seen in the previous section, there are different ways of commencing self-assessment in L2 vocabulary testing: self-scoring answer sheets, self-checking lists, word familiarity surveys, simple self-rating scales with typically three options (Known, Not Sure, Unknown). These types of L2 lexical self-assessment aim at establishing whether a particular word is known or unknown by the learner. They do not provide the learner with information on the qualities of that knowledge (i.e. do not measure vocabulary knowledge in depth). Though it is essential to assess size or breadth of a L2 learner's vocabulary, it is equally important to establish how well the word is known. In our opinion, the L2 lexical self-assessment descriptive scales may be one of the possible forms of commencing lexical self-assessment comprehensively (i. e. measuring both breadth and depth). These self-evaluation instruments should be able to

establish how wide a learner's vocabulary is, and at the same time inform the learner which areas of knowledge of a certain word they already possess and which areas require attention. In other words, these self-assessment methodologies should be able to embrace the whole of knowledge about each word tested.

Overall, we argue that L2 learners should carry out regular self-measurements (in size and depth) of their knowledge of the words which they acquire on a daily basis. These checks will highlight the areas of knowledge which need to be corrected or enhanced which, in turn, may accelerate the process of word acquisition in general. Emphasising the importance of self-measuring the depth of knowledge of acquired words it is important to be certain what the construct of depth implies.

In reality, the issue of "word knowledge depth" is not easy. One might ask: what does it imply? How well can the word be known? What are the constituents of "word knowledge"? Let us take a look at the concept of depth in L2 lexical acquisition in detail.

1.4.3 The concept of depth

Traditionally, in vocabulary acquisition, researchers distinguish between breadth and depth of word knowledge. In simple terms, breadth of vocabulary knowledge is determined by how many words are known, whilst depth of word knowledge is measured by how well these words are known (i.e. what is known about these words). The importance of distinguishing between these two concepts is shown in Meara and Wolter (2004; 95) "we might find learners with similar vocabulary sizes, but very different degrees of organisation in their lexicons...". While the concept of breadth of vocabulary knowledge does not seem to cause major disputes in the field, the concept and framework of vocabulary depth appear to be rather unclear.

Generally speaking, there is no consensus in contemporary vocabulary acquisition and assessment in regards to the definition of depth of word knowledge (as opposed to breadth). Though most scholars agree that lexical knowledge is a rather complex construct, the constituents and levels of vocabulary knowledge differ in various approaches. Moreover, as pointed by Nation (2001), there has been "enormous debate" about certain aspects of word knowledge that constitute the phenomenon "depth of lexical knowledge".

In fact, in L2 vocabulary acquisition, various aspects/kinds/types of word knowledge have been determined, such as: knowledge of meaning senses, ability to use, knowledge of syntactic properties, knowledge of different word forms and derivations, associative abilities and so forth (Cronbach, 1942; Evans et.al., 1986; Gass, 1989; Richards, 1976). Nation (1990) determines four dimensions of word knowledge: word's form, word's position, word's meaning and word's function. He also presents a list of eight word knowledge types:

1. The spoken form of a word; 2. The written form of the word; 3. The grammatical behavior of the word; 4. The collocational behavior of the word; 5. The frequency of the word; 6. The stylistic register constraints of the word; 7. The conceptual meaning of the word; 8. The associations the word has with other related words (pp.30-33).

Schmitt and Meara (1997) develop a word knowledge framework based on that list. They argue that vocabulary tests should be able to assess different aspects of learners' word knowledge rather than entirely focus on measuring their meaning knowledge.

Read (2000) investigated how well learners knew high-frequency English words (i.e. depth of learners' word knowledge) by using a written version of the interview procedure described in his earlier work (1989). He tested various aspects of learners' word knowledge including meanings, use, collocations and derived forms emphasizing the importance of establishing what knowledge of the target word the learners have.

Schmitt (1998) investigates this matter further. Exploring how individual words are acquired diachronically, he argues that it is vital to employ measurement procedures that can identify various degrees of lexical knowledge, and to involve tests of acquisition that are able to capture a wide range of word knowledge constituents, such as: spelling, associations, grammar forms, meaning senses, etc. Schmitt analyses two different ways of testing the depth of vocabulary knowledge: the developmental approach and dimension approach. Schmitt argues that there are advantages and disadvantages to both these approaches. Among the advantages of the developmental approach, he names the use of scales to describe the particular stages of acquisition of a word. However, at the same time, Schmitt emphasises that defining the stage boundaries within this approach might be vague and problematic. Furthermore, uneven intervals between the stages, ambiguity in determining the number of

stages required to describe the acquisition of a word, as well as lack of balanced attention to both receptive and productive knowledge within a scale, are mentioned as serious disadvantages of the developmental approach.

The dimension approach, on the contrary, receives mostly positive comments from Schmitt. He considers that describing word knowledge in terms of the eight categories suggested by Nation (1990; 30) as "the most complete and balanced" approach. It is interesting what characteristics of the dimension approach Schmitt names as advantages. First, it is comprehensiveness which implies that a researcher measures different aspects of knowledge of a word when assessing how well this word is known. However, at the same time, Schmitt notes that it could be time-consuming and quite complicated to test the knowledge of each of individual words according to all the parameters. Second, Schmitt notes that the dimension approach allows a researcher to test developmental sequence of various kinds of word knowledge.

The most elaborated taxonomy of what is involved in knowing a word was suggested by Nation (2001). In his latest classification (2001), Nation develops his earlier model of word knowledge (1990) subdividing its three main aspects: form, meaning and use into finer components. This is illustrated in Table 1-1.

Table 1-1: What is involved in knowing a word (Nation, 2001; 27)

FORM	Spoken	R	What does the word sound like
		P	How is the word pronounced?
	Written	R	What does the word look like?
		P	How is the word written and spelled?
	Word Parts	R	What parts are recognisable in this word?
		P	What word parts are needed to express the meaning?
MEANING	Form & Meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	Concepts & Referents	R	What is included in the concept?
		P	What items can the concept refer to?
	Associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
USE	Grammatical Functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	Collocations	R	What words or types of words occur with this or
		P	What words or types of words must we use with this one?
	Constraints on use	R	Where, when and how often would we expect to meet this word?
		P	Where, when and how often can we use this wor

Some further consideration of Nation's categories is in order here.

Knowing the form of a word in Nation's classification includes knowing its phonological form, its orthographic form and its morphological form. These three constituents of a word form are further subdivided into receptive (marked as R) and productive (marked as P) areas of knowledge each. Receptive knowledge about the spoken form of a word, for instance,

includes the ability to recognize the word when it is heard, while the productive knowledge of it includes the ability to pronounce the word correctly when required.

Similarly, the meaning aspect of word knowledge in Nation's categorization is divided into the following three areas: form and meaning, concept and referents, and associations. Each of these areas is subdivided into two parts: receptive and productive. The form and meaning constituent indicates a close connection between these two types of lexical knowledge. It includes the ability to infer the meaning from the form (receptive knowledge) as well as the ability to select the right form to express a certain meaning (productive knowledge). The concept and referents element indicates the importance of a cultural dimension to the meaning which should be learned when the word is acquired. The associations component of the meaning aspect provides information for understanding relations among words in the lexicon which in turn might provide some insight into how the lexicon may be organized.

Finally, knowledge of how to use a word is also divided into three areas: grammatical functions, collocations and constraints on use. And again receptive and productive types of knowledge are determined in each of these areas. Grammatical functions include knowledge of what part of speech the word belongs to and "what grammatical patterns it can fit into" (2001; 55). Knowledge of collocations relates to understanding with what kind of words the target word is normally used and includes knowledge of patterns of collocations. And, finally, knowledge of constraints on word use concerns awareness of sociolinguistic factors that limit where and when particular words can be used.

As seen from the description, Nation's classification of types of word knowledge is rather complex and rich. Indeed, a large number of qualities, 18 in total, are included in Nation's framework of the construct "word knowledge". This suggests that depth of word knowledge, as opposed to breadth, encapsulates all these categories determined by Nation and possibly more.

A different approach is offered by Daller et. al. (2007). These researchers describe word knowledge as a union of three main "dimensions": breadth, depth and fluency. In this model, a fair amount of attention is paid to fluency as opposed to breadth and depth of word knowledge. This is a strange approach as it seems to imply that depth is "an aspect of passive word knowledge". This would mean that the ability to use a word, which is one of the main

aspects in Nation's classification of word knowledge, is not considered to be a part of the framework of depth but fluency instead.

Reviewing research on second language vocabulary assessment, Milton (2009) points at the complexity and ambiguity of the concept of depth of word knowledge. He states that a variety of aspects of word knowledge can be included into the structure of this construct. However, at present, he continues, there is a lack of "clear, comprehensive and unambiguous definitions" of the concept of depth of word knowledge (2009; 150). Milton also points out that due to enormous complexity of lexical knowledge, a single test would not be able to measure all aspects of this construct, and multiple measures would be required instead.

There is also a viewpoint in vocabulary acquisition that depth of word knowledge or, at least the major part of it, should be considered as a network of links among words (Henriksen; 1999). This is an attempt to find an answer to the question of how lexical units that construct the depth of word knowledge associate and interrelate with one another. These links have been investigated in numerous word association studies which explored either existing or simulated word association behaviour in L2 vocabulary acquisition (for example, Wilks, Meara and Wolter (2005); Wilks and Meara (2007)). Emphasising the existence of the close connection between breadth and depth of word knowledge, Vermeer (2001) states that "a deeper knowledge of words is the consequence of knowing more words, or ... the more words someone knows, the finer the networks and the deeper the word knowledge" (2001; 222). This is an interesting assumption. Analysing different aspects (qualities) of word knowledge illustrated in Nation's taxonomy (Table 1-1), it can be seen how such types of knowledge as "concepts and referents", "associations", "grammatical functions", "collocations" and "constraints on use" might be based on the knowledge of other words. However, word form and its components ("spoken form", "written form" and "word parts"), as well as some qualities of the aspect of meaning ("form and meaning" for instance), do not seem to depend on how well other words are known. Instead, these types of word knowledge are acquired by learning a word itself.

Having reviewed different approaches to depth of word knowledge we can clearly see that this construct is rather complex and multi-dimensional. This suggests that the adequate assessment of how well a certain word is known (i.e. depth of knowledge of this word) would imply evaluation or measurement of a large variety of aspects (types, qualities) that constitute

word knowledge. Having established that, we are now curious to investigate how depth of word knowledge is self-assessed and measured in L2 vocabulary acquisition and testing.

This is the target of our next chapter.

CHAPTER TWO

Reviews of a number of studies that used the Vocabulary Knowledge Scale developed by Paribakht and Wesche (1993)

2.1. Reviews of VKS studies by Paribakht and Wesche

In the previous chapter, we discussed the concept of depth of word knowledge. We introduced different viewpoints on this rather complex construct. We highlighted the diversity of interpretations of what it means to know a word. We considered Nation's latest (2001) taxonomy of qualities of vocabulary knowledge noting that this classification appears to be the most elaborate and detailed model (suggested so far) of what might be included in the concept of depth.

Noting the enormous complexity of the concept of depth we raise the question of how depth of L2 lexical knowledge can be self-assessed and measured. Or is it generally possible to self-/measure the depth of knowledge of a particular word by means of a single scale? Indeed in order to obtain reliable information on how well (to what extent) a certain word is known, the whole range of qualities (types or aspects) that Nation includes in his concept of depth should be accurately self-/assessed.

When we look at other research on self-assessment in L2 vocabulary acquisition we immediately note that very little work has been carried out on this issue. Moreover, it appears that the Vocabulary Knowledge Scale (VKS) devised by Paribakht and Wesche in 1993 is the only elaborated self-assessment instrument which claims to measure the quality or depth of L2 word knowledge. Paribakht and Wesche (1996) seem to have been rather critical towards the existing self-assessment methodologies in L2 vocabulary testing, pointing out that they do not assess the quality of knowledge of words. They argue that their VKS "goes beyond the instruments currently available for naturalistic studies of vocabulary acquisition and instruction" (1996; 29). Indeed, the VKS is well-known and the most cited self-assessment scale. But the question is whether it is able to measure depth of L2 word knowledge? Is it a new revolutionary breakthrough in second language vocabulary self-assessment?

Within the VKS methodology, Paribakht and Wesche ask learners to evaluate how well they know each of the proposed words, i.e. to assess their own level in vocabulary depth. Bearing in mind the enormous complexity of the construct of depth (discussed in the previous chapter)

we have been a bit skeptical and cautious in regards to the ability of the VKS to assist L2 learners in self-assessing the depth of their own knowledge of words.

For this reason, we intended to review in detail a number of different studies by Paribaht and Wesche and also by other researchers which administered the Vocabulary Knowledge Scale as an assessment tool. By doing this, we intended to investigate how accurate Paribakht and Wesche's claims are in respect of the ability of their scale to measure depth of vocabulary knowledge.

Our interest in Paribakht and Wesche's VKS methodology has also been stimulated by contradictive feedback the VKS has received in L2 vocabulary acquisition and testing. Paribakht and Wesche argue that the VKS is a workable measure of both breadth and depth of word knowledge. However, a number of researchers emphasise serious drawbacks of Paribakht and Wesche's methodology. Read (2000), Wolter (2005; 29-33), Milton (2009; 161) point out that the full range of the scale is not used, the test is insensitive to many aspects of depth of word knowledge, the maximum score of 5 does not really indicate that the word is known to full extent. Moreover, Milton (2009; 161) concludes that "the VKS may not function as a scale at all, but may, ultimately, be a binary test (*I know this word/ I do not know this word*)." Furthermore, he states that the VKS was used in his own studies as a measure of the growth of vocabulary breadth rather than depth. These comments seem to undermine the significance of the VKS as a pioneering methodology for self-/measuring the quality or depth of word knowledge. However, we decided to keep an open mind and investigate Paribakht and Wesche's methodology by analysing a number of studies which employed the VKS as a measurement instrument of L2 lexical knowledge. We were interested to see how the VKS methodology was used in that research and whether the Vocabulary Knowledge Scale provided adequate and accurate information on a level of knowledge of targeted words. In other words, we wished to learn whether the VKS worked as a self-evaluation instrument especially in respect of the quality or depth of knowledge of the words being tested.

The points outlined in this section have defined our interest in Paribakht and Wesche's VKS methodology. Thus, we have decided to focus on this methodology and take the VKS as a benchmark in our research.

The following sections review articles by Paribakht and Wesche reporting on the studies which involved the authors' Vocabulary Knowledge Scale as an assessment instrument.

2.1.1 Paribakht and Wesche. 1993.

2.1.1.1 Summary

The article reports on a study of the acquisition of specific content vocabulary, discourse connectives and grammar knowledge by L2 learners. The study investigates the role of comprehension in L2 development. The objective was to explore the language learning outcomes (mostly gains in vocabulary knowledge) of a teaching approach based on global comprehension of written and oral texts.

The methodological issues discussed in this study were as follows:
--What kinds of measurement instruments and procedures can be used to track learner gains in specific aspects of target language proficiency?
-- Can a reliable self-report scale be developed to capture different levels of knowledge of vocabulary items?
-- Can introspection be used as a technique for exploring how learners deal with unknown vocabulary items while performing comprehension-based tasks?
Among the research questions highlighted were the following:
-- Do learners make measurable gains in vocabulary and grammar knowledge in comprehension-based classes?
-- Which type of vocabulary items is most easily acquired: content words or discourse connectives?
-- Are content vocabulary items related to instructed themes, acquired more readily than those from uninstructed themes?
-- Can different stages of vocabulary acquisition be identified?
Subjects were 37 university students from a variety of first language backgrounds. Subjects were divided into two groups: 19 in the Comprehension-based class and 18 in the Four-skill comparison class. Classroom treatment took up 54 hours. The Comprehension-based class concentrated on listening and reading activities on selected themes. There was no grammatical

instruction. Students practised reading skills and had to answer comprehension questions on the content of suggested authentic texts (1.5 hours). They also worked in the language laboratory listening to a recorded text or watching a video on the same theme as their current reading text followed by a new series of comprehension questions to answer (another 1.5 hours). One hour a week was allocated for newspaper or magazine reading.

The Four-skill group received a special classroom treatment based on an integrated four-skill approach. This included grammar instruction followed by exercises on prepositions and vocabulary (1.5 hours a week); listening and reading materials (another 1.5 hours per week), writing (0.5 hour a week) and speaking (at least 0.5 hour per week) activities. Hence, in this group the emphasis was placed on explicit grammar instruction as well as oral and written production.

Two sets of instruments were used to measure vocabulary acquisition: Cloze Tests and Vocabulary Knowledge Scale. They were administered to both groups at the beginning and end of the course. Two Cloze Tests with 35 blanks each tested content words and cohesive markers selected by the authors from the theme-related texts. Students filled in the blanks from a list which included all the missing vocabulary items as well as five distracters.
A self-report Vocabulary Knowledge Scale (VKS) developed by Paribakht and Wesche was administered as a measurement instrument in this study for the first time. It included five levels of description of the target word knowledge. The levels ranged from total absence of knowledge of the target word and some idea of its meaning, to the ability to use the target word in a sentence.

The self-report categories were as follows:
1. I have never seen this word.
2. I have seen this word before, but I do not know what it means.
3. I have seen this word before, and I think it means _____ (synonym or translation).
4. I know this word. It means _____ (synonym or translation).
5. I can use this word in a sentence:_____.
Learners were required to indicate on the VKS how well they knew each of the target words. This was completed before the cloze tests were administered. The authors claimed that VKS was utilized to capture different levels of self-perceived knowledge of particular words.

The VKS scoring was carried out as follows. A 6-level scale was administered to score the results. If subjects reported their knowledge of words in categories 3-5 on the VKS, that was considered as evidence of demonstrated rather than perceived knowledge. In Category 3, if a subject provided an incorrect synonym or translation, the response was scored as category 2, in case of a correct answer it was scored as category 4. Category 5 was also split into two levels in order to reflect vocabulary and grammar knowledge. If the word was misused according to both criteria the response was scored as category 2. In case of the correct meaning provided only, a category 5 score was awarded. Category 6 reflected both semantically and grammatically correct use of the word in a sentence.

Grammatical knowledge was also measured through two different tests: a general test of grammatical knowledge developed by Bialystok and her colleagues in 1981 and a test on knowledge of common prepositions.

The statistical analyses included calculations of frequency data for VKS variables. That was computed for all target words in order to establish the levels of word knowledge demonstrated by each subject for each vocabulary item at the beginning and end of the treatment. The authors carried out separate calculations for discourse connectives versus content words. Additionally, the pre-post status of each word on the VKS was scored dichotomously in order to capture any gains in vocabulary knowledge. Vocabulary Knowledge scale scoring categories 1, 2 and 3 were combined into a "not known" category, while scoring categories 4, 5 and 6 were allocated into a "known" category. Scores were computed for each participant on the total number of "known" words, both for content words and discourse connectives, at the beginning and end of the treatment. ANOVA analyses and t-tests were carried out on these scores.

Paribakht and Wesche report that subjects answered the VKS with certainty, and the results indicated that it had captured progression in the knowledge of the target words. The authors also report some variation in subject interpretation of "I think" and "I know". They also admit that sentences produced by subjects in category 5 did not provide enough information to judge whether subjects knew a specific meaning of the target word. They report their attempt to correct these problems by clarifying the instructions for the VKS through emphasizing the key instruction words. They also added instruction for students to provide a meaning in

category 4 if they produced a sentence for category 5. However, despite the changes made to the initial version of the VKS, Paribakht and Wesche doubt the validity of the VKS and admit that a separate validity study on the VKS was needed.

The VKS assessment was followed by a number of tests: cloze tests, grammar tests and an introspection task. The intended analysis was to compare the other test results with those of the VKS as well as obtain student feedback via written comments in the introspection exercise.

The VKS results indicated that the Comprehension-based group achieved significant gains both on content words and discourse connectives in relation to the instructed themes only. The Four-skill class, however, demonstrated gains only on content words. The Comprehension-based group also demonstrated significant gains on the cloze test (based on the instructed theme) as well as gains in grammatical knowledge.

In conclusion, Paribakht and Wesche point out that instruction influences on the level of gains on different aspects of language knowledge. They report superior gains by the Comprehension-based class in text comprehension and discourse processing.

2.1.1.2 Discussion

This section focuses on discussing the issue of appropriateness of the Vocabulary Knowledge Scale (VKS) in its current form developed by Paribakht and Wesche as a measuring instrument for this study. This includes a few points: 1) misleading instructions and vague categories; 2) unjustified scoring system; 3) administering a measuring instrument (the VKS) that had not been validated.

1) Paribakht and Wesche do not seem to provide clear instructions to learners in regards to how the VKS should be used, neither do they distinctively outline its categories. For instance, there is an obvious contradiction in their report: initially, the authors claim that learners responded to the VKS with certainty. According to Paribakht and Wesche, the recorded changes in subjects' knowledge of the target words indicated that the VKS "captured progression in development of vocabulary knowledge" (p.18). However, later on Paribakht

and Wesche admit that subjects found the instructions for the VKS confusing and misleading. They appeared to especially struggle with categories 3 and 4. The ambiguous instructions for these levels "I think I know" and "I know" are reported to cause subjects hesitation and uncertainty. Unsurprisingly, on many occasions, subjects could not choose between these categories and indicate their degree of knowledge of a target word.

Furthermore, Paribakht and Wesche agree on the fact that it was necessary to merge the VKS categories ("known" and "unknown" words) for quantitative analysis as gains between adjacent categories could not be demonstrated statistically. If this was the case, it is not clear why Paribakht and Wesche included such vague categories as category 2: "I have seen this word before but I don't know what it means" and category 3: "I have seen this word before. I think it means_____" into their measuring instrument in the first place.

Also, Vocabulary Knowledge Scale categories 1, 2 and 3 are described as levels of relative familiarity with words for which the correct meaning is not known. It is not clear what Paribakht and Wesche imply under "relative familiarity" if the correct meaning is reported or proved to be "not known" and none of the other constituents of the word knowledge had been tested. Furthermore, the VKS contains five levels of word familiarity. However, the assumption that the five steps have equal intervals seems to be unjustified.

2) With respect to the scoring system used by the VKS in this study, Paribakht and Wesche report scoring adjustments to the self-report categories. This resulted in a 6-level scale. One point in category one "I have never seen this word", two points in category two "I do not know what it means" and two points in category three "I think it means..." for the wrong answer provided, seem to be bizarre and totally unjustified. It is hard to understand why negative and wrong answers should score anything but nil.

The authors argue that category three "I think it means..." was retained for those cases when it was impossible to judge whether the word was known. This is an odd claim. No examples or further explanations were provided. Furthermore, in Category 3, if a subject provided a wrong translation, the response was scored as category 2; in case of a correct answer it was scored as category 4. This scoring method seems to be seriously problematical as well.

All in all, the VKS assigns arbitrary scores to different outcomes, and Paribakht and Wesche do not justify why these scores are awarded. Indeed, the use of different scoring values seems to be a problem. It is difficult to interpret the data it generates. It is also difficult to judge how significant the gains are.

3) Paribakht and Wesche acknowledge the fact that the validity of the VKS needs to be fully examined. In this respect, it seems unwise to evaluate their instrument and use it as a way of measuring improvements at the same time. The more sensible approach would be to establish whether the VKS can act as a valid measuring instrument (by carrying out a validity study on it) before administering it in this way.

Furthermore, Paribakht and Wesche report a number of adjustments they had to make in order to improve the initial version of the VKS. Among them, clarifying instructions which now required that subjects, writing a sentence for category 5, would also provide a meaning in category 4. In a number of instances, sentences produced by learners for category 5 were of such a general character that they could not signify whether the subject knew what the target word meant. This modification was meant to make it easier to judge whether learners knew a meaning of the target word. However, subjects were still not required to supply more than one meaning sense for the word or produce more than one sentence with this word. Hence, this alteration did not really provide the vital information on the depth of learners' meaning knowledge of the words being measured.

In summary, the adjustments to the VKS reported by Paribakht and Wesche failed to modify this measuring instrument in such a way that it would be capable of reflecting on the complexity of learners' word knowledge. Although they admit that "certain refinements are required" (p.18), the scale of these refinements appears to be significantly larger than acknowledged by Paribakht and Wesche.

2.1.1.3 Conclusion

Paribakht and Wesche show a critical approach to the self-assessment instrument they created. They acknowledge some of the apparent drawbacks in its structure and comment on difficulties in its use. However, despite the admitted downsides of their self-assessment instrument, Paribkht and Wesche administer the Vocabulary Knowledge Scale for measuring

gains in learners' word knowledge. Moreover, they make general conclusions in relation to methodologies they are testing by means of the VKS (for instance, a teaching approach which emphasises global comprehension of written and oral texts in the current paper).

2.1.2 Paribakht and Wesche. 1995.
Paribakht and Wesche. 1996.

2.1.2.1 Summary

Both articles report on the same study which investigates the validity of a classification scheme for reading-related vocabulary exercise types developed by Paribakht and Wesche. The classification scheme suggests "a hierarchy of the degree and type of mental processing by various kinds of vocabulary exercises" (1996; p.155). The categories of the classification scheme are: selective attention, recognition, manipulation, interpretation, and production. The exercises under these categories aimed at facilitating learners' perception of unfamiliar target words as well as achieving higher levels of knowledge of the words in future. The authors argue that this scheme reflects a view of initial vocabulary acquisition as a multistage process which involves repeated exposures to new words in meaningful contexts (1996).

The research investigates incidental learning of the target words through multiple exposures when reading. It aimed to compare vocabulary acquisition in two learning modes: through Reading Only and through Reading Plus which involved initial reading exposure followed by focused vocabulary activities. A number of questions were raised:

1. Does thematic L2 reading which provides multiple exposures to the new words lead to measurable gains in the knowledge of these words?
2. Can instructional intervention enhance such incidental L2 vocabulary acquisition?
3. What types of text-related vocabulary exercises are appropriate, and according to what rationale?
4. How can gains in vocabulary knowledge be measured, both in number of new words acquired and the depth of knowledge about them?

The Vocabulary Knowledge Scale (VKS), developed by Paribakht and Wesche was administered in this study to self-check the results in both modes. It included five self-report categories ranging from total unfamiliarity with the word and some idea of its meaning,

to the claimed ability to use the word with grammatical and semantic accuracy in a sentence (Table 2-1).

Table 2-1: VKS self-report categories, scores and their meanings

Self-report Categories	Possible Scores	Meaning of Scores
1. I don't remember having seen this word before	1	The word is not familiar at all
2. I've seen this word before, but I don't know what it means	2	The word is familiar but its meaning is not known
3. I've seen this word before, and I think it means ____(synonym or translation)	3 2	A correct synonym or translation is given (A wrong answer is given)
4. I know this word. It means__ (synonym or translation)	3 2	A correct synonym or translation is given (A wrong answer is given)
5. I can use this word in a sentence___ (Write a sentence) *(If you do this section, please also do section 4)*	5 4 3 2	The word is used with semantic appropriateness and grammatical accuracy in a sentence The word is used with semantic appropriateness in a sentence A correct synonym or translation is given A wrong answer is given

For the Reading Plus treatment, a system of vocabulary exercises was developed. The exercises were grouped into five categories according to "the presumed type of mental processing" each required.

Proposed Classification Scheme For Vocabulary Exercises:

Selective Attention: draws learners' attention to target words (ex. Underlining, bold-facing, circling);

Recognition: requires association of the written target word form with at least one of its meanings (ex. Matching word with definition or synonym);

Manipulation: requires structural analysis of the target word to rearrange given elements (ex. Constructing words using stems and affixes);

Interpretation: involves semantic and syntactic analysis (ex. Multiple choice cloze exercises);

Production: requires retrieval and production of the target word in appropriate context (ex. Open cloze exercises, answering a question requiring the target word).

In the Reading Plus Treatment, only three exercises of the proposed scheme were used for each target word.

The experiment was carried out in 1992-93. Subjects, 38 L2 students, received both instructional treatments and acted as their own controls. The number of the target words the subjects were able to recognize and the depth of their knowledge were assessed before and after treatments.

Target words included 28 items (12 nouns and 16 verbs) for the Reading Plus Treatment and 30 items (12 nouns and 18 verbs) for the Reading Only Treatment. Four topics were selected for treatments: two topics, Media and Environment, for the Reading Plus Treatment and another two, Fitness and Biological Revolution, for the Reading Only Treatment. In each treatment, subjects received four main texts of approximately 500 words each.

The experimental procedure was arranged as follows. In the Reading Plus Treatment, students read the texts, answered comprehension questions and completed vocabulary exercises created for the target words according to different levels of the classification scheme. In the Reading Only Option, subjects also read texts and answered text-related comprehension questions. They were not given vocabulary exercises. Instead, they received another text on the same topic which included the same target words. After reading the new text, subjects had to answer a new set of comprehension questions based on the text. The authors point out that the number of exposures to the target words in each treatment were approximately equivalent. T-tests were employed by the authors to check the gains in vocabulary knowledge for both treatments.

Paribakht and Wesche report that the total results of the study indicated that while both treatments led to gains in vocabulary knowledge, the Reading Plus Treatment provided significantly greater improvement in the subjects' word knowledge, both quantitative and qualitative. In other words, subjects acquired more words in the Reading Plus Treatment as well as achieving "greater depth in their knowledge of the target words". The overall results for both treatments are illustrated in the table below (p.54).

Table 2-2: Overall vocabulary gains for both instructional treatments

Variables	Reading Plus Vocabulary			Reading Only		
	N	Pre	Post	N	Pre	Post
Nouns	31	47.2	69.5	27	45.0	61.2
Verbs	31	45.9	65.0	27	51.6	60.1
Total Content Words	31	46.4	66.9	27	48.9	60.6
Discourse Connectives	33	51.8	64.2			

These greater lexical gains in the Reading Plus Treatment were attributed to the use of exercises. The authors claim that suggested exercises, different in type, directed learners' attention to specific lexical items. Furthermore, the exercises required learners to analyze meanings and functions of the target words. This, in turn, according to Paribakht and Wesche, required more varied kinds of processing of the words by subjects (than did multiple encounters through reading) which led to greater gains in learners' vocabulary knowledge.

2.1.2.2 Discussion

This section discusses the issue of the vocabulary assessment instrument, the VKS, used in the study. Two points are emphasised. The first point is our argument that the VKS cannot be used to measure the depth of word knowledge. The second point outlines further drawbacks of the VKS as a vocabulary assessment tool.

1. Among the research questions set up by the authors for the current study, question number 4 asks: How can gains in vocabulary knowledge be measured, both in number of new words acquired and depth of knowledge about them? Paribakht and Wesche's answer to this question seems to be obvious -- using the Vocabulary Knowledge Scale (VKS) which was elaborated by the authors themselves in 1993. However, as was specified in the research question, the vocabulary assessment instrument should be able to measure both quantitative (amount of new words learned) and qualitative (depth) gains in the word knowledge.

As for the VKS, there is an apparent contradiction in Paribakht and Wesche's outline of its functions. First, at page 48, Paribakht and Wesche claim that the Vocabulary Knowledge Scale (VKS) was used to track vocabulary knowledge gains in the two treatments. That included changes in both the number of words learned and the depth of knowledge attained. They argue that the Vocabulary Knowledge Scale was able to estimate gains in learning of particular words as well as discriminate between different levels of knowledge of these words. In other words, the VKS was meant to capture quantitative (number of words known) and qualitative (depth of vocabulary knowledge) changes in initial vocabulary knowledge. However, later on, at page 50, the authors admit that the VKS does not measure different meanings of a multiple-meaning word, neither does it measure different types of word knowledge.

Then, the question rises what meaning Paribakht and Wesche implied by "depth of word knowledge". As discussed in the previous chapter the concept of depth is complex and multi-dimensional. The knowledge of meaning senses of the word as well as other aspects of word knowledge, such as spelling, word class, derivatives, grammar characteristics etc. are essential constituents of the depth of vocabulary knowledge (as discussed earlier in this thesis, section 1.4.3). The authors claim that their scale could be expanded to measure other aspects of word knowledge, however no further explanation or examples are provided. Furthermore, it should be taken into account that the more points the scale provides the more difficult it is to ensure that it works effectively.

2. The Vocabulary Knowledge Scale is a multi-leveled self-assessment instrument. Let's look at each of its levels in turn. The first level (or self-report category) "I do not remember having seen this word before" is awarded 1 point with the commentary "The word is not familiar at all". It is not clear why the authors chose to award 1 point for the total absence of knowledge of the word. Clearly, the more appropriate score for this level is 0 points since there is no knowledge to be assessed.

The second level "I have seen this word before but I do not know what it means" is awarded 2 points with the explanation: "The word is familiar but its meaning is not known." Generally speaking, option 2 in the self-report categories does not seem to be different from option 1, though awarded two points, in terms of demonstrating word knowledge. It does not seem wise

to make conclusions whether the word is familiar to the student judging on their comment that they have seen the word before. Since none of the aspects of word knowledge, such as spelling, derivatives, grammar forms, etc. has proved to be known by the subject or even claimed as known, it does not seem right to state that the subject is familiar with the word. It may be the case that the learner got the target word mixed up with a different vocabulary item or chose this option simply because it sounds "more knowledgeable" than option 1 and does not involve any checks of the actual word knowledge. Furthermore, it does not seem reasonable to award two points for non-existing knowledge of the word. In fact, two points implies that this knowledge is twice the zero knowledge recorded previously.

The third level: "I've seen this word before, and I think it means_____ (synonym or translation)" and the forth level: "I know this word. It means_____ (synonym or translation)" are awarded the same three points for the correct answer and unjustified two points if the supplied synonym or translation was incorrect. Again, it is far from clear why Paribakht and Wesche need to differentiate between the statements "I think it means...." and "It means....." which represent different levels of word knowledge in their elicitation scale. It does not seem logical to emphasise this difference since degrees of certainty in the knowledge of the target word are not reflected in the score.

As for the fifth and final level "I can use this word in a sentence:_____ (Write a sentence.) *If you do this section, please also do section 4*" which is supply a correct synonym or translation for the target word. Five points (the maximum score) are awarded for using the word with semantic appropriateness and grammatical accuracy in a sentence. Bearing in mind that Paribakht and Wesche reward total absence of knowledge of the target word with one point (Level 1) and a disputable trace of this knowledge with two points (Level 2), five points does not seem to be a fair score for a complete right answer provided. Also, I must stress that the maximum score of five points does not reflect on students' knowledge of different meanings of the same target word, neither does it show their knowledge of different aspects of the word knowledge. And again, the VKS does not justify the scores awarded.

Generally, the Vocabulary Knowledge Scale suggested by Paribakht and Wesche does not take into consideration the complexity of vocabulary knowledge, the numerous aspects (such as, spelling, derivatives, synonyms, etc) which should be tested as integral constituents of

word knowledge. A further problem here is the fact that Paribakht and Wesche are both evaluating their scale and using it as a measurement instrument to investigate the assumptions of their research. This seems to be an odd approach. The normal practice would be to establish that the instrument actually works before it is utilized in this way.

2.1.2.3 Conclusion

In conclusion, we would like to note that self-assessment of vocabulary knowledge is an idea with great potential. We believe self-assessment might become an effective and reliable instrument of measuring learners' word knowledge. Having said that, we need to emphasize that the Vocabulary Knowledge Scale used in this study suffers from several drawbacks:
1) the outlined self-report categories do not reflect learners' actual levels of word knowledge;
2) the scoring system is not motivated;
3) the meanings of scores do not always correspond to the descriptions of the categories (e.g. self-report category No2);
4) the knowledge of different meaning senses of the target word is not tested;
5) the knowledge of different aspects of word knowledge is not assessed.

2.1.3 Wesche and Paribakht. 1996.

2.1.3.1 Summary

The article surveys second language vocabulary measures, "depth" and "breadth" tests, and describes the measuring instrument, the Vocabulary Knowledge Scale (VKS), developed by Paribakht and Wesche. The authors argue that it was designed to assess levels of familiarity with the target words. The article is based on theoretical speculation around Paribakht and Wesche's VKS. No experimental research is reported here.

Analysing the existing measuring instruments, Wesche and Paribakht point out that they all focus on breadth vs. depth. Among the breadth measures, they discuss the following vocabulary testing methods:
1) identification of written synonyms or definitions for each target word, such as a multiple choice test; 2) writing a dictated test containing target words; 3) C-tests requiring learners to

fill in the missing half of every other word in some sentences of a text; 4) listening tests -- judgment of real words; 5) matching tests; 6) error recognition; 7) compositions; 8) oral interviews with vocabulary components; 9) checklists; 10) descriptive scales. The last two methods estimate the self-perceived knowledge of certain words.

Describing tests using the checklist format Wesche and Paribakht note that learners have to indicate whether or how well they know each word on the list presented. In other words, it is a kind of self-reporting test: "known" versus "unknown" or "real" versus "nonsense" words. The imaginary words are included for adjusting scores against overestimates: "a correction factor based on a percentage of these is calculated into the final score" (p. 20). Wesche and Paribakht emphasise the positive sides of a checklist format: easily administered, computerized, automated and self-scoring. However, at the same time, the authors point out that the checklist tests are seriously criticized for the following shortcomings: unreliability, overestimation and lack of ability to test different aspects of word knowledge.

Another instrument for self-perceiving measurement of the word knowledge discussed by Wesche and Paribakht, is descriptive scales. They review the use of descriptive scales which is mostly to guide rater judgments of vocabulary features. They present Fulcher's (1988) criticism of this instrument. Fulcher (1988) argues that the terms in descriptive scales are not specific and the underlying concepts are too vague.

Continuing to speculate on vocabulary breadth measures, Wesche and Paribakht note that these evaluating instruments seek to assess the size or range of the students' vocabulary measuring demonstrated or self-perceived knowledge of particular (target) words. However, the authors emphasise that these breadth measuring instruments would be of limited use in measuring how much is known about the target words. For this purpose, vocabulary depth measures are used.

The paper discusses a number of issues related to partial knowledge of words, aspects and levels of word knowledge. Summarizing research on depth of vocabulary knowledge, Wesche and Paribakht note the importance of partial knowledge of words for L2 users. They mention various aspects of word knowledge described in lexical research, such as: meanings, appropriate uses, syntactic properties, forms and derivations, association network and

connotations. They also speculate on different levels of vocabulary knowledge ranging from initial recognition of the form of a given word, to an understanding of its meanings in context, to the ability to use this word in context. They briefly describe levels in the word knowledge suggested by Beck, Perfetti and McKeown (1982): unknown, acquainted and established knowledge. Wesche and Paribakht cite some rare efforts to assess different kinds of word knowledge. They describe Nation's (1990) four dimensional table of word knowledge measuring form, position, function and meaning. They outline the previous attempts to elicit self-assessment of vocabulary knowledge by Read (1989) pointing out that no testing procedure was suggested. They go on to describe word associates' tests carried out by Read (1993 and 1994) which investigate different kinds of word knowledge categories of familiarity with given words, ranging from being able to define the word, through selecting an appropriate use, to being able to use it (Cronbach, 1942).

They also describe the four-point self-assessment scale proposed by Dale (1965) to use with L1 students. These stages are:
Stage 1: "I never saw it before."
Stage 2: "I've heard of it, but I don't know what it means".
Stage 3: "I recognize it in context -- it has something to do with..."
Stage 4: "I know it".
Having analyzed the research carried out in this field, Wesche and Paribakht propose their own instrument for self-assessing learners' vocabulary knowledge, the Vocabulary Knowledge Scale (VKS). The authors claim that the VKS captures certain stages in initial development of core knowledge of particular words. These stages, according to Wesche and Paribakht, represent gains which are large enough to be meaningful on a self-report scale but too small to reflect changes in the actual knowledge. This seems to be vague and arbitrary as the border line between "large enough" and "too small" is totally unclear.

In the VKS description, Wesche and Paribakht state that the instrument is based on a scale which combines self-report and performance items to establish both self-perceived and demonstrated knowledge of particular words. The levels of the self-report scale range from total unfamiliarity, through recognition of the word and some idea of its meaning, to the ability to use the word with grammatical and semantic accuracy in a sentence. The VKS seems to bear some resemblance to the Dale's scale (1965) described earlier. However,

Wesche and Paribakht argue that their scale was developed without knowledge of the latter. The article briefly describes a pilot study in which the VKS was used for the first time (see Paribakht and Wesche, 1993). The authors continue describing the VKS and present its elicitation scale and the five scoring categories:

VKS Elicitation Scale Self-Report Categories

Self-report Categories

1. I don't remember having seen this word before.
2. I have seen this word before, but I don't know what it means.
3. I have seen this word before, and I think it means_____ (synonym or translation).
4. I know this word. It means_____ (synonym or translation)
5. I can use this word in a sentence:_____ (Write a sentence.) *(If you do this section, please also do Section 4).*

The VKS scoring categories and meaning of scores are shown in table 2-3.

Table 2-3: VKS Scoring categories: meaning of scores

Self-report Categories	Possible Scores	Meaning of Scores
1. I don't remember having seen this word before	1	The word is not familiar at all
2. I've seen this word before, but I don't know what it means	2	The word is familiar but its meaning is not known
3. I've seen this word before, and I think it means ____(synonym or translation)	3 2	A correct synonym or translation is given (A wrong answer is given)
4. I know this word. It means__ (synonym or translation)	4 3 2	The word is used with semantic appropriateness in a sentence A correct synonym or translation is given (A wrong answer is given)
5. I can use this word in a sentence___ (Write a sentence) *(If you do this section, please also do section 4)*	5 4 3 2	The word is used with semantic appropriateness and grammatical accuracy in a sentence The word is used with semantic appropriateness in a sentence A correct synonym or translation is given A wrong answer is given

With regards to the scoring system, Wesche and Paribakht state that VKS scoring uses a combination of self-reported and demonstrated knowledge. They carry on explaining

the meanings of scores: "Elicitation categories 1 and 2 lead to level 1 and 2 scores respectively. Category 3 may lead to a score of 2 (if the synonym or translation is wrong) or of 3 (if it is correct). Wrong responses in category 4 and category 5 likewise result in a score of 2". (p.30). Furthermore, if category 5 is chosen to report the knowledge of a word meaning but the word is inappropriately used in a sentence, a score of 3 is awarded. A score of 4 is awarded if the word is used with semantic appropriateness in the context but with some grammar errors. A score of 5 indicates that the word is used both semantically and grammatically correctly in a sentence.

In order to determine the extent to which self-report data might be relied on, the self-report scores were correlated with the actual scores for demonstrated knowledge of the target words. The authors claim that these scores were highly correlated (.96/.97 by text and by theme). Wesche and Paribakht also report on a VKS reliability estimate through test-retest administration at a Canadian university during the 1992 ESL summer courses. They claim that the resulting Pearson test-retest correlation was .89 for the 24 content words and .82 for the eight discourse connectives. However, later on, in Validity section, citing the validity analysis of VKS results with the Eurocentres 10k Vocabulary Size Test, Wesche and Paribakht present the results which show low correlations between VKS and EVST scores. In conclusion, Wesche and Paribakht admit a large number of limitations for the use of the VKS and speculate about a possible extension of the scale. However, at the same time they insist that the VKS captures initial stages and levels in word learning. They emphasise that an advantage of the VKS over the YES/NO type of self-assessment tests is the possibility of verifying actual knowledge against perceived knowledge.

2.1.3.2 Discussion

This part of the review highlights the issue of inconsistency and ambiguity in descriptions of tools and terminology used in the Vocabulary Knowledge Scale developed by Wesche and Paribakht. It discusses a number of examples of this ambiguity and inconsistency within the scale: 1) scoring system; 2) terminology; 3) inconsistent claims in regards of VKS abilities.

1) Let us take a look at their scoring system, for instance. Undoubtedly, a scoring system is one of the vital mechanisms in any assessment instrument. However, Paribakht and Wesche's scale suffers from a number of drawbacks, some of which are discussed in my reviews of

Paribakht and Wesche's earlier studies (1993). The VKS scoring presented in the current report does not appear to follow a logical sequence. Thus, for instance, self-report category No 4: "I know this word. It means_____ (synonym or translation)" may lead to a score of 2 (if the synonym or translation is wrong), a score of 3 (if it is correct), and also to a score of 4 (if the word is used with semantic appropriateness in a sentence). There is an obvious lack of correspondence between the description of self-report category No 4 which states: "I know this word. It means_____" (only one meaning is required!) and the score of 4 awarded for the ability to use the word in a sentence which is not required within this category. Furthermore, Wesche and Paribakht note that in some situations, unconverted self-report categories can be used as scores (p.34). This adds even more confusion to the scale's scoring system.

2) The inconsistency can also be traced in the terminology used for the VKS, for example the notions: "section, "category" and "level". They seem to substitute each other in different parts of the VKS description. Although it might be guessed that the notion "level" corresponds to the notion "scoring category", since Wesche and Paribakht state that elicitation categories lead to particular level scores. However, no explanation is provided. Moreover, the terms "category" and "level" seem to be used interchangeably in the authors' descriptions of the scale.

Furthermore, Wesche and Paribakht report their "striking finding" discovered during the VKS testing. The scale was administered twice within two weeks using the same target words. The authors note that more than 50 percent of the subjects who chose category 1("I have never seen this word before") in the first round, chose it again in the second round rather than the expected category 2 ("I have seen this word before but I don't know what it means"). This raises a question of vague category descriptions when learners struggle to differentiate between very similar categories of the Vocabulary Knowledge Scale.

3) In this study, Wesche and Paribakht declare that the primary goal in developing the VKS was to capture initial stages in word learning. In other words, according to Wesche and Paribakht, the VKS could capture progression in the development of knowledge of particular words at the initial stage. Generally speaking, the VKS sought to reflect gains in vocabulary knowledge during a brief instructional period. However, later on, in the same report, and in

their previous (1993) and subsequent studies in which the VKS was used as a measuring instrument, Wesche and Paribakht argue that the VKS was used to measure depth of word knowledge. There seems to be a contradiction here. It is readily apparent that gains at initial stages of development of the word knowledge represent superficial (initial) information about the word. Hence, it seems unwise to speculate on the depth of knowledge of the word at this stage.

Furthermore, reviewing the research on the depth of vocabulary knowledge, Wesche and Paribakht emphasise the importance of measuring various aspects of word knowledge, such as: meanings, grammatical properties, derivations, association network and collocations. They acknowledge the point that the depth of word knowledge is determined by the fact of how well learners know different aspects of this knowledge. In other words, it is defined by how much information they possess in regards to the words claimed as known.

Bearing this in mind, it needs pointing out that the VKS in its form suggested by Wesche and Paribakht does not tap different meanings of the same word, neither does it tap knowledge of other different aspects of word knowledge. This was admitted by the authors themselves (p.33). However, if this is the case, the VKS can hardly be referred to as a depth measuring instrument. Although, Wesche and Paribakht argue that an extension of the VKS might improve its ability to measure more constituents of learners' word knowledge, they also admit a number of difficulties this might lead to, e.g. reduction of administrative feasibility and scoring difficulties. Furthermore, it might not be wise to entirely agree with Wesche and Paribakht's claim that the VKS captures certain stages in initial development of knowledge of given words. The initial development of knowledge about a certain word might not necessarily begin with acquiring its meaning senses (the VKS assessment is based on). Quite the contrary, totally different aspects or components of knowledge of a particular word, its sound or written form for instance, may be learned first.

2.1.3.3 Conclusion

In conclusion, we emphasise that Wesche and Paribakht's VKS suffers from a lack of consistency as well as inadequate use of notions. Moreover, there is an apparent contradiction in regards to the VKS description. These issues need further consideration and alterations.

The limitations of use of the scale are pretty obvious and admitted by the authors themselves.

2.1.4 Paribakht and Wesche. 1997.

2.1.4.1 Summary

The article reports on a number of studies conducted by Paribakht and Weche. They explore the issue of instructional intervention aimed at supporting the process of word learning through reading. The classroom experiments reported here involve instructional procedures which were designed to increase the salience and cognitive processing of the words encountered by L2 students in reading texts. The Vocabulary Knowledge Scale was used in these studies to track the acquisition of the targeted words.

The purpose of this research was to investigate the role of various vocabulary instruction techniques that were based on reading texts. The research question was: Can tasks be designed that will increase the effectiveness of vocabulary learning through reading practice? In other words, the authors investigate whether reading comprehension accompanied by vocabulary exercises would lead to greater gains in selected words than reading additional texts. There were four hypotheses outlined for the studies. Among them, were the following: under the same time conditions, gains for the reading plus vocabulary instruction will be greater than for the reading only treatment; vocabulary gains will be both quantitative (the number of words known to some degree versus not known) and qualitative (increased depth of knowledge of the target words).

The studies were conducted in the context of the comprehension-based ESL program at the University of Ottawa. Subjects, students at the bilingual University of Ottawa's Second Language Institute, were exposed to each of the two instructional treatments: Reading Only conditions and Reading Plus vocabulary instruction. In the Reading Only treatment, students read selected texts on two themes and answered comprehension questions. After reading each main text subjects also read a supplementary text which included the target words from the main text. In the Reading Plus treatment, subjects likewise read selected texts on two themes and answered comprehension questions. Then instead of reading supplementary texts they completed a series of vocabulary exercises based on the target words from the main texts. The Vocabulary Knowledge 5-point scale was used as a measuring instrument to elicit self-

perceived and demonstrated knowledge of the targeted words. Again, subjects served as their own controls. They were given a list of target words and asked to indicate their level of knowledge for each word. They were also required to demonstrate their knowledge for self-report levels 3-4. The scores for a word and the self-report categories are illustrated in table 2-1 (p.31).

As seen from the table, wrong responses in self-report categories 3, 4, or 5 led to a score of 2. A score of 3 was awarded for an appropriate synonym or translation in self-report categories 3 or 4. A score of 4 was given if the word was used in a sentence which demonstrated subject's knowledge of its meaning in that context but with incorrect grammar (e.g., a target noun used as a verb: "*This famous player announced his retire*"). A score of 5 was awarded for both semantically and grammatically correct use of the target word. This table differs from the one published in Wesche and Paribakht (1996) (Table 2-3, p.39) in which an obvious error occurred for self-report category No 4: "I know this word. It means_____ (synonym or translation)". The authors claimed the response may lead to a score of 2 (if the synonym or translation is wrong), a score of 3 (if it is correct), and also to a score of 4 (if the word is used with semantic appropriateness in a sentence). The score of 4 was given for the ability to use the word in a sentence which was not required within that category. However, as we can see that was amended in the version of VKS used in this study.

Paribakht and Wesche report a pilot study carried out with 17 students in 1992. The materials were based on four themes and included two main reading texts and four additional texts for the Reading Only treatment. The word list consisted of 77 Reading Plus treatment words and 61 Reading Only treatment words. Paired t-tests were administered on the pre-post data. The authors report significant gains on the Reading Plus treatment word list ($p < .05$). They claim that the vocabulary enhancement techniques used in the study promoted increase in the acquisition of the targeted content words. However, the authors did not seem to administer their Vocabulary Knowledge Scale in this pilot study as no reference to the VKS is provided.

The main study, reported next, was carried out in 1993. This study was also reported in Paribakht and Wesche (1995 and 1996) and has been reviewed earlier in our research. However, since we discuss the information on the VKS provided in this article, a brief recap of this experimental study is necessary here. Thirty eight subjects were involved in two

treatments: the Reading Only mode and the Reading Plus treatment.

In the Reading Plus treatment subjects read the given texts and answered the comprehension questions at home. In class, the comprehension exercises were corrected, and the vocabulary exercises for the given texts were completed and corrected. Vocabulary exercises in the Reading Plus were arranged into five categories: selective attention, recognition, manipulation, interpretation and production.

In the Reading Only treatment, subjects also read the selected texts and answered the comprehension questions at home, followed by correction in class. However, the vocabulary exercises were replaced by supplementary texts accompanied by the comprehension questions. That was completed in class. The idea of these additional texts was to expose the subjects to the target words again through reading and completion of the comprehension questions.

Gains in subjects' word knowledge were measured using the VKS and t-tests. The following results were reported by the authors:

1. Both treatments produced gains for all types of the target words.
2. The Reading Plus treatment brought significantly greater gains for content words than the Reading Only.
3. Gains in the Reading Plus treatment were greater for content words than for discourse connectives.
4. Vocabulary gains were both quantitative and qualitative.

At the end of the instructional period, subjects were asked to indicate the usefulness of vocabulary activities on a 5-point scale and explain their reasons for their ratings. Subjects were also asked to state how many new words they had learned. The questionnaire included the following types of vocabulary exercises offered to the subjects:

A. Read word list and notice the same words in the text (selective attention).
B. Match word list with list of definitions (recognition).
C. Find words in the text and match them to definitions (interpretation).
D. Replace words with target words from text (interpretation).
E. Classify connectives by type (interpretation).

F. Unscramble words to form sentences (production).

Paribakht and Wesche report that in general, subjects found the vocabulary exercises suggested by the authors helpful and useful, and produced the following order of usefulness: B, C, D, A, F and E. They also stated that the percentage of the target words students thought they had learned ranged from 30 to 100, averaging 65. That matched teachers' estimates. Students in the Reading Only treatment acquired a number of target words. However, their knowledge of many of those words was described as the knowledge at the recognition level. While in the Reading Plus treatment, learners were reported to have learned more words and reached higher levels of knowledge of these words.

The overall conclusion was that focused vocabulary instruction provided by Paribakht and Wesche based on theme-related reading texts appeared to be more effective than reading comprehension alone.

2.1.4.2 Discussion

This part of the review further discusses one of the major weaknesses of the VKS: the fact that it does not reflect on the variety of aspects of word knowledge. This becomes an even more serious drawback of Paribakht and Wesche's self-assessment scale in the light of their repeated claim that the VKS was used to measure the depth of the subjects' word knowledge. The authors claim that the VKS revealed dramatic decreases in the number of words allocated under category 1 (never seen) in both treatments. Most of those words are reported to have moved to category 2 (seen but still unknown) after the Reading Only treatment. It is not clear why Paribakht and Wesche call this category the recognition level knowledge. It is understood that the word placed into this category remains unknown to the subject. Furthermore, an even higher number of words are reported to have moved to categories 3 through 5 after the Reading Plus treatment. Based on this data, Paribakht and Wesche claim that many subjects in the Reading Plus treatment passed the recognition level for most of the target words. Again, this claim does not seem to be correct. Let's take another look at category 3: only a simple translation of a target word is required here. Hence, technically this stage in subject's knowledge should still be regarded as the recognition level knowledge.

Moreover, the conclusion drawn by Paribakht and Wesche from this data is even more

unsubstantiated. The authors claim that judging from the word relocations shown on the scale, subjects in the Reading Plus treatment achieved greater depth in their knowledge of the target words. Surely, moving from category 1(never seen before - totally unknown) to category 3 or category 4 or even category 5 is moving from "unknown" to "known to some extent". This does not seem to indicate the depth of knowledge of a word. Furthermore, assessment of the latter is meant to involve measuring different features or constituents of the word knowledge (as discussed earlier in this thesis, section 1.4.3) However, as argued in our previous reviews and admitted by the authors themselves, the VKS, in its current state, does not measure different features of word knowledge.

With regards to the importance of measuring various types of word knowledge, the authors list a number of errors made by learners while completing the VKS. Had they been investigated in more detail, those errors would have provided valuable information on how different aspects of word knowledge established themselves when a new word was learnt. However, no further analysis was carried out. For instance, explaining the scoring system Paribakht and Wesche wrote: "A score of 4 is given if the word is used in a sentence demonstrating the student's knowledge of its meaning in that context but with inaccurate grammar" (p. 180). "Inaccurate grammar" sounds like a very general statement. Inaccurate grammar might imply a variety of errors within a large number of aspects of word knowledge which need to be tested separately in order to establish the depth of this knowledge. Let's take a look at some errors cited by the authors: "*This famous player announced his retire.*" This example indicates a lack of knowledge of the derivative forms for the verb *to retire*. It also hints that the learner might have failed to determine the word class of this target word. Thus, this error points, at least, at two aspects of the word knowledge: derived forms and word classes. Had Paribakht and Wesche analyzed this data separately and reflected on the results in their categories and scoring system, they could have measured the developing knowledge of the target words more precisely. Furthermore, had that been achieved, the VKS would have been of more use in terms of assessing the depth of word knowledge. Similarly, an error "losed" for "lost" could be considered as a valuable source of grammar information on the target word "to lose". That should have been reflected on the VKS and analyzed in more detail.

Overall, these errors appear to be a good source of information on how learners acquire new

words and how the word knowledge develops. Unfortunately, Paribakht and Wesche did not provide a list of the most common mistakes that occurred in students' self-report categories. However, judging by the list of the target vocabulary items selected for each text, such as *to arouse, to loosen, to be aware of, to trigger, behaviour, disapproval, retaliation* and others, the data collected by the authors might have shown gains in different aspects of the word knowledge if analyzed separately. Furthermore, it would be interesting to see how errors spread over different word classes and whether more than one meaning sense had been learnt. Moreover, since the focus of this study was vocabulary acquisition through reading, it would have been beneficial for the research to establish how different aspects of word knowledge were acquired in different Reading Treatments. Since Paribakht and Wesche report superior gains in the Reading Plus treatment, it may be reasonable to investigate whether these gains occurred in different aspects of word knowledge and how it is compared to the results in the Reading Only treatment. In this respect, post-VKS interviews might shed some light on these issues.

2.1.4.3 Conclusion

Having reported significant gains in vocabulary knowledge in both Reading treatments, Paribakht and Wesche did not provide sufficient evidence of qualitative changes which were claimed to have occurred in the developing knowledge of the targeted words.

2.2 General conclusions on the structure and use of Paribakht and Wesche's VKS

The survey of Paribakht and Wesche's studies which administered the Vocabulary Knowledge Scale designed by the authors highlights a number of common problematic issues in regards to the VKS methodology. These issues are summarized in the current chapter.

2.2.1 Issue 1.

The VKS as a valuating and evaluated instrument at the same time.

In this section, we argue that it does not seem wise to evaluate the instrument and use it as a way of measuring improvements at the same time.

As admitted by Paribakht and Wesche (1993-1997), the validity of the Vocabulary Knowledge Scale was tested while it was used as a measurement instrument to verify hypotheses or collect data which were not directly related to the VKS itself. This approach does not seem to be justified. Furthermore, this seems to be a serious issue for the following reason. In the case of employing a measurement instrument which has not passed all the preliminary checks on validity, any information collected by means of this instrument can hardly be regarded valid or reliable. Let us recap a number of instances, illustrating this issue.

The study conducted by Paribakh and Wesche in 1993 investigates a teaching approach based on global comprehension of written and oral texts using the VKS as a measurement instrument. In that study, however, at the same time, Paribakht and Wesche doubt the validity of the VKS and admit that a separate validity study on the VKS was needed. For that reason, the assessment by the VKS was followed by a number of tests. The intended analysis was to establish the validity of the VKS results by comparing them with the results of the other tests. Discussing the results of that study Paribakht and Wesche acknowledge the fact that the VKS required "certain refinements". The adjustments were made. However, despite the authors' earlier comments, no separate validation study on the VKS was conducted. In subsequent studies, Paribakht and Wesche continue using the scale as a means of verification. Thus, for instance, in their study, published in 1997 (4 years after their first admission that the VKS needed separate validity checks), Paribakht and Wesche test the validity of a classification scheme for reading-related vocabulary exercise types developed by Paribakht and Wesche. And again, the VKS was used as a validation instrument despite the fact that it had not undergone validity checks itself, in the first place.

Conclusion. It is important for the researcher to employ recognized (acknowledged) means of assessment in their research in order to be able to argue the validity of the results obtained.

2.2.2 Issue 2

The VKS as a self-assessment instrument to measure the depth of word knowledge.

This section argues that the knowledge assessing functions of the VKS are very limited. Briefly, they consist of measuring knowledge of one meaning sense of the given word as well as the ability to use that word in the meaning sense tested, in a sentence.

The claim regarding the ability of the VKS to measure depth of word knowledge appears to be unjustified due to the following reason. As discussed earlier (section1.4.3 of this thesis) the phenomenon "depth of word knowledge" implies the whole range of information about the word in question, in other words, all aspects or kinds of knowledge in regards of this particular word stored in the lexicon (Nation, 1990 and 2001). In this respect, it seems the more questions asked about the word (by the researcher), a more accurate picture of learners' knowledge of this word emerged. Paribakht and Wesche, however, ask two questions only: Question No1 is: Can you provide a translation or a synonym for a given word? And Question No 2: Can you write a sentence with this word? These questions do not seem to completely cover even two types of word knowledge. Question No 1 aims at checking knowledge of only one meaning of a target word. Likewise, question No 2 requires learners to produce a sentence with the word illustrating one of its meanings only. Thus, the two kinds of knowledge measured by the VKS are: knowledge of one meaning sense (including polysemous words) and knowledge of how to use this word in a phrase. This does not seem to incorporate even minimum information about the target word. Despite the fact that Paribakht and Wesche speculate about a variety of word knowledge types while surveying research on depth of vocabulary knowledge (1996), the diversity and complexity of the phenomenon "word knowledge" has not been reflected in the framework of their scale. In other words, none of the further aspects of word knowledge, such as further meaning senses (for polysemous words), inflectional and derivational forms, word class belonging, synonyms/antonyms, words' sound and written forms, and so forth are assessed by Paribakht and Wesche's VKS. Though the authors claim that their scale could be expanded for more precise word knowledge testing, no further information is provided.

Also, discussing the issue of word knowledge depth, Wesche and Paribakht (1996) emphasise the importance of identifying partial knowledge of words which is meant to be elicited by means of their VKS. However, again, the questions: How much knowledge? and which knowledge is enough?... remain unanswered.

What we should be looking for here are measurement procedures that could capture degrees of knowledge of various word knowledge aspects, in other words, establish how deep this knowledge is, for each of the targeted words. Indeed, in order to reveal how much is known about a certain word different aspects or types of this knowledge should be measured by a self-assessment scale. Provided this information is gathered and evaluated, further conclusions on the depth of lexical knowledge could be drawn.

Furthermore, in the discussion sections of our survey, we noted a contradiction between Paribakht and Wesche's claims that:
1) the VKS aimed at capturing gains at initial stages of development of knowledge in regards to a particular word and
2) the VKS is able to measure depth of word knowledge.
To start with, it does not seem reasonable to speculate about depth of the knowledge at the initial stages of its development. Also, the initial knowledge of a word might not be limited to the word's meaning knowledge, the VKS focuses on. The acquisition of foreign language words might begin with learning their written or sound forms or inflectional or derivation forms and so forth. Thus, in order to be able to measure qualitative gains in vocabulary knowledge the VKS needs facilities to assess knowledge of various constituents of word knowledge. However, in its current state, the VKS is unable to provide detailed information on how much knowledge about the word has been acquired. Paribakht and Wesche's scale, however, can be used to measure the growth of breadth (once the other problematic issues discussed below have been solved) rather than depth.

Conclusion. It does not appear to be possible to measure depth of word knowledge (in the sense determined by Nation, 1990; 2001 and agreed by other scholars in the field, see section 1.4.3 of this thesis) by means of the VKS, which is contrary to Paribakht and Wesche's claim. In order to assess the depth of knowledge of a certain word, valid measurements of various constituents of that knowledge would be required.

2.2.3 Issue 3

The self-report categories designed by Paribakht and Wesche.

This section discusses unreliability of the VKS categories. It is difficult to differentiate between some of these categories and, more importantly, test their validity.

The proposed self-report categories do not seem to reflect learners' actual levels of word knowledge. Moreover, learners, as acknowledged by Paribakht and Wesche (1993), found some categories confusing and misleading. Let's take a closer look at the VKS self-report categories again.

Let's start with Category 2, for instance: "I have seen this word before, but I don't know what it means." The question arises: How valid are the responses in this category? There are no means of verifying learners' judgments about their knowledge of the target words in this category. Generally speaking, even the definition (the wording) of this category appears to be problematic. The question arises whether it is really important to establish the fact that the learner has seen the word before if they are not required to provide any kind of knowledge regarding this word. The description of learners' knowledge in this category "The word is familiar but its meaning is not known" is also problematic, since none of the components of word familiarity are even tested. Furthermore, there does not appear to be any essential difference between Category 2 and Category1: "I don't remember having seen this word before" with the conclusion: "The word is not familiar at all". In both cases, the word is technically unknown. In this respect, we need to emphasise that although it is a good idea to record a vague sense of familiarity with a word, more distinctive categories as well as clear instructions for learners are required.

Another example of problematic self-report categories in the VKS is Category No 3 opposed to Category No 4. Category 3 states: "I have seen this word before, and I think it means..." Category 4 declares: "I know this word. It means..." This is another attempt by Paribakht and Wesche to elicit different degrees of certainty in subjects' knowledge of a certain word. However, according to Paribakht and Wesche's report (1993), their subjects struggled with those categories. The authors note that on many occasions, learners could not choose between categories 3 and 4 and were unable to specify their degree of knowledge of a given word. The problem here is the fact that knowledge and certainty are two separate dimensions,

whereas Paribakht and Wesche treat them as a single continuous one.

Paribakht and Wesche also report difficulties in carrying out a statistical analysis of the data provided in adjacent self-report categories on the VKS which forced the authors to combine the data collected into two categories only: "known" vs. "unknown".

In addition, Paribakht and Wesche do not seem to have carried out qualitative analysis of the data on different degrees of learners' confidence in their knowledge of words. This raises a question regarding Paribakht and Wesche's idea to include the certainty dimension in their Vocabulary Knowledge Scale, in the first place.

Conclusion. Paribakht and Wesche's format of the self-report categories within the Vocabulary Knowledge Scale consists of a series of similar categories which do not reflect learners' actual knowledge of a word. The designed categories are confusing for learners and misleading for researchers in data analysis.

2.2.4 Issue 4

The problematic scoring system.

This section shows that the application of the VKS scoring system provides misleading results which are difficult to interpret.

This issue is closely related to the previous issue - issue No 3: problematic self-report categories. As seen in the previous section, the self-report categories in Paribakht and Wesche's VKS do not appear to be reliable in assessing learners' word knowledge. The scoring criteria for these categories also suffer from a number of problems. There seems to be a considerable amount of subjectivity in scoring. Let's take a look at the following extracts from the scoring table:

1)1 point is awarded at Level 1for total absence of knowledge. The explanation of meaning for this score states: "The word is not familiar at all". The question is why Paribakht and Wesche award points for "0" knowledge. Logically, if the word is totally unfamiliar to the learner, they should be awarded 0 points for not knowing this word at all.

2) 2 points are awarded at Level 2 for the claim: "I have seen this word before but I do not know what it means". Two points in category 2 implies that this knowledge is twice "the zero" knowledge recorded in category 1? However, besides the knowledge of meaning senses, the learner is not required to self-assess or demonstrate any other type of knowledge of the word. Thus, at Level 2, the learner does not provide any information of the word. Hence, awarding 2 points just for a learner's claim that they have seen this word before, does not seem to be justified. Clearly, the more appropriate score for the both levels (Level 1 and Level 2) would be 0 points, since the knowledge, if any, is not self-/assessed. However, it should be noted that a learner might have been able to provide some information in regards of the word claimed as seen before (for instance, its sound or written form, word class or grammar forms, etc.). That information (if correct) would have been worth some points. Unfortunately, that was not requested within Category 2 or any other category of the scale.

3) 2 points are also awarded in self-report categories 3 and 4 for a provided wrong answer. And, again, it seems unwise to reward learners for wrong information provided in their self-assessment test. Furthermore, in general, unjustified (discussed under issue 3) creation of similar self-report categories causes confusion within the scoring system applied to these categories. In category 3, a learner guesses that they have seen the word before, and might know what this word means. In category 4, a learner claims that they are definitely familiar with the word and able to provide a synonym or translation for this word. In both categories, however, the right answer receives 3 points, while a wrong response scores 2. The question is why Paribakht and Wesche do not differentiate between degrees of certainty in terms of scoring points awarded.

Overall, there does not appear to be a considerable gap in terms of scoring points between "entirely unknown" awarded 1 point and "well known" awarded 5 points.

It has also to be noted that the scoring scale in Paribakht and Wesche's study published in 1996 suffers from a breach of logical sequence. In that scale, there is an obvious contradiction between the description of Category 4 and scores awarded in that category. A general serious issue is that the scores are not properly defined over a given set of words. For instance, in case of 20 words tested and the score of 10 points received, it is not clear what the scores mean since any combination of points that adds up to 20 is possible.

Conclusion. The scoring criteria for Paribakht and Wesche's VKS do not appear to have been logically thought through. Illustrations as above suggest that the scoring system suffers from serious defects and, therefore, requires serious attention in terms of its structure and scoring criteria.

2.2.5 Issue 5

Confusing and misleading instructions for the VKS.

On numerous occasions as reported by Paribakht and Wesche, subjects admitted that they were not able or struggled to understand the instructions. That might have had a negative impact on the overall data provided.

And again, this issue is closely related to issue No 3: unreliable self-report categories. As reported by researchers who employed the VKS in their studies, learners struggled to make their choice between the VKS categories. Thus, the vague instructions for categories 3: "I think I know" and 4: "I know", for instance, were reported to cause subjects hesitation and uncertainty. Other examples of ambiguous instructions include category 2: "I have seen this word before but I don't know what it means" as well as Category 1 which states: "I don't remember having seen this word before". Surely a simple straightforward instruction for category 1, such as, for instance: I am not familiar with this word, would not cause any doubt. Similarly, the instruction for category 2: "I have seen this word before but I don't know what it means" could be changed as follows: "The word looks familiar. Although I am unable to explain its meaning, I can provide some information about this word".

In many instances, as acknowledged by Paribakht and Wesche, subjects could not choose between the VKS categories and indicate their degree of knowledge of the target word. The existence of such cases suggests that there are serious weaknesses in the instructions to the self-report categories.

Conclusion. Generally speaking, the design of the overall scale needs to be reassessed in terms of developing self-report categories which would be able to test more aspects of word knowledge, with clear instructions.

2.2.6 Issue 6

Ambiguity and inconsistency in using notions and terminology.

This section argues that a number of notions in Paribakht and Wesche's framework of the VKS appear to be totally mixed up. This confusion makes it difficult to judge the results reported by the authors.

The notions used by Paribakht and Wesche: "Levels and scoring levels", "Known and unknown categories", "Unknown and recognition levels of knowledge" seem to cause confusion in the framework of their approach. Thus in Paribakht and Wesche's study published in 1993, the authors state that they developed a five-level description Vocabulary Knowledge Scale (p.15). However, on the next page they claim that the adjustments, they made, resulted in a 6-level scale. Further information indicates that Paribakht and Wesche might have meant the creation of 6 scoring levels. However, if that was the case, it is not clear what each of the levels indicates. Paribakht and Wesche report that they split category 5 into two levels, to reflect both semantic and grammatical abilities to use a target word in a sentence. In this case, category 1 choice seems to correspond to level 1: "The word is not familiar at all"; category 2 indicates level 2: "The word is familiar but its meaning is not known". However, it is not clear which level is indicated by category 3. As stated in the framework of the VKS, a response in category 3 could be scored either as category 2 or category 4: "A correct synonym/translation is given". This suggests that a category 3 response can be regarded as indicator of either level 2 or level 4. Assuming that category 4 signifies level 4 and category 5 is split into level 5 (semantic knowledge of the use of the target word in a sentence) and level 6 (semantic knowledge and grammatical exactness), it is apparent that level 3 is missing.

More confusion is caused by the way Paribakht and Wesche use the notions "Known and unknown categories". The researchers note (1993, p.17) that categories 1, 2 and 3 were grouped into a "not known" category, while categories 4, 5 and 6 were combined into a "known" category. However, as we remember, there are only 5 self-report categories on the scale (1993, p.15). Did the authors mean "levels" or "scoring categories" rather than "descriptive categories"? No explanation was provided. Furthermore, category 3 can hardly be placed in the "unknown category" group since a correct response in this category is also possible. Likewise, "unknown and recognition levels of knowledge" appear to be used

interchangeably by the authors in respect of categories 2 and 3 of the Vocabulary Knowledge Scale. Similarly, categories 1, 2 and 3 of the Vocabulary Knowledge Scale are described as levels of "relative familiarity" with words for which the correct meaning is not known (1993; p.17). It is not clear what Paribakht and Wesche imply under "relative familiarity" since category 1, as we remember, states "The word is not familiar at all."

Conclusion. Paribakht and Wesche should have been more cautious about the terminology they used. Ambiguity and inconsistency in using notions and terminology make it difficult to understand the authors' approach.

2.2.7 General conclusion

The issues discussed in this section indicate that Paribakht and Wesche's VKS methodology does not appear to be as advanced as it has been claimed to be by its authors. The main problem is that the VKS does not seem to provide sufficient and accurate information on how well a word is known (i.e. the depth of knowledge of a particular word). It also suffers from a number of technical problems highlighted in this section. However, having noted that, it should be acknowledged that the VKS is the first serious attempt in L2 vocabulary assessment, a pioneering initiative, to design a measurement instrument which would allow learners to self-assess the quality of their own lexical knowledge. Generally speaking, the VKS can be regarded as another methodology for evaluating breadth of vocabulary knowledge rather than depth.

The Vocabulary Knowledge Scale was employed by some other researchers as an evaluating instrument in their own studies. We are now interested to explore why other scholars adopted the VKS in their research, and whether the VKS methodology has been developed over time. In the next section, we will review a number of other highly cited studies that used the VKS approach to evaluate L2 vocabulary knowledge. We will investigate whether any alterations to the Vocabulary Knowledge Scale improved the original methodology suggested by Paribakht and Wesche.

2.3 Reviews of some other studies that employed Paribakht and Wesche's VKS as a self-assessment instrument

This section explores how other researchers use Paribakht and Wesche's Vocabulary Knowledge scale in their studies. We are interested to know whether and how they modify the original scale to adapt it to the purposes of their research. We are curious to see whether the VKS works in those studies. It is essential to note that although Paribakht and Wesche's studies were referred to and cited in many research papers, there were only a small number of studies which employed the VKS as a measurement tool. In this section, we are reviewing four different papers by Wolter (2001), Rott (2005), Folse (2006), and Zareva (2007).

2.3.1Wolter, B. 2001

2.3.1.1 Summary

This study explores the possibility of L1 and L2 lexicons being structurally similar. The author suggests that depth of individual word knowledge might determine a degree to which a certain word is integrated into the mental lexicon.

The paper begins by reviewing previous research on the structure of the L2 mental lexicon. Wolter states that a large majority of studies in the field support the view that the L2 mental lexicon is fundamentally different from that of the L1. He argues against this viewpoint by claiming that the data obtained through word association tests for native speakers (which backs up this point of view) was limited to fairly common prompt words. He carries on trying to prove his point by citing the results obtained by Postman (1970) and Stolz and Tiffany (1972). These studies found out that when responding to low-frequency (and presumably less familiar) prompt words native speakers produce a large number of non-native responses including clang associates.

Summarizing findings in the field, Wolter argues that there is generally strong evidence for a structurally similar L1 and L 2 mental lexicon:

1. Both native speakers of English and L2 learners demonstrate syntagmatic-paradigmatic shifts in responses.
2. They both produce clang responses, mediated responses, and completely unrelated (to the

prompt words) responses.
3. A large diversity of responses can be found in the data of word association tests collected for native and non-native speakers.

The purpose of this study was to devise and test a model of the L2 mental lexicon based on word associations of native and non-native speakers. The hypotheses of the study were as follows:
1. The L2 mental lexicon of a non-native speaker is structurally similar to the L1 mental lexicon of a native speaker.
2. Depth of word knowledge is a key component for determining the degree of integration for the individual words in both the L1 and the L2 mental lexicon.

Wolter continues by asking a question regarding the factors of word arrangement in both lexicons. Among these factors, he names word frequency, language proficiency and the so called underlying factor which is depth of individual word knowledge. Wolter presents his own model for the last factor: a depth of individual word knowledge model. He argues that this model (DIWK) views the word connections as conditioned by how well particular words are known to a particular learner. Describing the model, Wolter sets up a number of assumptions:
1) The mental lexicon of all learners is unstable: new words are acquired, some previously known words are lost and generally, words are known to different degrees at any time.
2) The mental lexicon is not a fixed structure. It is no more than "the sum of its parts" (individual words).
3) Words in the mind do not all possess the same status.

Developing these ideas, Wolter points out that words in the mental lexicon might be acquired individually and undergo developmental shifts separately from other words in the lexicon. In this sense, he suggests that one should view the mental lexicon as "consisting of a core vocabulary containing well-known words and several layers of peripheral vocabulary consisting of words that are known to varying degrees..." (p.47) Furthermore, he argues that the proximity to the core vocabulary determines how well the particular word is known. This leads him to a further assumption that paradigmatic associations would be predominantly formed between words in the centre circles, syntagmatic responses -- between words slightly

further out and phonological associations -- between words on the periphery. Therefore, the circles or layers of DIWK are as follows (moving from the centre): well-known words; fairly well-known words, moderately well-known words, slightly known words; unknown words. These issues were investigated in the course of Wolters' experimental study reported in this paper.

The subjects for this study consisted of 13 Japanese speakers of English as a second language and 9 native speakers of English. The group of NNS participants included University students and ESL teachers. The native speakers served as a control group and also provided data for analysis. The methodology of the study was as follows. The aural-oral method of collecting data was administered within a word association test. Two lists of prompt words selected from word frequency data (from the Bank of English corpus) were used in this study, one for both groups of participants (48 words) and the other for the native speakers only (48 additional words of low frequency). All responses on the word association test were classified as "paradigmatic", "syntagmatic" or "clang-other" responses or "no response". The scoring procedure for responses was based on the assumption that a paradigmatic response was superior (an indicator of a higher degree of lexical or cognitive development) to a syntagmatic response, which, in turn, was regarded to be superior to a clang or nonsensical response. Hence, a paradigmatic response received a score of 3, a syntagmatic response a score of 2, clang-other responses were awarded a score of 1. If the subject failed to respond they were given a score of 0.

A depth of individual word knowledge test in the form of Paribakht and Wesche's VKS was used to establish how well each of the participants knew each of the prompt words. The underlying hypothesis was that DIWK scores would correspond with "a particular word's integration into the mental lexicon" (p.53). Furthermore, Wolter assumes that this, in turn, would account for patterns of response type and the connections between the words in the mental lexicon indicated by the response types. Paribakht and Wesche's VKS with a scoring scale of 1 --5 was selected to assess the DIWK. It was slightly modified by Wolter in terms of replacing *seen* with *heard* since the test was applied orally. Wolter claims that the maximum score of 5 would indicate that the word belongs to the core of the mental lexicon, a score of 4 would signify that the word is fairly known, a score of 3 would show that the word is moderately well known, a score of 2 would demonstrate that the word is only slightly known

and a score of 1 would represent an unknown word outside the mental lexicon. In Wolter's study, the VKS was applied immediately after the association test was completed. Each participant received a copy of the scale, and had to rate each of the words on the word association test according to the scale. The words were presented in random order using the oral-aural method: PWL 1 for the NNS group of participants and PWL 2 for the NS group.

The first hypothesis was that the two groups (non-native speakers and native speakers) would not show a significant difference in patterns of responses within each category of the VKS. To test this hypothesis, patterns of responses between groups for each VKS category were compared and analyzed. The second hypothesis of the study was that depth of word knowledge is a key factor in determining the structure of both L1 and L2 mental lexicons. In order to check the second hypothesis, the mean proportion of response types (paradigmatic, syntagmatic, clang-other and no response) was assessed in relation to the five VKS scoring categories. The key assumption was that "... VKS scores would have a significant effect on how participants in both groups tended to respond" on the association test (p.56).

However, contrary to expectations, the results for NNS and NS groups appeared to be different. Despite the fact that both groups showed a tendency to respond in accordance with the VKS scoring patterns ($p<.001$), Wolter admits some variation between the two groups, particularly in respect of the well-known prompt words. The native speaker group produced a significantly greater proportion of paradigmatic as well as syntagmatic responses than the non-native speaker group. All in all, he acknowledges that the two groups did not supply enough data to support the hypothesis that the L1 and L2 mental lexicons are structurally similar, especially with regards to well known words. Furthermore, Wolter reports three general patterns in distribution of responses within each VKS category:

1) the similarity in patterns of responses in regards to the words which were not well known (for example, VKS categories 1 and 2);

2) the high proportion of clang-other responses produced by the NNS group for moderately known words (VKS category No 3) which was not the case within the NS group. This led the author to the conclusion that phonology plays an important part in structuring the NNS mental lexicon for moderately known words.

3) the proportion of paradigmatic to syntagmatic responses in VKS category No 5 was described as "a mirror effect" (a reversal of each other) for the two groups. The native

speakers provided 48.9% paradigmatic and 39.8% syntagmatic responses, while the non-native speakers showed the preference for syntagmatic responses 54.1% over paradigmatic responses 35.4%.

Thus, in conclusion, Wolter states that the syntagmatically dominated L2 mental lexicon can be regarded as an underdeveloped form of the L1 mental lexicon. The phonological connections between words in the L2 mental lexicon have priority over semantic connections for moderately well-known words. However, as more knowledge of the word is gained, the syntagmatic connections start to dominate.

Wolter's overall conclusion is that L2 mental lexicon is structurally different but not necessary functionally inferior to the L1 mental lexicon. He points out that L2 mental lexicon is not randomly structured, but there are significant differences between these lexicons.

2.3.1.2 Discussion

In this section, the issue of appropriateness of administering Paribakht and Wesche's VKS for measuring the depth of individual word knowledge is raised. Arguing against this, we consider two points: 1) aspects of word knowledge tested in the study and 2) the scoring system administered.

1) As stated in the summary section, Wolter argues that the depth of individual word knowledge (DIWK) determines a degree of integration of a particular word into the mental lexicon. Hence, according to Wolter, DIWK indicates the location of an individual word in the mental lexicon. Thus, DIWK is central for Wolter's theoretical framework, and the way it is measured might account for the credibility of the whole concept tested. How was DIWK measured in this study? It was measured by means of Paribakht and Wesche's VKS only. Let's briefly recap the procedure. The results of the VKS testing were compared against those obtained from the word association test in order to check the model of the mental lexicon suggested by Wolter in this study. In order to investigate Wolter's assumptions, the results of the association test were compared against the results of the VKS. The main emphasis was placed on the VKS scoring categories 3: "A correct synonym or translation is given" and 5: "The word is used with semantic appropriateness and grammatical accuracy in a sentence".

Thus, in measuring depth of individual word knowledge, Category 3 was meant to indicate that the word was moderately known, while category 5 was sought to represent a well-known word in the core of the mental lexicon. With regards to the results of the association test, a paradigmatic response was supposed to indicate that the word is known to the highest degree, that is, according to Wolter, located in the core of the lexicon. Likewise, a syntagmatic response was thought to indicate that the word is located somewhere in the middle layers of the mental lexicon. Hence, in Wolter's framework, paradigmatic responses ought to signify that the word belonged to category 5 while, a syntagmatic response could be considered as an indicator that the word belongs to category 3. However, as seen from the summary, that turned out not to be the case: Wolter's assumptions were not supported by the results of his experimental study. However, a possible explanation of this failure might lie in the way DIWK was measured in the first place. As usual, within the VKS, the subjects were asked to indicate how well they knew each of the target words by simply providing a translation or a synonym for the target word and by using the word in the meaning (provided in category 3 or 4) in a sentence.

Analyzing the description of Scoring Category 3, the question arises: "If a learner is able to provide as little information about the target word as one of its meaning senses, would it be enough evidence to conclude that the word is moderately known?" And, furthermore: "Would there be enough evidence to make conclusions regarding the place of this word in the mental lexicon?" Similarly, for Scoring Category 5, the question is: "If a learner is able to use a target word (apparently, in one of its meaning senses only) in a sentence, would it be enough information to claim that the word is well-known? ...and, moreover, to conclude that the word is in the core layers of the mental lexicon?" Thus, generally speaking, in this study, the depth of individual word knowledge (DIWL) was checked by requesting subjects to submit one meaning of a word and demonstrate the ability to use it in a sentence. However, it does not seem reasonable to make any judgments in regards to the learners' DIWL on the basis of this limited data (see our discussion on the concept of depth in section 1.4.3). There is even a possibility that learners can produce sentences without actually knowing what a word means. What's required here is a valid measurement of the various kinds/aspects of individual word knowledge for each of the target words. Our review of studies on depth of L2 vocabulary knowledge (section 1.4.3) suggests that if a wider range of the constituents of individual word knowledge are tested, then more accurate information on the depth of this knowledge would

be obtained.

2) The second point against the appropriateness of administering Paribakht and Wesche's VKS in this study is a certain amount of subjectivity in scoring. Wolter points out that it was vital for the DIWK test to provide a wide range of DIWK scores. That was not achieved in his study, however. Paribakht and Wesche's VKS scores range from 1-5 (not really a wide range), from which 1-2 scores indicate absence of knowledge of a given word. In particular, we argue for inappropriateness of VKS scoring category 2 ("The word is familiar but the meaning is not known" as discussed in my previous reviews) in the scoring system. Furthermore, we totally disagree with the Wolter's claim that a score of two represents a slightly known word. If a learner is unable to provide any evidence of his slightest knowledge or has not even been asked for that evidence (apart from a simple translation) it does not seem right to declare the word to be known (even to a slight degree). Moreover, the whole scoring system seems to indicate that full knowledge of a word implies only a small step forward from "no knowledge".

Generally speaking, since the depth of individual word knowledge was not assessed adequately, the assumptions of the study could not have been checked in full. In other words, had the DIWK been measured in a different (assessing more aspects), more precise way, the outcomes of the study might have been totally different. And, moreover, the results of the association test could well be considered as indicators of the word location in the mental lexicon. In this respect, it seems worthwhile to re-test the hypotheses of the study using different DIWK measuring instruments.

2.3.1.3 Conclusion

It does not seem wise to make any kind of conclusions regarding positions of words in the mental lexicon on the grounds of the results of this study. The VKS (in its original form) which was used to measure the depth of individual word knowledge, failed to provide valid evidence for the hypotheses of the study because:

1) it did not test different kinds of word knowledge and

2) its scoring system appeared to be subjective and inaccurate. It failed to provide a sufficient range of scores in DIWK measurement.

2.3.2 Rott, S. 2005

2.3.2.1 Summary

The study explores why certain vocabulary interventions are more facilitative for word learning than others. The author investigates the effect of word processing strategies employed by second language learners on establishing and strengthening lexical form-meaning connections as well as text-comprehension. Two conditions were created: reading a text with 1) multiple-choice glosses (MCGs) and 2) single-translation glosses (STGs).

The study addresses a number of issues, among them the following:
1. What are the qualitative characteristics of word processing strategies of L2 readers who encounter MCGs as compared to readers who encounter STGs?
2. What is the effect of the gloss condition on the robustness of entries in the mental lexicon?
3. What is the effect of the gloss condition on text comprehension?
The incentive of the current study was to record quality and quantity of word processing strategies as learners established form-meaning connections during subsequent encounters with a new word.

The methodology of the study was as follows. Rott asked 10 English learners of German to read a short story which contained four unknown German words each repeated four times. The input passage was an adapted Chinese tale "Shade for Sale". Besides the four target words, seven further words were glossed to assist with text comprehension. All subjects read the same text and were randomly assigned to one of the two treatment conditions: the multiple-choice gloss condition (MCG) or the single-translation gloss (STG) condition.
In the multiple-choice gloss condition, the text was enhanced with multiple-choice glosses for each of the four target words at their first occurrence which were placed in the margin of the corresponding passage. Seven additional glosses chosen for the most difficult words served as distractors so that the subjects did not focus exclusively on the target words. The target words and the additional glosses were bolded in the text. Each gloss had four options: the correct meaning of the word, two additional meanings (quite close to the correct one), and a "don't know" choice. Subjects were asked to circle the option whose meaning fits best in the context.

In the single-translation gloss condition, learners received the same text, however this time it was enhanced with glossed Language 1 translations rather than multiple-choice glosses, for

each of the four target words at their first occurrence only. Likewise, in this condition, the target words and the seven additional glosses were bolded and placed in the margin of the text. After the reading, in order to ensure that subjects processed the reading text for meaning, they were asked to retell the content of the text in writing. Rott recorded the strategies that students used to handle these words using a think-aloud procedure. All think-aloud protocols were analyzed for subjects' processing behaviour. Later on, in order to assess an immediate vocabulary knowledge gain, two tests were administered to the subjects: a slightly modified version of Paribakht and Wesche's Vocabulary Knowledge Scale followed by a word recognition test (a multiple-choice test which was the exact copy of the multiple-choice glosses in the treatment passage).

The adapted version of the Vocabulary Knowledge Scale was as follows:
-- a) I don't remember having seen this word.
-- b) I have seen this word but I don't know what it means.
-- c) I think it means_____(English translation).
-- d) I know this word. It means_____(English translation).
-- e) I can also use this word in a sentence in German.
In computing gains in subjects' word knowledge, the individual categories of the VKS were added up separately. One point was awarded for a correct answer, 0 points -- for an incorrect response.

The word recognition test was administered immediately after the VKS, and was aimed at further assessing "the recessive word gain". Retrieval clues were provided in the form of multiple choices. Generally speaking, this test might be regarded as a follow-up check for the results provided by the VKS.

The experimental procedure included three phases. At phase 1, participants completed the vocabulary checklist test to ensure that the target words were totally unfamiliar to all the subjects. At phase 2 (a week later), subjects received the treatment text without glosses. They were asked to read the text silently. Then, Rott demonstrated to participants what it meant to think aloud. After that, the main treatment was administered, and subjects were engaged in a think-aloud procedure. That was followed by the retelling of the content of the text in writing, which, in turn, was immediately followed by two vocabulary tests (VKS and WRT).

At phase 3 (four weeks after), the same vocabulary tests were administered unannounced.

The results were stated as follows. The think-aloud protocols were reported to show that learners used only a limited number of strategies. These strategies were combined into two categories: meta-cognitive word processing behavior (the glosses, monitor and verbalization) and semantic elaboration (for example, the use of the context, a synonym or background knowledge). Subjects in the multiple-choice gloss condition were reported to integrate meta-cognitive and semantic-elaborative processing strategies when establishing a connection between the lexical form and its meaning. In contrast, subjects in single-translation gloss condition were identified as users of meta-cognitive strategies in order to process the target word.

The results of the vocabulary tests were reported to demonstrate that subjects in both conditions performed very similarly on the immediate post-test. On the VKS, for the majority of the target words, readers demonstrated receptive word knowledge by translating the target words in English (VKS level **c/d**: 60% in the MCG and 55% in the STG condition). For the VKS level **e --** the ability to use the word in a sentence in German, the results were a little lower (55% in the MCG and 45% in the STG condition).

On the subsequent word recognition test, subjects were reported to reach a ceiling effect of 90% (MCG) and 95% (STG) of correct answers. However, the retention scores appeared to be different. On the VKS, Rott recorded a decrease in target word knowledge over four weeks. Though the subjects in the MCG condition performed slightly better in categories **c/d** and **e** on the VKS than those in the STG condition, their target word knowledge also dropped (categories **c/d:** from 60% to 55%; category **e:** from 55% to 45%). At the same time, the VKS results revealed that category **b** scores increased notably in both conditions. It signified that after four weeks subjects were only able to state that they had seen the word before.

General conclusions were as follows. Both, single-translation and multiple-choice glosses triggered essential learning mechanisms. In both conditions, learners established initial form-meaning connections which were measured immediately after the reading treatments. However, subjects who read the text enhanced with multiple-choice glosses were reported to retain significantly more word knowledge than those in the single-translation gloss condition.

On this basis, Rott argues that multiple-choice glosses triggered additional learning mechanisms that fostered word-retention. In general, the conclusion was that glossed sentences received significantly more attention and proved to be useful in directing readers' attention to key ideas of a text. Furthermore, Rott states that multiple-choice glosses proved to be more effective in eliciting strategies than strengthening the links between form and meaning.

2.3.2.2 Discussion

This part of the review analyses of how Rott altered the original VKS by Paribakht and Wesche for the purposes of her research. As mentioned above, Rott's VKS includes the following five categories:

-- a) I don't remember having seen this word.

-- b) I have seen this word but I don't know what it means.

-- c) I think it means_____(English translation).

-- d) I know this word. It means_____(English translation).

-- e) I can also use this word in a sentence in German.

To compare with the original version by Paribakht and Wesche:

1. I have never seen this word.

2. I have seen this word before, but I do not know what it means.

3. I have seen this word before, and I think it means _____ (synonym or translation).

4. I know this word. It means _____(synonym or translation).

5. I can use this word in a sentence:_____.

As we can see, only minor verbal changes were made to the original VKS (Paribakht and Wesche) by Rott (for example, in Paribakht and Wesche's version, category ***1)*** sounds: "I have not seen this word before" compared to Rott's category a): "I don't remember having seen this word"). It is not clear why Rott altered the wording of the original VKS since the meaning of descriptions for VKS categories do not appear to have changed.

Having noted that, it is important to point to considerable differences in the scoring systems of the two versions. While Paribakht and Weshe use a 5-level scoring scale, Rott suggests only 3 levels to score the results of her research. She adds up all the individual categories of the VKS separately, awarding one point for a correct answer and 0 points for an incorrect one. Consequently, unlike Paribakht and Wesche, Rott established three levels (instead of 5) of word gain from the results provided on the VKS. They were as follows: Level 1 = category **b).** Responses at this level indicated that subjects recognized the word form but not its meaning. Level 2 = category **c)** and category **d**). These categories were combined into Level 2 since subjects' responses on those two categories were alike. Rott confessed that subjects did not make a distinction between their levels of certainty of word knowledge ("I think" and "I know"). This is a very fair comment, emphasised by us in our reviews on Paribakht and Wesche's Vocabulary Knowledge Scale. Once again, category 3: "I think it means…" proved to be pointless as a separate category in the framework of the VKS. At this level, the subject was reported to gain receptive knowledge and could provide an English equivalent.
Level 3 = category **e).** The author reports that at this level, subjects gained additional syntactic knowledge about the target word and were able to use it in a sentence in German. She also emphasises the fact that none of the students chose category **a) I don't remember having seen this word.** Something seems to be missing here. There are five self-report categories in Rott's VKS, and only three levels. As we see, Rott does combine two categories (c and d) into one level (level 2), however there is still no level for self-report category **a).** As we remember, in the original VKS, self-report category ***1: "I have not seen this word before"*** (which corresponds to category **a)** in Rott's VKS) represents Level 1 (out of possible 5) -- **The word is not familiar at all.** It's not clear why Rott does not suggest any level (for instance, Level 0) to embrace self-report category **a).**

Generally speaking, there is no use changing the VKS scoring categories if the list of self-report categories remains the same. In other words, the categories which are not going to be scored should not be retained in the scale. Otherwise, it is difficult to interpret the final results.

Furthermore, as seen from the description, in Rott's scale, category **e)** (Level 3) does not reflect on semantic and grammatical criteria in subjects' word knowledge. In Paribakht and Wesche's VKS the score of 4 was awarded in cases when the word was used with semantic

appropriateness in a sentence. The score of 5 was given for the use of the word with semantic appropriateness and grammatical accuracy in a sentence. However, it is understood that in Rott's scale, semantically and grammatically correct use of the word in a sentence was awarded the score of 1. Likewise, if the word was misused according to both criteria the response was scored as 0. However, in case of correct semantic use but failure with grammar, the response was still scored as 0. Within this approach, even fewer aspects of word knowledge seem to have been taken into account by the researcher while measuring levels of gains in learners' vocabulary knowledge.

Moreover, it is unclear whether 1 point was granted to the subject for a choice of category **b:** "I have seen this word but I don't know what it means". If that was the case, it did not appear to be reasonable. Having chosen category **b),** the learner did not demonstrate any knowledge of the word, either receptive, or productive. Rott's claim that category **b)** responses indicated subject's ability to recognize the word form seems to be purely subjective. In addition, clearly, a score of 1 obtained in category **b)** cannot be considered the same value as a score of 1 received in category **d)** or **e),** for instance.

On the other hand, if a choice of category **b)** received 0 points on Rott's scale, it still did not seem to be entirely fair. A refusal or inability to provide any answer (category **b)** and an attempt to submit the answer (but failure) in categories **c/d** and **e** were not rewarded in favour of the latter. This does not appear to be justified either. A learner should always be encouraged to make an effort and recall the maximum information they possess about the word. Furthermore, if category **b)** was not awarded any scoring points, it is not clear why it was labeled as Level 1in subjects' word knowledge achievements. In addition, the scoring system of Rott's VKS suggests that full knowledge (category **e)** is only 1 point forward from total absence of knowledge **(category a).**

Basically, all this confusion within the scoring system does not promote Rott's version of the VKS. Generally speaking, the simplifications made by Rott to the scoring system of the VKS did not improve the accuracy in measuring gains in learners' word knowledge. Quite the contrary, these alterations seem to have brought more confusion and mismatches between self-report categories, levels of knowledge and scoring categories within Rott's VKS.

2.3.2.3 Conclusion

The alterations to the original VKS, mostly in terms of the scoring categories did not add value to the original version of the Vocabulary Knowledge Scale. Rott's version of the VKS seems to suffer from most of the drawbacks inherent to the original VKS, Rott's version of the VKS measures even fewer aspects of learners' word knowledge.

2.3.3 Folse, K., 2006

2.3.3.1 Summary

This study investigates the effect of the type of written exercise on L2 vocabulary retention. Target vocabulary was practiced under three types of written exercise conditions: one fill-in-the-blank exercise, three fill-in-the-blank exercises and one original writing exercise.
A modified version of Paribakht and Wesche's Vocabulary Knowledge scale was used to test the meaning of the target words and usage of the word in student-written sentences.

Subjects were 154 ESL students representing 14 different native languages and functioning at 3 different levels of proficiency: 50 - lower intermediate, 51 - upper intermediate and 53 advanced level students. 15 target words, all verbs, plus 3 distractors (18 in total) were divided into 3 equal groups. The distractors were never right answers and were added to the list of target words to reduce guessing. Subjects practiced each word group under one of the three exercise conditions. Folse emphasizes the fact that all of the target words were supposed to be unknown to the participants. A pilot test with 11 advanced ESL students was conducted to ensure that the words would be unknown by the actual subjects. A mini-dictionary (4 pages) was created for this study to provide input for the meaning of each of the 18 words.

Three written exercise conditions were administered in the study. In Condition 1, subjects were required to fill in the blanks in 5 provided sentences with one of the six target words shown in a table. Each word was used once. Under Condition 2, the words were practiced in three different exercises. However, each of those exercises was similar to the exercise in Condition 1. In Condition 3, subjects were asked to write a sentence with each of the target words.

The pre-test and post-test vocabulary knowledge was assessed by means of a modified version of the vocabulary knowledge scale originally suggested by Paribakht and Wesche. Unlike the original scale, the modified version of the VKS included 3 levels of word knowledge (or self-report categories). The scoring system graded from 0 to 2. One point was awarded for providing a correct meaning. Another point was granted for producing a correct sentence with the target word.

Modified Vocabulary Knowledge Scale:

1. I don't know what this word means.
2. I know this word. It means_____
(provide an English synonym or a translation in your native language).
3. I can use this word in a good example sentence. Write your sentence here.
(If you do No3, you must do No2 also.)

For the pre-test, participants completed the VKS for 24 words. This included 15 target words, 3 distractors and 6 easy words. After the pre-test was completed, "a filler activity" (an association test) was used in order to distract subjects and decrease the chance of remembering the words from the pre-test. The following day, subjects completed another word association task before they started on the actual exercise treatment. Participants received a mini-dictionary and a task booklet with three different exercise conditions. After completing the exercise treatment (40 minutes), students were given 30 minutes to complete the post-test which was exactly the same as the pre-test. Participants were awarded a score of 0, 1, or 2 for each target word. At level 2 of the VKS, participants were asked to provide an English synonym or a L1 translation. One point was given for the correct answer. The third level requested a written example sentence with the target word. Two points were awarded for a good example. It was explained to students that if they chose option 3 they also had to complete option 2. Strict and Lenient scoring were applied to interpret the data obtained. Folse points out that incidental learning (which was the matter of investigation in his study) cannot be full or deep learning. Thus, any learning that took place under the circumstances was not expected to be comprehensive or deeply processed learning. Taking this fact into account, lenient interpretation was used to reveal smaller increments of learning. For instance, Folse reports a case when a student translated the word *toil* as *work* instead of *work hard* or *work with great effort.* Emphasising the fact that learning often occurs in increments, Folse

awarded this response a 1 in Lenient scoring (0 in Strict scoring). Follow-up interviews were conducted to clarify any questions raised by subjects' responses in the VKS.

A repeated measures analysis of variance (ANOVA) was conducted on the data with regards to the effect of exercise conditions. The results are shown in table 2-4.

Table 2-4: Descriptive statistics for retention (Lenient Scoring) by exercise type

Condition	n	Mean	Standard Deviation
1 (one completion)	154	2.18	2.36
2 (three completions)	154	4.78	2.78
3 (original sentences)	154	2.39	2.48

The main effect of exercise type was reported statistically significant, $F(2,306) = 87.01$, $p < .0001$.

The overall conclusion of the study was: vocabulary exercises which required multiple encounters with- or multiple retrievals of the target words (for example, multiple sentence-completion exercises) proved to be a strong factor in L2 vocabulary learning regardless of the type(s) of exercises involved. The mean score for subjects who completed three fill-in-the-blank exercises was 4.78, compared to 2.39 for those who wrote sentences and 2.18 for subjects who did one completion exercise. This gives Folse grounds to argue that multiple encounters using fill-in-the blank activities appeared to be a more effective and efficient L2 vocabulary learning task than writing original sentences.

2.3.3.2 Discussion

This section discusses the measuring instrument, a modified version of the VKS, used in the study to measure breadth and depth of word knowledge. Two points are highlighted: 1) self-report categories or levels and 2) the scoring system of Folse's version of the VKS.

1) The original VKS by Paribakht and Wesche contains 5 self-report categories (see Table 2-1, p. 31). Let us compare it against Folse's 3-level version of the VKS (Folse uses word "levels" for "self-report categories") shown in table 2-5.

Table 2-5: Modified vocabulary knowledge scale by Folse:

Level	Level Description	Scores
1	I do not know what this word means.	0
2	I know this word (the correct synonym/translation is given)	1
3	I can use this word in a sentence (correctly used in a sentence)	2

As we can see Folse removed self-report categories 2 and 3 from the original scale: the categories that indicated learners' certainty in their word knowledge. This implies that the author does not consider that a learner's vague sense of familiarity with a word is worth recording. This simplification, however, does not make the VKS any more effective in measuring learners' knowledge of the words being acquired. That is to say, these alterations do not enable the VKS to measure the depth of subjects' knowledge of the target words which is contrary to Folse's claim. The author argues that the VKS was used in his study to measure breadth and depth of word knowledge (p.289) before and after the treatment. Folse argues that the suggested version of VKS was able to detect even small gains in learners' word knowledge. However, the paper does not provide any results of the VKS testing.

Judging on the levels descriptions, we can note that Folse's VKS superficially tests only two constituents of word knowledge: meaning and ability to collocate (in its simplest form). In other words, general conclusions regarding subject's word knowledge are drawn from the fact whether they can remember a word meaning and whether they can use it in a simple sentence. It is understandable that in the case of incidental learning students might not instantly acquire such aspects of word knowledge as: grammar characteristics or ability to derivate.

Nevertheless, some other types of word knowledge, for instance, the word form or knowledge of more than one meaning (where applicable) could have been included in subjects' self-assessment reports within the VKS. That would have created extra levels within the modified version of the VKS suggested by Folse and led to more accurate testing of gains in subjects' word knowledge.

Let's take a look, for instance, at perspectives of developing and measuring the multi-meaning aspect of word knowledge in Folse's study. Describing the mini-dictionaries created for this study, Folse pointed out that they contained pertinent information regarding the

meaning of each of the 18 target words. The definition was followed by two sentences illustrating the target word. The entry for the verb *bolster* was given as an example:

bolster (verb) To support; to make something feel strong again.

Example Sentence 1. When Joe was in the hospital, I sent a nice card to *bolster* his spirits.

Example Sentence 2. The players practiced very hard to *bolster* their chance of winning.

Immediately questions arise: If a target word has more than one meaning, are all of them shown in the mini-dictionaries used in the research? Are illustrative sentences on each meaning provided? Unfortunately, Folse does not present any examples of entries for multiple-meaning words from his mini-dictionaries. Neither does he provide any information on these types of words listed in his mini-dictionary. What was, for instance, the entry or entries for the word to *burst* (one of the 15 target words)? According to Collins dictionary and thesaurus (2000), the verb to *burst* possesses the following 4 meanings:

1. To fly asunder
2. To break into pieces
3. To rend
4. To break suddenly into some expression of feeling.

Does Folse's mini-dictionary contain all these four entries or does it provide only one definition chosen by the author? If there is only one definition, on what grounds has it been selected?

Similarly, it is far from clear what entries were provided for the word to *ponder:*

1. To muse, meditate. 2. To consider, deliberate (on)?

Or to *launch:* 1. To set afloat. 2. To set in motion. 3. To begin. 4. To hurl, send?

Generally speaking, the multi-meaning aspect of word knowledge was not discussed in the paper or reflected in the levels of his Vocabulary Knowledge scale. Also, the question of collocations arises: Do the examples of sentences shown in Folse's mini-dictionaries (only two for each entry) cover all principal collocations with the introduced word (for example: *to launch a spacecraft; to launch a course or to launch into an argument)?*

This information is crucial for establishing associative links between the new words being acquired and those stored in the mental lexicon. Had that been reflected in the VKS levels, it would certainly have improved the original version of the scale.

2.Similarly, the scoring system of Folse's VKS suffers from a number of drawbacks. Let's compare this scoring scale illustrated in table 2-6 with the original one by Paribakht and Wesche presented in table 2-1 (p. 31).

As we can see the scoring systems of these two versions of VKS differ considerably. On the positive side, it should be noted that Folse abolished a number of totally unjustified scores from Paribakht and Wesche's original version. They are:

a) 1 point -- "The word is not familiar at all" in category 1
b) 2 points – "The word is familiar but its meaning is not known" in category 2
c) 2 points – "A wrong answer is given" in category 3
d) 2 points – "A wrong answer is given" in category 4.

Thus, in Folse's version of VKS, points are awarded for demonstrated rather than perceived knowledge. On the one hand, this seems to make the scoring system fairer since the actual knowledge only is rewarded. However, on the other hand, the VKS scoring remains quite ambiguous in terms of learners' final scores. Thus, for instance, if a learner scores 18 points on an 18 item test suggested by Folse, this score might be the sum of different combinations. This could be 18 words – one point, or 9 words – two points, or various combinations of ones, twos and zeros that add up to 18. It makes it impossible to interpret the summary data. Furthermore, the alterations and simplifications made by Folse did not improve the scale's ability to measure the depth of learners' word knowledge (contrary to Folse's claims). For instance, it failed to reflect on the cases when/if the participant was able to retrieve more than one meaning or produce a number of collocations with the target word. An alternative scoring scale could have been used to detect these larger gains in vocabulary knowledge, such as:

-- I don't know what this word means -- 0 points
-- I know this word. It means.... -- 1 point for each meaning provided.
-- I can use this word in a sentence -- 2 points for each sentence created.

Furthermore, additional 0.5 points might have been awarded for each correctly produced word form. Since the knowledge of the word form is one of the constituents of the word knowledge and could be acquired incidentally, it was worth testing. That might have provided more precise information on the gains in subjects' word knowledge. However, this would need further elaborating and discussing in terms of presenting the summarized results.

2.3.3.3 Conclusion

The modified version of the VKS created by Folse to assess knowledge of the words acquired incidentally, does not appear to have improved the original VKS. In particular, more testing levels (to measure the depth of learners' word knowledge) as well as a better elaborated scoring system are required. Overall, it is essential to provide students with comprehensive (on various aspects of knowledge of a given word) information on each target word during a treatment period, for instance, well-constructed mini-dictionaries. It is also equally important to comprehensively measure the word knowledge acquired.

2.3.4 Zareva, A. 2007

2.3.4.1 Summary

The study analyses quantitative and qualitative characteristics of the patterns developed by L2 learners to structure their lexical knowledge. The word association test was administered to investigate how L2 learners organize their lexical knowledge.

First, Zareva reviews lexical research on the use of word association tests (WA). She focuses on distinctions L1 researchers make between qualitative and quantitative features of WA domains. She points out that those distinctions have been overlooked in L2 research. Then, she carries on speculating about the ways WA tests have been administered in L2 research and findings in regard to L2 learners' vocabulary structure. Finally, she reports on her own study on the issues related to the organization of the L2 mental lexicon. Zareva raises a number of questions which guided her research:

1. Where should the difference between NSs' and L2 learners' mental lexical structure be looked for?
2. How does an increase in language proficiency affect the organization of L2 learners' lexical knowledge?
3. Do the quantitative and the qualitative features of the mental lexicon interact?
4. Which set of features is more sensitive to an increase in language proficiency?

5. Do WA tests have potential for use as assessment tools?

The quantitative characteristics of the WA domains of 3 groups of subjects (native speakers (NSs), L2 advanced and L2 intermediate learners) were compared. Associative strength, size of the associative domain and heterogeneity of the response domain were measured. The quantitative features were tested in the following way:
--associative strength was measured by response commonality;
-- size of the associative domain was measured by total number of responses;
-- heterogeneity of the response domain was measured by number of different responses.
The qualitative patterns of WAs: proportion of paradigmatic; proportion of syntagmatic and proportion of phonological associations were examined. The qualitative and the quantitative characteristics of the WA domains of the three groups were compared in order to investigate the strength of the relationship between them. Eighty-seven adults (NSs (n=29) and L2 learners of English (n=58) took part in the research. The L2 subjects were divided into two groups based on their performance on three proficiency tests: an advanced group (n=29) and an intermediate group (n=29).

The procedure was as follows. The participants completed in writing a vocabulary test containing 73 sample words (SWs) selected from a learner's dictionary (Hornby, 1978) by a spaced sampling procedure --"systematic sampling with a random start"(p.132). Each SW was accompanied by a modified version of the VKS (Paribakht and Wesche, 1993), referred to in the study as "a word familiarity scale" (p.133). Into her version of VKS, Zareva includes four main options which, she argues, are meant to identify four degrees of word familiarity. They are as follows:
1) I have not seen this word before;
2) I have seen this word before but I don't remember what it means;
3) I think this word means_____ (provide a synonym or brief explanation);
4) I know that this word means_____ (provide a synonym or brief explanation).
The fifth option -- I associate this word with _____ was added by the researcher to collect WA data.

The subjects were required to produce three associations if they stated their familiarity with the SW in option 3 or 4. They were asked to respond to every SW according to its lexical

category. The target words were presented in alphabetical order with their lexical categories stated next to them. A total of 7854 WAs were collected during this study.

All the associations generated by the subjects were combined into a total list (one for each group, three in total). An association was included in the list if the subject responded correctly to the lexical class of the target word, along with, at least, one of its meanings in option 3 or 4 of the word familiarity scale. For example, for the target word *hard,* which was specified as an adverb, synonyms like *strenuously, laboriously, persistently* provided by participants in option 3 or 4 and all associations (generated in option 5) following such synonyms were included in the analysis.

Zareva provides her list of criteria for scoring quantitative and qualitative characteristics of subjects' WA responses. Two one-way ANOVAs were administered to compare the quantitative and qualitative associative patterns submitted by the three experimental groups. The author reports that overall, the results of the analysis confirmed differences in the quantitative features of NSs', L2 advanced and intermediate learners' mental lexicons. However, she points out that the size, commonality and heterogeneity of the advanced learners (in contrast with the L2 intermediate learners) meaning connections closely resembled "the NSs' quantitative patterns of associative links" (p.140). At the same time, she argues that differences between the NSs' and the L2 learners' mental lexicons can be revealed primarily at an intermediate level of proficiency. She continues by pointing out that smaller vocabularies (she suggests, 6000 words for an intermediate level) are likely to be characterized by "fewer links among words, a lower degree of commonality and lesser heterogeneity of meaning connections" (p. 144). In contrast, she states, larger vocabularies (over 9000 words) have significantly richer connections in respect to size, commonality and heterogeneity.

With regards to the qualitative features of subjects' lexicons, Zareva reports that no phonological associations were found among the responses generated to known words. On these grounds, she concludes that word familiarity is "the factor that motivates the elicitation of more phonologically than semantically or syntactically linked associations" (p.140). The overall conclusion was that the study proved the existence of quantitative but not qualitative differences in the patterns developed by subjects to structure their lexical

knowledge. Zareva reports that the quantitative differences appeared to be most prominent in the intermediate group. As for the qualitative aspect, adult L2 learners like NS indicated a preference for paradigmatic rather than syntagmatic connections for familiar words.

2.3.4.2 Discussion

This section discusses the modified version of the VKS altered and administered by Zareva in her study. Zareva points out that the "word familiarity scale" (as she refers to VKS) was used in her study to establish subjects' levels of familiarity with the target words. In her version of the VKS, Zareva retains the four self-report categories suggested by Paribakht and Wesche. However, the fifth category of the original VKS: "I can use this word in a sentence_____ Write a sentence." (Paribakht and Wesche, 1993) is omitted. At the same time, a new category: "I associate this word with…" was added by the author pursuing the goals of her word association study. Thus, category 5 in Zareva's VKS is a word association task which appears to be compulsory for those subjects who submitted their answers in categories 3 and 4. What follows from here is the fact that the status of Category 5 in Zareva's scale is entirely different from that in the original scale. In other words, Category 5 on Zareva's scale cannot be regarded as a separate higher level of the scale, unlike Category 5 in the original version of the VKS by Paribakht and Wesche. This decision to replace the original Category 5 does not seem to be justified. Had Zareva retained this category within the VKS, it would have been possible to explore how different levels of word familiarity influenced the characteristics of L2 WA domains.

It's clear from the beginning that the main focus of Zareva's study was associative abilities of the L2 mental lexicon. The question is: how the Vocabulary Knowledge Scale (VKS) fits into the frame of this WA research. Zareva notes that the self-report categories of the VKS were used to establish whether the subject was familiar with the words, selected by her for the word association test. As stated in the summary, subjects were instructed to indicate on the word familiarity scale whether they knew each of the given words. Only if the subject submitted their response in option 3 or 4, they proceeded to the next stage: a word association test. Immediately a question arises: how Zareva graded the answers provided in option 3: I think this word means_____ (provide a synonym or brief explanation) compared to option 4: I know that this word means_____ (provide a synonym or brief explanation), provided they

were both correct. In other words, what criteria applied to differentiate the confident answer (option 4) from the uncertain one (option 3)? Similarly, it is not clear what impact option 2 had on the research: I have seen this word before but I don't remember what it means. Bearing in mind that only correct answers in option 3 or 4 allowed students to participate in the main part of the study: the WA test (category 5), the retention of categories 2 and 3 within "the word familiarity scale" seems totally pointless.

Therefore, it is not clear why the author chose VKS to establish subjects' familiarity with the words rather than assigning, for instance, a simple translation test. Self-report categories 2 and 3 on Zareva's VKS do not seem to be included in the data analysis in any way. Moreover, degrees of certainty in the word knowledge (shall I say meaning knowledge) which are meant to be revealed in those categories do not appear to play any part in the framework of Zareva's study.

Also, it is not shown in the research how participants responded within the self-report categories. Furthermore, no scoring system was suggested within Zareva's version of the VKS: subjects' responses provided on the scale (excluding Category 5) do not appear to have been scored in any way. In section "Scoring procedures", Zareva describes norms for compiling a list of word associations generated by the participants in option 5, however, scoring for the first four levels of "the word familiarity scale" is not considered by the author. This must have occurred due to the fact that the researcher focused entirely on learners' correct translations/explanations for the given words and completely ignored other responses submitted in categories 1--3. And again the question of VKS appropriateness for this study arises. Without doubt, it would have been beneficial for the research had the author given more thought as for her choice of the assessment tools as well as consistency in their application throughout the study.

2.3.4.3 Conclusion

Briefly mentioned in the section "Materials and procedure", the VKS appear to be totally unconnected to the rest of the study. Since the information provided by students within the self-report categories was neither scored, nor included in any kind of analysis, it seems unreasonable to employ this instrument for the goals of the research.

2.3.5 Conclusion on the use of VKS by other researchers.

In this section, we investigated how Paribakht and Wesche's VKS had been used by other researchers in the field. We were curious to learn why they employed VKS as a measurement instrument in their studies. We intended to investigate whether and how they developed the original scale to adapt it to the purposes of their research. It has been of particular interest to see whether any alterations to the Vocabulary Knowledge Scale improved the original methodology suggested by Paribakht and Wesche. In order to achieve this, we have reviewed a number of highly cited studies that used the VKS approach to evaluate L2 vocabulary knowledge: four different papers by Wolter (2001), Rott (2005), Folse (2006), and Zareva (2007). The analysis of these studies has shown a common trend which will be discussed below.

As seen from these papers, Paribakht and Wesche's VKS was used by the aforementioned scholars either to establish whether particular words were known by the subjects (Zareva) or to measure how well the words were known (Wolter) including vocabulary knowledge gain in treatments (Rott) and incidental gain (Folse). None of these researchers explain why they decided to employ the VKS in their studies. Their decision might have been influenced by the claims Paribakht and Wesche made in regards of the abilities of their VKS: to measure both breadth and depth of vocabulary knowledge. It is understandable that such a comprehensible measurement instrument as Paribakht and Wesche's VKS was claimed to be, would be of particular interest to researchers in terms of collecting complete and accurate information about L2 learners' knowledge of words. However as seen from our reviews this does not appear to be the case. In none of the studies reviewed, does the VKS seem to be the right choice of the measurement instrument within the framework and goals of those studies (see relevant discussion sections). Moreover, we had the impression that the wrong conclusions may have been drawn (e.g. Wolter's study) due to a wrong choice of the assessment methodology.

Has the VKS been developed over time? The original self-assessment scale was used either unchanged (Wolter), slightly modified (Rott) or significantly altered (Folse and Zareva). Thus there have been some rare attempts to adjust the VKS to the goals and aims of particular studies. However, as argued in our reviews the alterations made to the VKS by these researchers do not appear to have improved the original. The altered scales seem to suffer

from the same problems as Paribakht and Wesche's VKS, i.e. 1) no accurate, substantial information on vocabulary depth, 2) confusion and mismatches within self-report categories and scoring systems, 3) unclear instructions. At the same time, the reviews of these studies show that changes made to the VKS by the aforementioned researchers (mostly in terms of its simplification) led to even more confusion and ambiguity within its levels or self-assessment categories and scoring.

Also, the simplified versions of the VKS appear to provide even less data on how well a certain word is known by a learner since less information on word knowledge gets extracted and analysed (e.g. Folse's scale abandons the dimension of certainty in knowledge whilst Zareva replaces the aspect of use with the aspect of association knowledge and abandons the scoring system within her version of the scale). Thus, the ability of the VKS to measure depth of learners' word knowledge has not been improved as it remains insensitive to many constituents of vocabulary depth (see section 1.4.3 of this thesis).
Our overall conclusion is that Paribakht and Wesche's VKS does not develop over time. The alterations made to the original VKS by other researchers, mostly in terms of simplification of its self-assessment categories and scoring system did not add value to the original version of the Vocabulary Knowledge Scale.

2.4 General conclusions

In this chapter, we have reviewed in detail a number of studies by Paribakht and Wesche, which develop and administer the VKS methodology. We have also reviewed a number of highly cited studies in which adaptations of the VKS have been used to evaluate L2 learners' vocabulary knowledge. All these studies appear to be problematical in one or more respects. The main issue seems to be that the VKS simply does not adequately measure the depth of a subject's word knowledge. Clearly, then, there is a need for further research into self-assessment of vocabulary knowledge, and the next section of this thesis will attempt to do this through a serious of experimental reports.

CHAPTER THREE

Self-Assessment of L2Vocabulary Knowledge

Introduction

General Methodology administered in this research

In the series of studies we run within this thesis, we have adopted the case study research methodology. There are a number of reasons why we have administered this approach.

First of all, due to the nature of our research, we have been looking for a research method which would allow us to investigate the complexity of single cases in full detail. In contrast with large-scale experimental studies, case study is described as a study of particularity and complexity of a single case when a holistic, in-depth investigation is required (Stake, 1995; Yin, 2003; Duff, 2007). Since, in our research, we intended to explore the complexity of learners' individual word categorizations, a case study approach seemed to be ideal for the purposes pursued.

Secondly, we were interested to examine each learner's detailed explanations of the ways they measure and classify their vocabulary knowledge. And again, a case study approach appears to perfectly suit this purpose. As noted by Stake (1995) case studies are designed to bring out the details from the viewpoint of the participants. Gall et al. (2003) describe case study research as "in-depth study of instances of a phenomenon in its natural context and from the perspective of the participants involved in the phenomenon" (p.436).

Thirdly, the case study approach allows the researcher to focus on a particular issue or issues that are fundamental for understanding the system that is being investigated. In our studies, these issues are quantitative and qualitative features of the categorization systems of lexical knowledge of our participants.

The case study methodology allows the experimenter to test the existing hypotheses and, at the same time, it generates new ideas, sometimes counter-evidence to existing viewpoints and claims (Duff, 2007). In our research, we were attempting to find some evidence in support of our critics of the existing self-assessment methodologies which, in our view, provide rather superficial information in regards to the depth of L2 learners' vocabulary knowledge.

Also, a case study approach seems to create an ideal research environment for carrying out a detailed longitudinal examination of a single case. According to Stake (1995) in-depth longitudinal examination of single cases provides an ideal way of systematic observing, analysing and reporting results over a long period. Indeed, we believed it would be more feasible at the final stage of our research to investigate a chain of possible changes in a learner's categorization systems of her L2 lexical knowledge by using a longitudinal design.

In this thesis, we report on six case studies: two - multiple and four – single case studies. A multiple-case design has been used in our study to trace "numerous sources of evidence through replication" (Stake, 1995). In general, multiple-case design is claimed to enhance and support the evidence obtained in these studies which may not be captured through full-scale experimental research (Stake, 1995; Yin, 2003). In this study, we were interested to investigate how different learners measure and classify their own knowledge of the same proposed words. In order to achieve this target we intended to compare and analyse categorization systems of vocabulary knowledge created by different subjects in our first multiple case study (reported in Chapter 3).

Besides these major advantages of the case study methodology there are also a number of other reasons why we have adopted the case study approach in this thesis, rather than the more traditional approach that make use of large subject groups. One of them is purely pragmatic -- it is extremely difficult to get hold of large numbers of learners of Russian in Wales, particularly learners of Russian who are not beginners, and this effectively precludes an experimental approach to the questions we are interested in. The other reasons are: the possibility of investigating research questions on a larger number of words; one-to-one interactions (subject—researcher) in order to better understand subject's responses and no time restrictions.

Experimental Study 1

3.1 Introduction

Having analyzed Paribakht and Wesche's approach to self-assessment of L2 lexical knowledge we established a number of weak points in their methodology which were discussed in the previous chapter. Our main argument relates to the fact that Paribakht and Wesche's VKS cannot be used to measure depth of lexical knowledge (which is one of Paribakht and Wesche's major claims in relation to their scale) since it focuses on testing only two types of learners' word knowledge: meaning and use. We felt that a self-assessment methodology should be able to assess considerably more types or aspects of the learners' word knowledge than the aforementioned two. This is especially important if the target of self-assessment is to measure the depth of learners' lexical knowledge (see section 1.4.3 of this thesis: the concept of depth). We also felt that the best way to explore this issue would be to ask learners themselves to rate their knowledge of words.

The experimental work, reported in the six chapters that follow and the appendix at the end of this thesis, presents our attempt to establish how learners themselves self-assess their L2 vocabulary knowledge. The research is based on the information provided by learners of Russian as a second or foreign language. We chose Russian because it is one of the so-called "unusual" languages: there is an apparent shortage of research on Russian (as a foreign language) language vocabulary acquisition. Our choice is also explained by the fact that the author of this thesis is a native Russian speaker with a large experience of teaching Russian to foreigners.

Our research is very exploratory and unique in terms of engaging learners themselves in investigating how they classify their own knowledge of words. This information might assist in understanding how learners perceive and structure their knowledge of words. Furthermore, we believe that "this look from inside the system" might enable us to better understand how word knowledge is acquired and develops. This has never been explored before. However, we are certain that this information is vital for establishing how lexical self-assessment instruments should be structured.

This chapter presents our first attempt to explore how learners classify their L2 lexical

knowledge. We intended to establish whether learners of Russian would differentiate between various aspects of lexical knowledge or simply focus on their meaning knowledge. Thus we asked the participants to categorize their knowledge of Russian words. The purpose was to explore how self-report categories suggested by our learners while assessing their own knowledge of certain words, would differ from the self-report categories developed by Paribakht and Wesche within their Vocabulary Knowledge Scale.

Bearing in mind the self-assessment categories suggested by Paribakht and Wesche, we posed a number of questions.

Research Questions

1. How many self-report categories will be created by the learners in order to classify their knowledge of the given words?
2. Which criteria/features will be considered by the subjects when they assess their own knowledge of words? Will learners entirely focus on their meaning knowledge (Paribakht and Wesche's viewpoint) or will they look into other aspects of word knowledge?
3. How do different degrees of certainty about the knowledge of a target word (which received a lot of attention in Paribakht and Wesche's scale) manifest themselves in self-assessment categories created by learners?
4. Will categorizations submitted by different learners differ from each other in terms of their quantitative and qualitative features?

These questions guided our current study.

3.2 Study

3.2.1 Method

3.2.1.1 *Target Words*

Fifty Russian words were used in this study. The words were selected from the Word Frequency List (Vakar, 1966; also checked with Агрaев et. al., 1977) and fell under the following 3 levels: 2k, 3k, 4k. Each word was typed onto a single card: 50 cards in total. The target words represented four parts of speech: nouns, verbs, adjectives and adverbs. These words are shown in table 3-1 in the alphabetical order.

Table 3-1: Target Words

Вернуться *To return*	Журнал *Magazine*	Мир *Peace; World*	Послать *To send; to move*	Случай *Case; event; opportunity; chance*
Взглянуть *To look at (perf)*	Завидовать *To envy*	Надеяться *To hope(for); to rely(on); to hope(to*	Последний *Last; the latest; the latter; worst*	Собрание Meeting; assembly; collection
Взять *To take (perf)*	Замуж *Marry(females)*	Находить *To find*	Предлагать *To offer; to propose; to set (put); to order*	Счастье *Happiness; luck*
Врач *Doctor*	Звонить *To ring*	Нужный *Necessary*	Придумать *To think (of); to imagine*	Терять *To lose*
Всё *Everything*	Каждый *Every; everyone*	Объяснить *To explain (perf)*	Принимать *To take; to accept; to admit; to recieve*	Уезжать *To leave*
Дорога *Road; journey; way*	Красивый *Beautiful(fine)*	Повторить *To repeat; to revise*	Прощать *To forgive; to remit*	Улица *Street*
Дорогой *Dear; expensive*	Красный *Red*	Поднимать *To raise; to pick up; to stir up; to improve*	Прямо *Straight, directly; frankly; really*	Умереть *To die*
Душа *Soul; spirit*	Кушать *To eat*	Позволить *To allow*	Разговаривать *To talk (speak)*	Учить *To learn; to teach*
Ждать *To wait; to expect*	Лежать *To lie (to be situated)*	Получить *To receive*	Сердце *Heart*	Честный *Honest*
Жить *To live*	Лицо *Face; exterior; person; identity*	Порядок *Order; succession; manner; customs*	Сидеть *To sit; to fit; to be in prison*	Читать *To read; to give a lecture*

3.2.1.2 Participants

Participants in this study were 12 Russian foreign language students: males and females attending Russian language courses at Cardiff University. Students' proficiency levels ranged from lower intermediate - Year 2 (5 people) to upper intermediate - Year 3 (7 people). The subjects were from a variety of L1 backgrounds: English, Italian, Portuguese, Spanish, Arabic and Polish.

3.2.1.3 Procedure

Participants received 50 cards each, with one target word printed on a card (50 words in total). They were asked to divide this pile of cards into groups according to how well they thought they knew a word on the card. Participants were not advised on which type of criteria they should base their categorizations. The time allocated for the self-assessment task was 30 minutes. Each subject had two desks at their disposal to provide maximum space for their

creativity. No discussion between subjects was allowed.

After all the words on the cards had been divided into categories, subjects received strips of paper of different colour and were required to label each of the created groups. They were asked to explain the criteria they chose for their classifications. At the end, subjects filled in the final form and copied the categories they had created, onto the form. Since some of the labels contained vague explanations we asked a number of students to clarify their labels and explain the categories they had created. That was recorded.

3.2.2 Results

Addressing the research questions formulated in section 1 of our current study, we asked the subjects to categorize the targeted words according to how well they knew them. The basic results are presented in tables 3-2 through 3-3. Table 3-2 shows the total number of self-report categories created by the subjects.

Table 3-2: Number of self-report categories suggested by the subjects.

Year 2 Subjects	No of Categories	Year 3 Subjects	No of Categories
TS	11	OP	7
BG	8	ID	6
JM	6	SM	6
HE	5	AC	6
SL	5	RP	4
		MB	3
		CH-M	3

As seen from the table, most subjects created rather a large number of categories. In fact, the average number of the submitted categories was 6 (5.83) which is more than included in Paribakht and Wesche's scale. Three subjects arranged the targeted words into more than six categories (with the largest number of categories, namely eleven, produced by subject TS). At the same time, another three subjects managed to create just four or three categories.
The summarized analysis of the qualitative characteristics of the submitted categorizations is

illustrated in table 3-3. This table demonstrates how often a particular criterion was mentioned in the subjects' descriptions of the self-report categories they created.

Table 3-3: Distribution of features among the categories submitted by the subjects

Subjects	Meaning		Use		Word Class		Grammar Info		Written Word Form		Sound Word Form		Multiple-meaning		Four Skills		Certainty in Knowledge + Guessing	
	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
TS	6	1	3		4		6	1									6	2
JM	5	1	3		1		1		1	1					4			2
SL	3	1	1				2										2	2
HE		1									2							1
BG	2	2	3												2		2	1
RP	2	1															2	
OP	5	2			6													2
SM	4	2			1	1	1							1				2
AC	2	1			1		1						1				1	
MB	2	1	1				1											2
ID	3	1	1				2										2	2
CH-M	2	1	1												1			1

+indicates that certain features were mentioned in a positive context; - signifies that certain features were mentioned in a negative context.

As seen from the table, subjects referred to a large number of different features when categorizing their knowledge of words. In fact, nine different descriptors were used by the learners to characterize their knowledge of the given words. The main features subjects' classifications are based on seem to be meaning knowledge and confidence in knowledge. The second main block of classification criteria includes grammar knowledge, word class and use in a sentence. In general, all the features listed in table 3-3 seem to fall into two main dimensions: familiarity and certainty in knowledge. Overall, as indicated by the data, the number of features engaged by our subjects to describe their knowledge of words significantly exceeds the number of self-assessment criteria considered by Paribakht and Wesche. The results obtained in this study and their implications will be discussed in the section that follows.

3.3 Discussion

This section discusses the self-report categories suggested by our subjects while self-assessing their knowledge of the given words. First, we will look at the general characteristics of the sets of categories created by our learners. Then, we will compare the categorization systems to detect their differences.

3.3.1 General characteristics of the self-rating categorizations submitted by the subjects

In an attempt to find answers to the questions posed in this study we examined the classification systems created by the learners. First, we were interested to investigate the quantitative features of the created categorizations. As seen from table 3-2, our subjects created a fair amount of categories. The average number of the proposed categories was 6 which is more than suggested by Paribakht and Wesche. Furthermore, seven out of twelve subjects submitted six or more categories, with one learner having classified their targeted word knowledge via eleven categories. This individual variation in regards to the number of the created categories raises a question in relation to an individual approach in lexical self-assessment.

Second, we intended to establish which criteria or features would be chosen as grounds for the subjects' categorizations of the targeted words. The results of this study seem to indicate that learners rate their knowledge of words in different ways. This means that learners use a variety of criteria to describe their knowledge of words. This also means that different learners use different sets of classification criteria. In regards to the former, it should be noted that despite a large variety of features our learners refer to, all descriptors seem to fall within the two dimensional space: 1) aspects of knowing a word and 2) degrees of confidence in knowledge (Table 3-3). Evidently, the most common way of rating was consideration of both dimensions.

Interestingly, most of the descriptions for the created categories referred to more than one feature. For example, in the description: "Can spell these words, use in a sentence, can recognize written and in speech and translate…", at least four different features can be determined. They are: meaning, use, written forms/spelling and language skills. Tables 3-4 through 3-5 illustrate further examples of distribution of features among the categories submitted by different subjects.

Table 3-4: Distribution of features among the categories in the set submitted by subject TS

No of the Category	Features referred to
1	Grammar, word class, certainty
2	Grammar, use
3	Grammar, certainty, use
4	Grammar, certainty (-), use, meaning
5	Word class, meaning, grammar, certainty (-)
6	Meaning, grammar, certainty (-)
7	Meaning, certainty (-), grammar(-)
8	Word class, meaning, certainty (-)
9	--
10	Word class, meaning, certainty (-)
11	Meaning (-)

Note. (-) indicates that the feature was used in a negative context.

Table 3-5: Distribution of features among the categories in the set submitted by subject JM

No of the Category	Features referred to
1	Word class, grammar, meaning, use, lang. skills
2	Meaning, use, lang. skills
3	Spelling, use, meaning, lang. skills
4	Meaning, lang. skills, spelling (-), certainty (-)
5	Meaning, certainty (-)
6	Meaning (-)

Note. (-) indicates that the feature was used in a negative context.

As seen from these tables, descriptions of the created categories are normally based on a combination of different features. Moreover, category 1 in JM's classification system refers to five different features. These multi-featured descriptions of the self-rate categories might be taken as evidence for the existence of some kind of links among different aspects of word knowledge. Furthermore, it might be the case that one feature triggers another or a set of others linked to a particular word. And eventually, the whole range of information that is stored in the mental lexicon in regards to the selected words becomes unveiled. However, this needs further exploring on a larger set of words.

Another interesting finding relates to the use of features in a negative context. The data presented in tables 3-3 through 3-5 indicate that on a large number of occasions, learners

referred to the criteria (in their category descriptions) within which they were unable to provide information on the word. For instance, one of the subjects wrote: "I know what this means but can't put it into different cases without help". As seen from this description, characterizing their knowledge of certain words, the subject considered the grammar aspect as a criterion for the particular categorization, pointing out that they were unable to perform certain grammar operations with the words allocated under that category.

Another interesting point that is worth noting is creation of categories (within the submitted systems) that contain just one or two of the targeted words. This might indicate that learners consider a fair amount of properties for each of the targeted words while measuring their knowledge of those words. Moreover, this might imply that learners examine each word separately. Indeed, our subjects tended to determine certain features in their knowledge of a word rather than grouping words according to pre-selected features.

Overall, the results of this study seem to indicate that most of the learners used a comprehensive (or multi-featured) approach while assessing their knowledge of words. Though meaning knowledge appears to be one of the major criteria for learners' categorizations (as shown in table 3-3), a large number of other constituents of word knowledge (Schmitt, 1998) also received some degree of learners' consideration. Moreover, in a number of cases when meaning senses were indicated as totally unknown, learners still assessed and categorized their knowledge of words on the grounds of some other criteria (e.g. "Words I do not know the meanings of but can say what part of speech they belong to"). This suggests that words might be found in the lexicon by their different properties which, in turn, implies that these properties should be determined and tested within the process of self-assessment of learners' knowledge of a word. This may be achieved by developing a new comprehensive approach to the L2 vocabulary self-assessment which would aid learners in estimating the depth of their word knowledge. In summary, the experimental data obtained in this study suggests that self -assessment of L2 vocabulary knowledge is considerably more complex than the abilities of the Vocabulary Knowledge Scale by Paribakht and Wesche and its current modifications. In other words, self-report categories of a vocabulary knowledge self-assessment instrument should be designed in such a way as to enable learners to reflect on different aspects or features while assessing their knowledge of words.

The third question posed in our study addresses the issue of certainty in knowledge. It is noteworthy that this dimension was used as a major (together with meaning knowledge) classification criterion by our subjects. The learners expressed different degrees of certainty in knowledge of a word or a particular aspect of knowledge of a word, e.g. "Words I think I know one meaning…", "I would guess that these are masculine words but I would not be confident about this." Different degrees of confidence in knowledge are reflected in the "positive" and "negative" columns in table 3-3.

Interestingly, one classification was based purely on the subject's certainty in her meaning knowledge of the target words:

----1. I know; 2. Have a good idea; 3. A pretty good guess; 4. No idea

A number of subjects tended to distinguish between slight degrees in their word knowledge:

--Familiar words vs. Less familiar words

--Confident vs. Fairly confident

This might imply that different degrees of certainty expressed by subjects in their word knowledge descriptions indicate different phases in the development (or regress) of that knowledge. If this is the case, different degrees of certainty about the knowledge of the target word should be reflected in the scoring system in order to obtain a more complete picture of the depth of a learners' word knowledge. However, if a learner claims that he might know the word but fails to demonstrate any actual knowledge, they should not be awarded a score higher than if they admit a lack of knowledge. This is contrary to Paribakht and Wesche's approach.

3.3.2 Differences between the submitted categorizations.

The fourth question of the present study required a comparison between the created classifications. In the previous section, we discussed general (common) characteristics of the submitted categorizations. Here, we will focus on their differences. When twelve categorizations are compared, the results reveal some differences in quantitative and qualitative features of these classifications. In relation to the quantitative differences, we note that the numbers of self-rate categories in the submitted categorizations fall within the range from 11 to 3 (Table 3-2). However, as stated in the results section, half of the submitted classifications contain 5 to 6 categories.

As for the qualitative differences of the created categorizations, they appear to be quite distinctive. The data obtained in this study indicate that different learners use different sets of classification criteria i.e. refer to different features. Though all subjects accept meaning knowledge and certainty in knowledge as criteria for their categorizations, other features do vary.

Let's take a look at tables 3-3 through 3-5. As seen from the tables, subject TS, for example, classifies her word knowledge according to the grammar aspect, subject JM evaluates his abilities within the four language skills, subject SL and subject RP categorize according to their certainty in their knowledge, subject HE refers to phonological information while subject OP categorizes according to word class. The individual variation revealed in our study is a rather important finding since it indicates that self-assessment methodologies should approach the issue of lexical self-assessment on an individual basis.

Summarizing the discussion held in this section, it must be noted that the process of self-categorization of learners' word knowledge appears to be considerably more complex than generally assumed in the field. Though it is hard to make generalizations on the basis of the results obtained from a small group of subjects, the findings of this study seem to have some rather important implications for lexical self-assessment and vocabulary acquisition in general.

3.4 Conclusions

This study was our first attempt in exploring how learners categorize their own knowledge of words. The data indicates that a variety of features were taken into account when learners assessed their word knowledge. This has never been noted in previous research and runs counter to the general assumption in the field.

The findings of this study are encouraging since they substantiate our idea of the complexity of the process of lexical self-assessment. In light of the insights gained in this chapter, it seems reasonable to continue investigating the way/s learners classify their knowledge of words.

Having discovered some connections among different features of word knowledge, we intend to further investigate those links via new case studies on a larger set of selected Russian words. However, first, in light of the fact that meaning knowledge seems to be one of the leading criteria for arranging words into self-rate categories, we are interested to explore in more detail how learners themselves measure their meaning knowledge. We intend to test this aspect of word knowledge in detail in the next chapter.

CHAPTER FOUR

Self-Assessment of Meaning Knowledge of Selected Polysemous Words Experimental Study 2

4.1 Introduction

In the previous chapter, we discussed the findings of our first experimental study into the issue of lexical self-assessment. We found that learners were generally very good at categorizing their knowledge of words. The classification systems submitted in the previous study on average appear to be more complex, in terms of their quantitative characteristics, and comprehensive (i.e. multi-featured) compared to the self-assessment scales developed in the field so far. However, the individual variation between the characteristics of the subjects' categorization systems indicates that further research on a larger set of words will be required. This will be commenced in studies three through six of our research.

The current study, however, focuses on self-assessment of meaning knowledge. Furthermore, we are curious to establish how L2 learners will measure their knowledge of polysemous words. There are two reasons for investigating this matter. First, before we continue to explore the way/s learners self-categorize their knowledge of words, we are interested to establish the reliability of data obtained through lexical self-assessment. Due to the fact that it is literally impossible, within the scope of this thesis, to test the reliability of self-assessment of each of the word knowledge types, we chose meaning knowledge.

Second, as indicated in the previous study, meaning knowledge appears to be the main criterion used by the learners to categorize their knowledge of words. Furthermore, it is relatively easy to validate the results of self-assessment of meaning knowledge by comparing them against the data obtained in the subsequent assessment. In addition, a high percentage of Russian words are polysemous (with meaning senses not often obviously related to each other) which is important to bear in mind when learning or teaching Russian.

Thus the object of the present study is self-assessment of multiple-meaning knowledge. We felt that self-measuring the multiple-meaning knowledge is an important part of evaluating the depth of learners' lexical knowledge (see section 1.4.3 of this thesis). Thus, we were curious to explore how learners themselves perceive their knowledge of multiple-meaning senses.

In this study, we intended to establish whether learners of Russian would be able to correctly measure their multiple-meaning knowledge of certain words.

By addressing this issue, our goal was to collect evidence on the prospects of using self-assessment for measuring learners' meaning knowledge, multiple-meaning knowledge in particular. We hypothesized that when asked to assess how many meaning senses of a given word are familiar to them, L2 learners will provide reliable information in regards to their multiple-meaning knowledge.

While the Vocabulary Knowledge Scale by Paribakht and Wesche (at least its original version) does not measure multiple-meaning knowledge, we feel that it is vitally important to obtain this data in order to establish the full picture (or the depth) of learner's word knowledge. In summary, the current study addresses the following questions:

Research Questions

1. Will learners of Russian be able to recall more than one meaning of a polysemous word when required?
2. Are learners of Russian able to provide reliable information regarding their meaning knowledge?

Thus the purpose of this study was to establish whether learners are able to assess their knowledge of multiple meanings. We believed that this information would provide some indications of how L2 learners perceive their knowledge of multiple meaning senses.

4.2 Study

4.2.1 Method

4.2.1.1 Target words

The target words were selected from the Russian Word Frequency List from levels: 2k, 3k, 4k and 5k (Vakar, 1966; also checked with Аграев et. al., 1977). The chosen words possessed at least two major meaning senses, which were verified by the Oxford Russian Dictionary (2000). Because the study explored multiple- meaning knowledge we needed to be sure that the participants were exposed to the words during their course of study. Therefore, the initially selected words were compared against the vocabulary in the Russian language teaching materials which were used to teach students on Russian courses at Cardiff University. The finally selected 20 words possessed at least two major meaning senses (checked with the Oxford Russian Dictionary, 2000) which were thought to be familiar to the students. These words are shown in table 4-1.

Table 4-1: Target words and their meanings

Word	Meanings
1. Лицо	1.Face; 2.Exterior; 3.Person; 4.Identity
2. Дорога	1.Road; 2.Journey; 3.Way (route)
3. Мир	1.World; 2.Peace
4. Собрание	1.Meeting; 2.Assembly; 3.Collection
5. Порядок	1.Order; 2.Succession; 3. Manner; 4.Customs
6. Месяц	1.Month; 2.Moon
7. Ручка	1.Hand; 2.Handle(arm); 3.Pen
8. Взгляд	1.Glance; 2.Opinion
9. Язык	1.Tongue(anat.); 2.Tongue(cul); 3.Language; 4.Clapper; 5.Prisoner; 6.Sole
10. Свет	1.Light; 2.Daybreak; 3.World; 4.Society
11. Право	1.Law; 2.Right; 3.Really
12. Совет	1.Advice; 2.Conference; 3.Council; 4.Harmony
13. Вечер	1.Evening; 2.Party
14. Вид	1.Look; 2.Shape; 3.View; 4.Prospect; 5.Sight; 6.Type; 7.Specious; 8.Aspect(gram)
15. Письмо	1.Letter; 2.Writing; 3.Script; 4.Style(of painting)
16. Учить	1.To teach; 2.To be a teacher; 3.To say; 4.To learn
17. Здоровый	1.Healthy; 2.Health-giving; 3.Robust; 4.Strong(sound); 5.Good at
18. Тяжёлый	1.Heavy; 2.Oppressive(smell); 3.Hard(difficult); 4.Slow(brain), 5.Severe; 6.Serious(bad); 7.Painful(heavy); 8.Difficult(personality); 9.Ponderous(heavy)
19. Полный	1.Full(of); 2.Stout(plump)
20. Повторить	1.To repeat; 2.To revise

4.2.1.2Participants

Nine Russian foreign language students attending the Advanced Russian language course at Cardiff University took part in this study. All students had been studying Russian as a foreign language for at least three years prior to the experiment. They represented different backgrounds and ages. Their proficiency levels ranged from intermediate to advanced. A variety of the first language backgrounds were represented here: English, French, Italian and Bulgarian.

4.2.1.3Procedure

The assessment was conducted in three steps. The first step was a self-assessment. In order to determine how many meanings of the target words the subjects knew, they were asked to fill in the self-assessment forms. Learners were requested to place a tick mark in the column which showed the number of meaning senses they thought they knew, ranging from 0 to 3+. For example, if the learner was not familiar with the word, they were supposed to choose the first column --"0 meanings". However, if a subject knew (or believed that they knew) one, two, three or more meanings of the target word, they had to choose their option and place a tick mark in the corresponding column.

At the second stage, learners were required to demonstrate their actual meaning knowledge. The same clean forms were given to subjects, however, this time they were requested to write down the meanings of the words which they had marked as known on the self-assessment form at the first stage. They had to provide all of the meanings they could think of, for each target word. No prompts were given.

Finally, the test-takers were asked to produce sentences with the target words: one for each meaning named by them on the assessment form. The format of the assessment was explained in detail. The Russian word "***коса***" (**plait; scythe; spit**) was used as an example and we went through "the entire battery" for this word. Students were not allowed to discuss the task or target words. Neither were they allowed to consult a dictionary. There was no time-restriction.

4.2.2 Results

The research questions posed in this study address the issues of learners' abilities to self-measure and demonstrate L2 multiple-meaning knowledge. We needed this information in order to argue our point regarding the importance of measuring multiple-meaning knowledge within lexical self-assessment methodologies. The results of the self-assessment and actual assessment of the subjects' multiple-meaning knowledge including their ability to use identified meaning senses in a sentence are presented in Table 4-2. This table shows the number of occasions on which students indicated that they knew 0, 1, 2 or 3 (or more) meaning senses of the twenty targeted words.

Table 4-2: Total results for nine subjects

Task	0 meanings	1 meaning	2 meanings	3+ meanings
Self-Assess.	9	35	101	35
Assessment	12	42	94	34
Use in Sent	17	55	90	18

In the actual assessment, the participant scored "0 meanings" if they did not provide any answer at all or if all the meaning explanations submitted by the participant were incorrect. Further on, erroneous responses were deducted from the total number of responses received. In the "use in a sentence" section, no attempt to create a sentence or complete misuse of the target word was regarded as "0 meanings". Any misuse of the target word (in any of its meanings) was considered as a failure for that particular meaning and the sentence was deducted from the total number of sentences provided. The individual results for each of the nine participants are shown in table 4-3.

Table 4-3: Individual results for nine subjects: self-assessment vs. assessment (use in a sentence is not included)

words	GC		ID		SQ		DG		KC		DT		LC		VL		PO	
	sa	a	sa	a	sa	a	sa	a	sa	a	sa	a	sa	a	sa	a	sa	a
1 (4)	2	2	2	2	3	3	3	3	3	2	2	1	1	1	3	2	1	1
2 (3)	2	2	2	3	3	3	3	3	3	2	2	2	1	1	2	2	1	1
3 (2)	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3
4 (3)	2	2	2	2	2	2	2	2	2	2	1	1	0	0	2	2	1	1
5 (4)	3	2	0	0	3	2	3	3	1	1	2	2	0	0	2	2	1	0
6 (2)	2	2	2	2	2	2	2	2	2	2	2	2	1	1	2	2	1	1
7 (3)	2	3	3	3	3	3	3	3	2	1	1	1	2	2	3	3	2	2
8 (2)	2	0	2	2	2	1	2	2	2	1	2	2	1	1	2	2	0	0
9 (6)	2	2	3	3	3	3	3	3	3	3	2	2	2	2	3	3	2	2
10 (4)	1	1	2	2	2	2	1	1	2	1	1	1	1	1	2	2	1	1
11 (3)	2	2	3	3	3	3	3	3	2	2	1	1	2	2	2	2	1	0
12 (4)	2	2	3	3	3	3	3	3	2	2	1	1	1	1	2	2	2	1
13 (2)	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	1	1
14 (8)	2	2	3	3	2	1	3	4	3	3	2	2	1	1	1	1	1	1
15 (4)	3	3	3	3	3	3	2	2	3	3	1	1	1	1	2	3	1	1
16 (4)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
17 (5)	2	2	2	1	2	2	2	2	2	2	2	2	1	1	1	1	1	1
18 (9)	2	2	2	2	2	2	3	3	2	2	3	2	0	0	2	1	0	0
19 (2)	2	1	2	2	2	2	2	2	2	2	2	2	0	0	2	2	0	1
20 (2)	2	2	2	2	2	2	2	2	2	2	1	1	1	0	2	2	0	0
Mean	2.1	1.95	2.2	2.2	2.4	2.25	2.4	2.45	2.2	1.95	1.7	1.6	1.05	1.05	2.05	2.0	1.1	1.0

Note. SA- known meanings in self-assessment, A-known meanings in assessment. Numbers in brackets (1(4); 2(3); 3(2) etc.) show the total number of meanings of a target word.

Table 4-3 indicates that the participants were generally very good at assessing their multiple-meaning knowledge. As seen from the table the mean scores from the self-assessment task were very close or matched those from the actual assessment. Though subjects failed to submit more than three meaning senses (apart from one case) where required, they demonstrated that they were generally able to provide information regarding their knowledge of multiple meaning senses as well as retrieve those meanings in the actual assessment.

To summarize the results of the actual (demonstrated) knowledge we used the following scoring system. If the participant provided the correct meaning sense for the given word and used it correctly in a sentence, they were considered to demonstrate Productive Knowledge 2 and were awarded 2 points. In case the participant submitted the correct meaning sense for the target word but failed to use it correctly in a sentence, their response was regarded as

Productive Knowledge 1 and was awarded 1 point. Unknown meaning senses received 0 points. The Meaning Proportion was calculated by dividing the learner's total score for each word by the number of possible points (number of meaning senses multiplied by 2 points each). The results for each of the nine participants are shown in table 4-4.

Table 4-4: Meaning proportion

words	GC	ID	SQ	DG	KC	DT	LC	VL	PO
1	.38	.50	1.0	.50	.50	.25	.25	.50	.25
2	.67	.83	1.0	1.0	.67	.67	.33	.67	.33
3	.50	1.0	1.0	1.0	1.0	.75	1.0	1.0	1.0
4	.50	.64	.50	.67	.67	.33	.00	.67	.33
5	.50	.20	.50	.75	.25	.50	.00	.50	.00
6	1.0	1.0	1.0	1.0	1.0	1.0	.50	1.0	.50
7	.67	.83	1.0	1.0	.33	.33	.67	1.0	.50
8	.00	1.0	1.0	.50	.50	1.0	.25	.75	.00
9	.25	.50	.50	.42	.42	.33	.33	.42	.33
10	.75	.37	.50	.13	.25	.50	.25	.38	.25
11	.33	1.0	1.0	1.0	.67	.67	.67	.50	.00
12	.50	.75	.75	.63	.38	.25	.25	.50	.25
13	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	.50
14	.19	.31	.22	.44	.38	.25	.12	.12	.20
15	.63	.50	.75	.50	.63	.25	.25	.63	.25
16	.38	.50	.50	.38	.50	.25	.38	.50	.38
17	.30	.50	.40	.40	.40	.40	.20	.20	.20
18	.22	.22	.22	.28	.22	.22	.00	.22	.00
19	1.0	1.0	1.0	1.0	.75	1.0	.00	1.0	.50
20	1.0	1.0	1.0	1.0	1.0	.50	.00	1.0	.00
Mean	.54	.71	.74	.68	.57	.52	.32	.63	.29

As seen from the table, the average meaning proportion was .60. A meaning proportion of 1.00 was diagnosed in 46 instances. There were 56 instances of a meaning proportion higher than .7. However, as indicated in the table, the majority of cases fell in the .2 to .6 range. The implications of these results will be discussed in the next section.

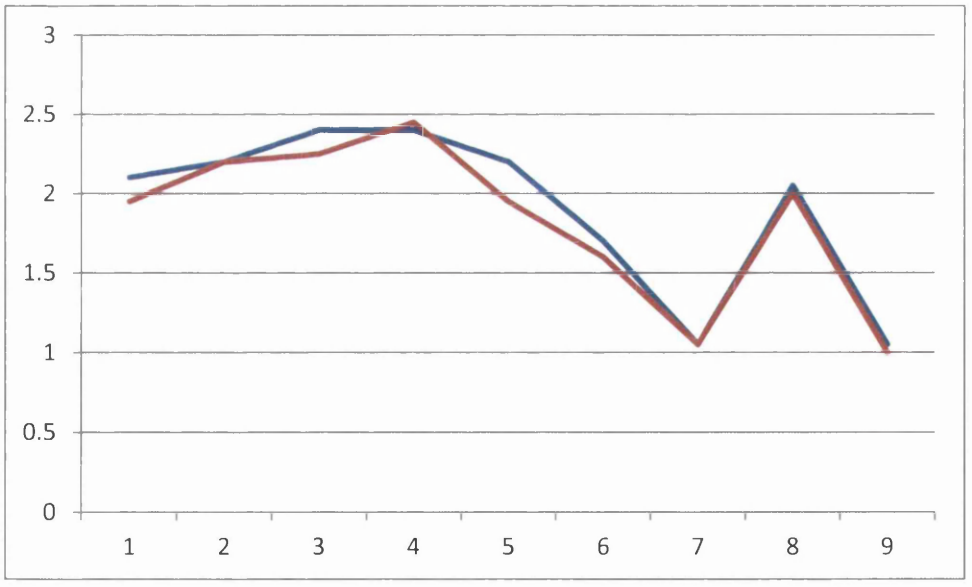
Pursuing the objective of our study, we analyzed the relationship between the self-assessment and assessment scores. In order to compare the results of learners' self-assessment of their own meaning knowledge against the results of the actual assessment of this knowledge, Pearson's correlation test was administered in this study. We did not run paired T-test or ANOVA in this study due to the small number of subjects. Table 4-5 shows the mean scores for self-assessment and actual assessment.

Table 4-5: Mean scores for self-assessment, assessment and difference.

	Self-assessment	Assessment	Difference
Mean Score	1.91	1.83	.078
SD	.528	.513	.094

As indicated in the table, the mean number of the self-assessment was very close to the mean number of the actual assessment. Furthermore, the statistical analysis revealed high correlation between the self-assessment and assessment scores: Pearson's correlation coefficient **r** = .984 which suggests a very strong positive relationship. The correlation was statistically significant: $p<.001$. The strength of the relationship between the self-assessment and assessment mean scores is further demonstrated by figure 4-1.

Figure 4-1: Self-assessment of multiple meaning knowledge (mean scores) vs. Assessment of multiple meaning knowledge (mean scores)



Self-assessment and assessment mean scores are shown on the vertical axis with a range of 1-2.4. The subjects (from 1 to 9) are shown on the horizontal axis. As we can see, the match between these two curves is apparent.

Thus the finding of this study (though carried out on a small number of subjects) confirms the research hypothesis in regards to the validity and reliability of self-assessment of Russian

language learners' multiple-meaning knowledge.

4.3 Discussion

This section deals with two issues posed as questions in the introduction section of the current study: demonstration of multiple-meaning knowledge and the ability to self-measure that knowledge. We will start by discussing the results received from the self-assessment and the teacher's assessment. Then, we will speculate on the reliability of the data obtained from the lexical self-assessment. Finally, we will talk about the limitations of this study.

4.3.1. Self-assessment vs. assessment

The study focused on self-assessment and assessment of knowledge of meaning senses of twenty Russian words by nine learners who had been studying Russian as a foreign language at Cardiff University. This study sought to identify how participants would respond on both (self-assessment and assessment) tasks since the importance of knowledge of more than one meaning of Russian polysemous words had not been emphasized to learners before, in the course of study. We felt that self-assessment of multiple-meaning knowledge should constitute an important part of the lexical self-assessment system. Our study was especially important in the light of the common tendency in L2 vocabulary self-assessment research to ignore (i.e. not to test) the knowledge of more than one meaning senses of polysemous words

The data collected in this study also shows whether learners of Russian were able to retrieve different meaning senses of a polysemous word as well as use their knowledge in practice. In terms of demonstrated multiple-meaning knowledge, the responses submitted appear to be a little disappointing. As noted in the results section, the average meaning proportion was .60 This indicated that in many instances, only partial meaning knowledge was achieved which is quite low for advanced students. A meaning proportion of 1.00 was diagnosed in 46 instances That suggested that all of the meaning senses for a target word were known productively and used correctly in sentences. In other words, in 46 instances, the meaning knowledge of certain target words was achieved to full productive mastery. In all the other cases, the knowledge of meaning senses was regarded as incomplete or partial. On the positive side, there were 56 instances of a meaning proportion higher than .7. However, as noted above, the majority of cases fell in the .2 to .6 range. Generally, the target words with fewer meaning senses seemed

to have been mastered the best: for example, nearly all the participants demonstrated full mastery of the words: ***мир (world and peace) (No 3), месяц (month and moon) (No 6)*** and ***вечер (evening and party) (No 13)*** (Table 4-3). On the other hand, the subjects fell far behind with 4- or more- meaning words, especially a 9-meaning word **тяжёлый (No 18)** or an 8-meaning word **вид (No 14).** Typical recall for those words was just 2 to 3 meanings. Sadly, none of the participants demonstrated the knowledge of all of the 76 meaning senses of the 20 targeted words. Overall, despite the fact that the average meaning proportion appeared to be higher than .5, the multi- meaning knowledge of the Russian vocabulary appeared to be low for the level tested.

However, at the same time, the data suggest that on most occasions, learners did manage to produce more than one meaning sense of the target polysemous words. Table 4-2 demonstrates that the number of occasions on which students indicated that they knew (self-assessment) and provided (assessment) two meanings of the target words was considerably greater than that for one meaning: 101 (SA) vs. 35; 94(A) vs. 42. The number of cases when learners stated that they knew and/or submitted 3 or more meaning senses for the given words was also considerably high: 35 (SA) and 34 (A). Generally speaking, while assessing their own knowledge of meaning senses, all the subjects (but one) chose a "more than one meaning" option and "three or more meanings" option at least for one of the target words. Only 3 participants admitted that they did not know one (ID) or a number (PO and LC) of the target words. In the assessment, the results were similar. As seen from table 4-3, 100% provided two or more meaning senses for at least one of the target words. Moreover, it is worth noting that among the submitted meanings, there were also those of low frequency, for example, ***язык -prisoner who would speak or право-really(No 9).*** Also, we have to bear in mind that only productive knowledge was tested. This means that some of the meaning senses might have been known to a different extent at the receptive level and might have been recalled after prompts had been demonstrated or in a context. This implies that learners' knowledge of all meaning senses of polysemous words can and should be self-/measured, otherwise only partial information of learners' word knowledge is extracted.

Unfortunately, the multiple-meaning knowledge did not receive proper consideration in Paribakht and Wesche's VKS. This is one of the arguments against their claim regarding the ability of their self-assessment instrument to test the depth of lexical knowledge.

It was also desirable within this study to measure learners' abilities to use target words in sentences. It is always important that learners show their abilities to use words appropriately in speech and writing rather than just explaining what a word could mean. Hence, we intended to establish whether learners who submitted correct meaning senses for the given words, were able to create sentences with these words in all their different meanings. The results (Tables 4-2 and 4-3) showed that on most occasions, learners were able to support the meaning senses, provided by them on the assessment form, with sentences illustrating these meanings: 4(P1) vs.90 (P2) for two meaning senses, 16 (P1) vs. 18 (P2) for three meaning senses. This suggests that learners mastered words in their certain meanings to a higher productive level (P2).

There were a number of occasions in which learners did not seem to give the correct meaning sense/s of the target word on the assessment form, however, later on, they used the word correctly in a sentence. In these cases, we asked the subjects to provide the meaning/s again. If the student proved their knowledge of the meaning in question by explaining it or by translation, the meaning sense was regarded as known and 2 points were awarded.

The implication of these results is the necessity of testing and self-testing learners' abilities to use a polysemous word in a sentence in each of its meanings. And again, it should be noted that contemporary self-assessment approaches do not appear to pay enough attention to this aspect of word knowledge. However, evidently, this information is vital for measuring the depth of the learner's knowledge of a certain polysemous word.

4.3.2 Reliability of the self-assessment results

In this study, we attempted to investigate the issue of self-assessment of multiple-meaning knowledge. Tables 4-3 and 4-5 illustrate how self-assessment results related to the results of the actual assessment. Even visual inspection of the means reveals a possible trend between the self-assessment and assessment scores. Figure 4-1 also shows an apparent match between the self-assessment and assessment curves. Furthermore, this is supported by the statistical analysis. A strong positive relationship between the self-assessment of multiple-meaning knowledge and the demonstration of this knowledge is confirmed by the results of Pearson's correlation test ($r = .984$; $p < .001$). Moreover, for two subjects (ID and LC), the complete match between the mean numbers of self-assessment and assessment was recorded.

Furthermore, for another three subjects (DG, VL, PO) the difference between the means was .05. Evidently, only 19 cases (1%) of overestimating of the target word meaning knowledge were noted. There were also 5 cases (0.3%) of underestimating of meaning knowledge. Furthermore, there was generally a high level of consistency in subjects' responses to the 20 content words tested.

It must be noted that though we cannot generalize the results received from nine subjects tested on twenty words, the trend for further research can be outlined. The findings of the present study clearly suggest that self-evaluation of meaning knowledge can be considered as a reliable, workable assessment instrument in measuring learners' word knowledge.

In summary, the implications of these findings seem to be as follows. The results obtained in the current study provide new insights into the self-assessment of word knowledge. It is clear that learners tend to assess their knowledge of different meaning senses (rather than just one) of the word. In light of these results, it seems unfortunate that this aspect of word knowledge has been neglected in self-assessment instruments designed so far. Evidently, this important constituent in the structure of vocabulary knowledge should be tested separately.

4.3.3Limitations

In this study, we have come across some difficulties. First, it was not always easy to determine the learners' knowledge of similar meaning senses of the same word. It also appeared to be difficult for subjects themselves to differentiate between similar meaning senses of a Russian polysemous word. It surfaced during the experiment that some meaning senses (different dictionary entries) of polysemous Russian words are regarded as the same meaning of a similar English word. Due to those lexical differences between Russian and English, subjects seemed to struggle to determine whether it was an additional explanation (a synonym) of the same meaning of a target word or a different meaning sense. Another issue that caused problems for our subjects was the fact that in Russian, a number of meaning senses of some polysemous words manifest themselves in set phrases (or idioms) only. This makes it literally impossible for a learner to produce that type of meaning sense unless they are familiar with a certain expression that meaning is used in. Second, of course, there was some degree of subjectivity on both sides (the researcher and the subjects) in measuring and

scoring the subjects' multiple-meaning knowledge.

4.4 Conclusion

In this chapter, we investigated how learners themselves assess their meaning knowledge of L2 polysemous words. In summary, the data collected shows that all the participants indicated that they knew more than one meaning of at least one of the targeted words. Furthermore, they could readily provide two or more meanings for these words. The data also show a good match between the knowledge of a word's meaning senses and the ability to use the word in its different meanings in a sentence.

Furthermore, the data strongly suggest the reliability of the information (though obtained from a small number of subjects) provided from self-measuring of multiple-meaning knowledge ($r = .984$; $p<.001$). This implies that multiple-meaning knowledge should be included into lexical self-assessment methodologies as one of its targets.

Having stated the importance of measuring multiple-meaning knowledge within lexical self-assessment approaches, we are now interested to explore other criteria (or features) of learners' categorizations of their lexical knowledge. Thus in the next chapter, we will continue exploring the issue of self-classification on a larger set of words in order to obtain a more detailed picture of this process and its outcomes.

CHAPTER FIVE

Self-Categorization of L2 Word Knowledge

Experimental Study 3

5.1 Introduction

In the previous studies, we explored how learners of Russian self-classified and self-assessed their lexical knowledge and multiple-meaning knowledge in particular. We established (though on small numbers of subjects) that learners were able to self-measure their knowledge of multiple-meaning senses and provide reliable information in regards to their knowledge. Now we are curious to investigate which other features (besides meaning knowledge) will be used by learners to categorize their knowledge of words.

The study reported in chapter 3 showed that the categories learners use to describe their own vocabulary knowledge are rather more complex than researchers assumed them to be. The majority of the subjects, tested in our first study, created more than five categories which is more than the number of categorizations suggested in the VKS approach. Many of these classifications were complex in terms of their qualitative features as well. It is also noteworthy that there was little overlap between categorizations used by different subjects – though some subjects referred to grammar aspect or word class, the subjects did not use these features systematically. The only features that appeared in the categorizations of all the subjects are meaning and certainty in knowledge.

As seen from above, the data we have collected in our research so far is rather interesting. It challenges the common assumptions in the field in regards to the existing methodologies in L2 lexical self-assessment. This suggests that this matter is worth pursuing further.

In the current study, we intend to receive more data on the issue of self-categorization of L2 vocabulary knowledge. We decided to test a single subject. This learner acted as part of the group in study 1. Our intention here was to compare categorizations produced by the same subject when tested under different experimental conditions (i.e. as part of the group of subjects vs. single subject).We also decided to increase the number of target words to 200 which include the original 50 words used in study 1. The advantage of using 200 words rather

than 50 lies in the opportunity to reveal smaller categorizations (probably containing just one word) which would not normally surface when a small number of words are categorized. We also assumed that a larger set of words might reveal new qualitative features in subjects' classifications. We also hope that a one-to-one experimental condition will create the right environment for the participant to relax and at the same time focus entirely on the task. Generally speaking, by making all these alterations we intended to explore the process of L2 lexical self-categorization and its outcomes in more detail since we believed that the new experimental conditions might provide more evidence on how L2 word knowledge is measured and categorized by learners.

Similar to our previous studies, this research was carried out on the basis of Russian as a second or foreign language. Since the Russian language (the Slavonic language group) is regarded to be an unusual language (as it differs significantly from Western European languages) it was of particular interest to see how an English native speaker would rate their knowledge of 200 Russian words.

We continue to investigate the issue raised previously in our research that is the assessment criteria considered by learners while measuring their own knowledge of words. We are looking for an answer to the question: "Will learners focus entirely on their meaning knowledge (Paribakht and Wesche's VKS approach) or will they look into the other aspects of their word knowledge?"

Summarizing the above, the current study aimed to establish how learners classify their own knowledge of words. We needed that information in order to establish how self-assessment of word knowledge is carried out and which features of that knowledge receive priority in the learner's mind. It was interesting to see whether the self-report categories suggested by the subject in the current study would differ (quantitatively and/or qualitatively) from the categorizations produced by the learners in our previous study. Furthermore, it was worthwhile to compare the classification system created by the subject in the present study against the set of categories he suggested previously, in study 1, when he acted as part of the group of subjects.

Research Questions

In summary, the current research was led by the following questions:

1. Will the learner be able to create a larger number of categories while assessing his knowledge of a larger number of words?
2. Will the newly created categories refer to a wider range of features?
3. Will the status (i.e. the original location) of the target words used in study No 1 change when self-assessed by the subject again in this study?

5.2 Study

5.2.1 Method

5.2.1.1 Target Words

Two hundred Russian words were selected for this study. They fell under the following 5 levels of the Word Frequency List: 1k, 2k, 3k, 4k and 5k (Vakar, 1966; also checked with Аграев et. al., 1977). In order to ensure that the targeted words match the subject's level of proficiency, the selected words were compared against the vocabularies in the teaching programs for the Russian language courses at Cardiff. Each word was typed onto a separate card. There were 200 word cards in total including fifty target words used in study No 1 (Chapter 3). The target words were selected from the four word classes: nouns, verbs, adjectives and adverbs. The words are presented in table 5-1.

Table 5-1: Target Words

Врач *doctor*	**Честный** *honest*	**Болезнь** *illness*	**Пора** *it's time*	**Думать** *to think*
Счастье *happiness*	**Дорогой** *dear*	**Город** *city, town*	**Слышать** *to hear*	**Знать** *to know*
Лицо *face, person*	**Краснвый** *beautiful*	**Извиненне** *apology*	**Государство** *state*	**Рожать** *to give birth*
Случай *case ,event*	**Последний** *last; latest; the latter*	**Подряд** *in succession*	**Письмо** *letter, manuscript, writing*	**Молодёжь** *youth*
Дорога *road, way*	**Каждый** *every(one)*	**Глаз** *eye*	**Хозяйство** *housekeeping*	**Образование** *education*
Душа *soul, spirit*	**Нужный** *necessary*	**Подпись** *signature*	**Общество** society	**Положеиие** *position*
Мир *world, peace*	**Красный** *red*	**Страна** *country*	**Вид** *look, view, form*	**Значить** *to mean*
Собрание *meeting;assembly*	**Всё** *everything*	**Справедливый** *just, true (correct)*	**Ручка** ***pen, arm***	**Понимать** *to understand*
Журнал *magazine*	**Прямо** *straight*	**Земля** *earth, soil*	**Бояться** to be afraid	**Казаться** *to seem*
Сердце *heart*	**Замуж** *marry (fem)*	**Слово** *word*	**Камень** *stone*	**Смотреть** *to look*
Порядок *order ;manner*	**Вечер** *evening, party*	**Правило** *rule*	**Приглашать** *to invite*	**Жить** *to live*
Улица *street*	**Завод** *plant (factory)*	**Огонь** *fire*	**Ответнть** *to answer*	**Приходиться** *to fit, to fall, to have to*
Разговаривать *to talk*	**Капитан** *captain*	**Женщина** *woman*	**Тут** *here*	**Оставаться** *to stay*
Предлагать *to offer*	**Друг** *friend*	**Вместо** *instead*	**Даже** even	**Жениться** *to marry*
Принимать *to accept*	**Молодой** *young*	**Мост** *bridge*	**Существовать** *to exist*	**Снова** *again*
Терять *to lose*	**Угол** *corner, angle*	**Место** *place, seat, room*	**Голос** *voice*	**Любимый** *favourite*
Поднимать *to raise*	**Отдых** *rest*	**Нота** *note*	**Берег** *shore*	**Домашний***domestic*
Придумать *to devise*	**Вопрос** *question*	**Удовольствие**pleasure	**Путь** ***way, journey***	**Довольный** *satisfied, contended*
Учить *to learn, teach*	**Конец** *end*	**Мысль** *thought*	**Дверь** *door*	**Здоровый** *healthy*
Лежать *to lie*	**Совсем** *entirely*	**Часть** *part ,section, unit*	**Война** *war*	**Возможно** *maybe*
Читать *to read*	**Тихо** *quietly*	**Книга** *book*	**Только** *only*	**Действительно** *really*
Сидеть *to sit*	**Быстро** *quickly*	**Народ** *people*	**Дело** ***matter, business, deed, cause***	**Вместе** *together*
Повторить *to repeat*	**Молодец** *well done*	**Вещь** *thing*	**Помещаться** *to be*	**Правый** *right, correct*
Спешить *to hurry*	**Совет** *advice, council*	**Великнй** *great*	**Согласиться** ***to agree***	**Детский** *children's*
Получить *to receive*	**Свет** *light; world*	**Пост** *post, fasting*	**Помощь** *help*	**Новый** *new*
Находить *to find*	**Взгляд** *look, view,*	**Рукав** *sleeve*	**Возместить** *to compensate*	**Тогда** *then*
Уезжать *to leave*	**Дети** *children*	**Солнце** *sun*	**Труд** *labour, work*	**Просто** *simply*
Ждать *to wait, to expect*	**Жизнь** *life*	**Зря** *in vain*	**Сила** *force*	**Личный** *personal*
Позволить *to allow*	**Мужчина** man	**Успеть** *to manage*	**Почти** *almost*	**Написанный** ***written***
Звонить *to ring*	**Сразу** *at once*	**Наука** *science*	**Заниматься**	**Прекрасный** *beautiful*
Кушать *to eat*	**Упасть** *to fall (perf)*	**Духота** *stuffiness*	**Деньги** *money*	**Скоро** *soon*
Вернуться *to return*	**Урод** ugly person	**Считать** *to count; to consider*	**Область** *region,field*	**Особенно** *especially*
Взять *to take (perf)*	**Отдыхать** *to rest*	**По-всякому***in any way*	**Писатель** writer	**Высокий** *tall (high)*
Взглянуть *to look at*	**Слушать** *to listen*	**Однако** *however*	**Спрашивать***to ask*	**Тяжёлый***heavy,hard*
Объяснить *to explain*	**Рука** arm, hand	**Мешать** ***to prevent, to disturb, to stir, to blend***	**Видеть** *to see*	**Школьный** *school*
Умереть *to die (perf)*	**Люди** *people*	**Решить** *to decide; to solve*	**Хотеть** *to want*	**Вдруг** *suddenly*
Завидовать *to envy*	**Время** *time*	**Просить** *to ask (beg)*	**Мочь** *to be able*	**Больно** *painfully*
Послать *to send*	**Право** law, *right*	**Давать** ***to give***	**Рядом** *near*	**Конечно** *of course*
Надеяться *to hope*	**Человек** *person*	**Стоять** ***to stand***	**Вроде** *like, as if*	**Также** *also (too)*
Прощать *to forgive*	**Игра** *game*	**Называть** *to call (name*	**Безусловно** *absolutely*	**Можно** *possible*

5.2.1.2 Participant

The participant in this study was a male (in his 30s) student of an Advanced Russian class at Cardiff University. He had been learning Russian as a foreign language for four years prior to this test. His native language is English. During the University course of the Russian language study, this student demonstrated good (for the level) knowledge of Russian vocabulary

according to the assessment reports. We tested this learner earlier as a member of the group of subjects in our study No 1 reported in chapter 3.

5.2.1.3 Procedure

The participant was given 200 cards in a random pile. Each card contained one word printed on it. The instruction was as follows: "Divide this pile of cards into categories according to how well you know the word on the card." Thus, the participant was required to arrange the target words into self-evaluation categories suggested by himself. The subject was not restricted in time. He did not receive any further advice on the criteria for his classification. After all the cards had been combined into groups, the participant was required to explain his classification system and the criteria he used to arrange the words into the categories, in detail. The subject's explanations were recorded. All the word cards for each of the created categories were placed in a separate envelope with the subject's description of the category written on it. The participant received a payment for participation in this study.

5.2.2 Results

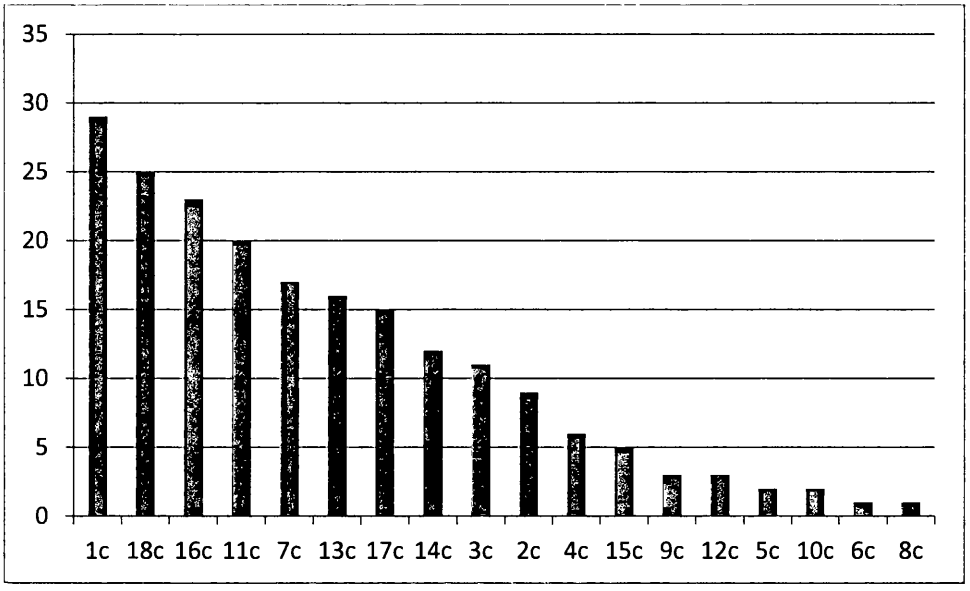
Addressing the research questions raised in this study, we asked the participant to categorize his knowledge of the 200 given words. The subject divided the 200 target words into 18 different categories. The created self-report categories are shown in table 5-2.

Table 5-2: Self-report categories suggested by the subject

Category Description	No of Words
1.Words which I know the meaning and feel confident using them, conjugating and declining, etc.	29
2. Words which I know more than one meaning of. I am able to use them in more than one context.	9
3. Verbs that I do not recognize. I would not feel confident conjugating because of irregular endings, etc.	11
4. Words which "have irregularities". I could use them in certain circumstances.	6
5. Nouns that I know are irregular. I would recognize them and might be able to use them in certain circumstances.	2
6. Expressions.	1
7. Words which I would be able to guess at their meanings but I would not be sure what they meant... similar (phonologically) to other words ...	17
8. Words I know the meaning but I would not be confident in using or I would not know which prepositions come before or after them.	1
9. Words which I recognize from them being similar to the other words I know; can use in more than one form, for example: noun--verb or verb-noun.	3
10. Words which I know the meaning of and I can write them down but beyond that I would not be able to do anything with them.	2
11. Words -- I do not know what they mean, I do not know what type of words they are, etc.	20
12. "Joining words" but I do not know exactly how to use them in sentences.	3
13. Adverbs. I know the meaning(s) and feel confident using.	16
14. Verbs. I do not know their meanings but would be able to conjugate them as they look regular.	12
15. Words which I would not be able to produce myself but I would be able to recognize when written down (only).	5
16. Verbs which I know the meanings of and can conjugate, etc.	23
17. Adjectives which I can "conjugate" and use in context.	15
18. Words I do not recognize and would not be able to do anything with them.	25

The table indicates that our participant created a rather complex classification system which contains 18 categories (though categories 11 and 18 can be combined in one category: "totally unknown"). This is nearly four times more than used in the VKS approach. The main features the subject refers to in his descriptions for the arranged categories are: meaning knowledge, use, and grammar. Generally speaking, the qualitative features of this system seem to be rather diverse. This will be discussed in the section that follows. Figure 5-1 shows how the target words were spread between the created categories.

Figure 5-1: Distribution of words among the suggested categories at stage 1.



Note. 1c, 18c, 16c, 11c ... on the horizontal axis indicate numbers of the created self-report categories. Vertical axis shows amount of words allocated under the categories.

The figure shows that approximately half of the words fall into categories 1, 18, 16 and 11. The second largest group includes categories 7, 13, 17, 14 and 3. The third group contains categories 2, 4 and 15 with several words each. The last block contains the remaining categories: 5, 6, 8, 9 and 10 with only 1 to 3 words in them.

5.3 Discussion

This section discusses the classification system created by the subject in this study. Addressing the first two questions of the study, we will look into the quantitative and qualitative features of the created categorization. Then, we will compare this categorization against the sets of categories produced by a group of subjects (including our current testee) in the first study of our research. Finally, we will talk about the issue of movements of some of the targeted words between the categorizations created in different studies of this thesis.

5.3.1 Quantitative and qualitative features of the created classification

The data submitted here supports the findings of our previous studies in relation to L2 learners' abilities to measure and categorize their knowledge of words. The number of categories created by the subject while classifying his knowledge of the given words was 18. This is a lot more than considered by other scholars and included in any lexical self-assessment methodologies proposed so far. In terms of the proportion of words allocated

under the suggested categories, it should be noted that a vast majority of words (namely, 111 or 56 %) were placed into various "known" categories. At the same time, only 45 words (23%) were named as totally unknown. There were rather large (e.g. cat.1, cat.18, cat.16 etc) and very small categories (e.g. cat. 6 and cat. 8). Interestingly, there were two one-word categories, two two-word categories and two three-word categories. The existence of such small categories might imply that some features can be detected (i.e. come to light) only on larger sets of words. This will be further addressed in the next section.

Generally speaking, a large number of categories used by a single subject to describe his lexical knowledge may suggest that the learner rates his knowledge of words in a detailed way. This assumption is also supported by the qualitative characteristics of the created system. Despite the fact that there was some sort of repeat (e.g. cat 11 vs. cat.18), the subject's descriptions for the developed categories refer to a rather large number of different features. This is shown in table 5-3 (p. 11).

As seen from the table, the main features of the learner's descriptions for the arranged categories in this study are: meaning knowledge, use in sentences and grammar information. In light of the findings of our previous study No 2 (chapter 4), in which we argued for the importance of self-measuring learners' multiple-meaning knowledge, it is worth emphasizing that the subject of our current study suggested a separate category for the words (nine words in total) he knew more than one meaning sense of. Furthermore, the learner noted that he was also confident in using those words in their different meanings in context. Thus, as indicated by the data, learners themselves are willing and prepared to measure their own knowledge of multiple meaning senses.

However, at the same time, the data revealed that the meaning knowledge was not the compulsory feature to form a category. A number of categories suggested by the participant (for example Cat.3 and Cat.14) indicate that even in the cases when the subject was not aware of the words' meaning senses, he was still able to supply some kind of information about those words. It means that he demonstrated other types of word knowledge rather than knowledge of meaning senses, for instance word class knowledge.

The learner seems to have paid special attention to context use. On a number of occasions, he

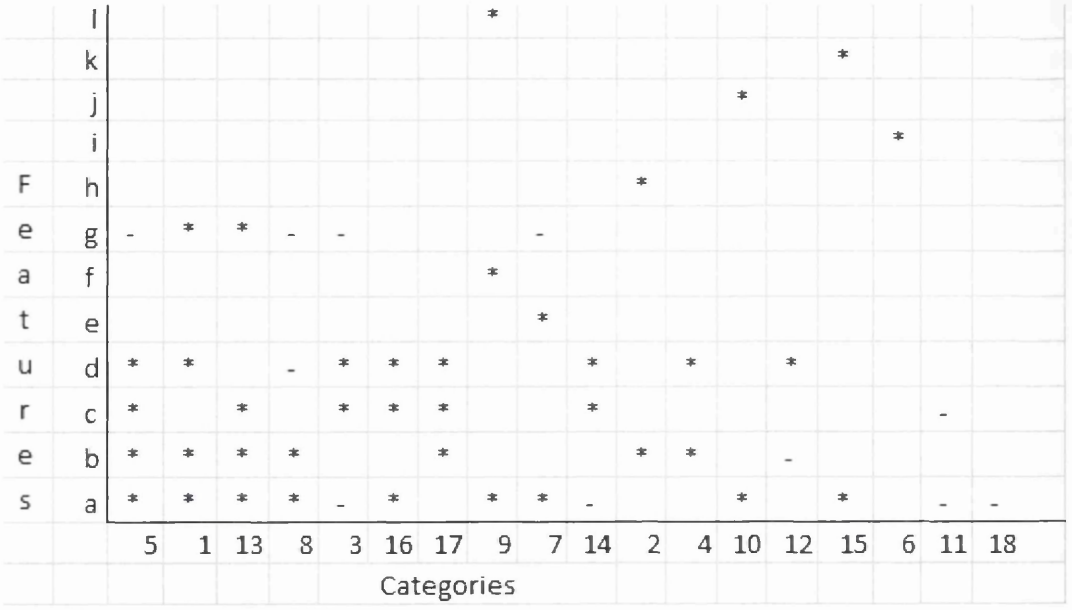
indicated his ability to use particular target words "in certain circumstances" only. This suggests the importance of testing learners' abilities of using a word in various types of context. In addition, he created a separate category for collocations/expressions. While linked to the ability to use a word under various circumstances, this involved knowledge of set phrases, sayings, proverbs, etc. Hence, it might be worthwhile to also assess this kind of learners' word knowledge within a lexical self-assessment scale.

For the first time in our research, knowledge of derivation forms was used as a criterion for a learner's categorization. Evidently, in Category 9, the subject states that he could recognize and use various derivation forms of the words placed by him into that category.

Also, it appears that the subject tended to differentiate between receptive and productive language skills. Some of the key words in his descriptions were "to recognize" and "to use". Furthermore, describing a number of categories created (for example, Cat.5 and Cat.9), the subject used both of these words, reporting that he was able to perform both activities with some of the targeted words. However, describing Cat.15, he states that he was able to recognize the words placed into this category "when written down (only)". This suggests that once again (see Exp. 1), the learner attempted to assess his word knowledge from the practical viewpoint by estimating his abilities to use the target words within the four language skills: speaking, listening, writing, and reading.

The data also revealed a dense distribution of features among the arranged categories. Figure 5-1 shows how different features were spread among the categories.

Figure 5-1: Distribution of features among the self-report categories



Note. Features shown on the vertical axis: a-meaning; b-use; c-word class; d-grammar info; e-phonetic info; f-morphological info; g-certainty in knowledge; h-multiple-meaning knowledge; i-collocation knowledge; j-spelling; k-receptive vs. productive skills; l-knowledge of derivatives.

As seen from Figure 5-1, the majority of the created categories are based on a combination of different features. Thus the description for category 5 refers to five various features including one feature (certainty in knowledge) used in a negative context. Categories 1, 13, 8 and 3 are based on four features. The third block contains categories 16, 17, 9, 7 and 14, the descriptions of which refer to three different features. The final block includes the remaining eight categories based on two or less features. The figure also indicates that the subject referred to different features both in positive and negative context. The implication of this is that various features seem to be closely interrelated. Furthermore, a certain word appears to be associated with particular features regardless of how well the learner knows the word.

5.3.2 A comparison of different categorizations submitted in our research so far.

Addressing the third research question of the present study, we carried out a comparison of the different categorizations created in our research so far. First, we compared the two categorizations produced by the same subject in studies 1 and 3. Then, we compared the classification created in the current study with the sets of categories submitted in study 1. When two classifications produced by the same learner are compared, the results reveal some

differences in quantitative and qualitative features of these categorizations. In terms of quantitative differences, the comparison indicated that in this study, the subject arranged given words by creating three times as many classification categories as he produced initially in study 1, namely eighteen vs. six. Thus in regards to its structure, the latest classification system appears to be far more complex than the set of categories suggested by the subject in study 1. A comparison of the qualitative characteristics of the two categorizations shows considerable differences in the number of features (or classification criteria) the learner referred to in his descriptions for the created categories. This is illustrated in table 5-3.

Table 5-3: Features reflected in the descriptions for the categories created by the same subject in studies 1 and 3

Features	Experiment 1		Experiment 3	
	+	-	+	-
Meaning	4	1	9	4
Use	1		7	1
Word Class	1		6	1
Grammar	2		8	1
Written Form			1	
Sound Form			1	
Morphology			1	
Derivatives			1	
Multiple Meaning			1	
Four Skills			1	
Collocations			1	
Certainty	4		2	4

+ indicates that the feature was mentioned in a positive context; - shows that the feature was mentioned in a negative context.

As indicated in Table 5-3, a large number of new features were considered by the subject as criteria for the classification of his knowledge of 200 given words. In fact, in the current study, the subject used more than twice as many features as he referred to in his initial self-assessment (namely twelve vs. five). Also, more features (compared to his first categorization) were used in a negative context to indicate a lack of knowledge or lack of certainty in knowledge in regards to certain word knowledge types. Analysing the data illustrated in table 5-3, we note that in his first categorization, the subject placed emphasis on meaning knowledge, though he did not separate words with multiple meanings, word class knowledge and grammar knowledge. He also referred to context use as a classifying criterion.

In his second categorization, the subject keeps and expands his previous "criteria" and, at the same time, employs a large number of new criteria for his categorization. They are: knowledge of multiple meanings, knowledge of derivation forms, spelling knowledge morphological information and knowledge of collocations. He also differentiates between his productive and receptive skills. Similar to his first classification, the subject's categorization criteria here fall within the two- dimensional space: knowledge and certainty.

Thus the classification system submitted in this study appears to be a lot richer (i.e. multi-featured) than the one produced by the same learner in study 1. The possible reason that would explain the difference between the two classifications lies in the different experimental conditions administered in these studies including the number of words tested. Presumably, some features (or word properties) e.g. collocation knowledge or multiple-meaning knowledge can only be revealed when knowledge of a rather large number of words is self-measured. It also seems very probable that one-to-one (with the researcher) experimental conditions enabled the subject to fully concentrate on the task and measure his knowledge of the words more scrupulously.

Having established such considerable differences between the two categorizations created by the same subject (acting on his own in the present study and as a part of the group of subjects in study one), we are now curious to investigate how this latest classification differs from the categorizations suggested by the other subjects in study 1. The average number of categories created in study 1 was six which constitutes only a third of the number of categories suggested by the learner in the current study. Let us take a look at the classification criteria used by learners in study 1 compared to those applied by the subject of our current study.

Table 5-4: Features referred to in the descriptions for the categories in studies 1 and 3

Subject	Meaning		Use		W/class		Gram		Spelling		Phonol		Morph		Deriv		Multipe-Meaning		4 Skills		Colloc		Certainty	
	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
TS	6	1	3		4		6	1															6	2
JM	5	1	3		1		1		1	1									4					2
SL	3	1	1				2																2	2
HE		1									2													1
BG	2	2	3																2				2	1
RP	2	1																					2	
OP	5	2			6																			2
SM	4	2			1	1	1											1						2
AC	2	1			1		1										1						1	
MB	2	1	1				1																	2
C-M	2	1	1																1					1
Current Subject	9	4	7	1	6	1	8	1	1		1		1		1		1		1		1		2	4

+ indicates that the feature was mentioned in a positive context; - shows that the feature was mentioned in a negative context.

Comparing the data presented in the table, we note that the current classification appears to reflect on a wider range of word knowledge types than the sets of categories submitted earlier (in study 1). Moreover, as seen from the table, the new categorization seems to have embraced the classification criteria used by the group of learners in our first study. At the same time, a number of totally new features were reflected in the descriptions for the categories developed by the current subject. They are: knowledge of collocations/set phrases, morphological knowledge and knowledge of derivation forms. Thus in summary, the comparison indicates that the classification system created in the present study appears to be the most complex in our research so far. It consists of significantly more categories and refers to considerably more aspects of word knowledge than all the categorizations submitted earlier in this research.

Again, we point out that the word knowledge types the subject referred to, are completely missed out in Paribakht and Wesche's Vocabulary Knowledge Scale. However, as the results of this study indicate, these aspects of word knowledge should be traced by a self-assessment scale.

5.3.3 Re-location of words

A new issue that emerged in this study relates to re-location of words between the self-rate categories. Since 50 out of 200 target words used in this study were also used in study 1, we were curious to see whether the subject's knowledge/ or his perception of his knowledge of those words changed over time. Indeed, the current study found that the status of a number of target words which had been employed in both studies did change. Nine words out of 50 used in study 1 seem to have changed their initial positions within the subject's categorization. The changes are shown in table 5-5.

Table 5-5: Re-location of words among the self-report categories

words	unknown	might be able to guess what the word means	can recognize when written	known	known as multiple meaning words & can be used in different context
Лицо *Face; person*				Exp 1	Exp 3
Случай *Case; event*		Exp 1		Exp 3	
Порядок *Order; manner*	Exp 1				Exp 3
Собрание *Meeting; assembly*	Exp 1				Exp 3
Дорога *Road; way*		Exp 1		Exp 3	
Последний *Last;latest; the latter*		Exp 1		Exp 3	
Душа *Soul; spirit*		Exp 1	Exp 3		
Звонить *To ring*	Exp 1			Exp 3	
Повторить *To repeat; to revise*	Exp 1			Exp 3	

This table indicates how the nine words changed their positions in study 3 compared to study 1. For example, word *лицо* moved from the category "known" (study 1) to the category "known as multiple meaning words and can be used in different context" (study 3). As seen from the table, the words followed different routes while changing their initial status. This might hint at the existence of different patterns of acquisition and development of these words in the subject's lexicon. This never emerged in other research. Thus it would be of particular interest to further explore this issue on a larger set of words in order to determine the most common patterns of word re-locations. We hope that this investigation will determine

possible stages in the development of knowledge of certain words.

5.3.4 Summary

What follows from the results of this study? The two main points should be highlighted here. They are:

l) the way/s learners measure and rate their own knowledge of words;

2) re-location of words among the self-report categories suggested by the learner.

This never turned up in the other studies and, hence, needs special attention.

1) The data collected here seems to support the findings of our previous study that learners use a comprehensive (i.e. multi-dimensional and multi-featured) approach while assessing their knowledge of words. The categorization system produced by our current subject is rather complex and rich. In fact, it is a lot more diverse than assumed by Paribakht, Wesche and other scholars in the field. In general the findings of our study imply that the existing self-assessment instruments for measuring the depth of learners' word knowledge do not reflect (or match) the complexity of the self-report categorizations produced by learners themselves. The question here is whether the large number of self-report categories created by the learner in this study is an exception, a one-off case or a normal way of word knowledge measuring? This will be addressed in the next chapter.

2) The issue of word re-locations among self-rating categories never turned up in other studies. This is another important finding of our research. The fact that some words changed their location (since they were self-assessed for the first time) by moving into different categories in a newly created categorization might indicate how lexical knowledge is acquired. Furthermore, patterns of these re-locations may illustrate possible directions or routes of L2 lexical acquisition. If it is so, the question arises: What might influence or initiate this change of location. In other words, what forces words to move from one category to another? To which extent do learning activities, for instance reading, promote word relocation? We will explore these issues in the chapter that follows.

5.4 Conclusion

In this chapter, we continued exploring how a L2 learner categorizes his knowledge of 200 target words. The data obtained in the present study reveals a rather complex nature of the learner's categorization system. This study shows that the learner was able to create a very

large number of self-rating categories while assessing his knowledge of a larger (compared to study 1) portion of words. In fact, he developed three times as many categories (18 in total) as he suggested in his previous self-assessment (study 1).

Furthermore, the results of this study seem to support the findings of our previous research that learners tend to use a comprehensive (i.e. multi-featured and multi-dimensional) approach while self-assessing their knowledge of words. The newly created categories reflected on a rather wide range of features (12 in total).

The issue of word relocation from one self-report category to another seems to be an interesting finding of this study. The status of a number of target words used in study No 1 changed when self-assessed again in the current study. This will be investigated further.

In the next chapter, we will continue investigating the issue of self-categorization of L2 lexical knowledge. We are interested to explore how a language activity (e.g. reading) will effect the way a learner categorizes their knowledge of words. We will also look into the issue that emerged from the current study, i.e. the re-location of words among the created categories in the learner's classification systems.

CHAPTER SIX

Self-Categorization of Word Knowledge: Further Research

Experimental study 4

6.1. Introduction

The study reported in the previous chapter indicated that the word categorization used by our subject to describe his knowledge of given words is rather more complex than it has been assumed by researchers in the field. In fact, both classifications of words created by the subject in studies 1 and 3 were complex. Moreover, the number of self-report categories included in his last categorization (chapter 5) is nearly five times as many as suggested in Paribakht and Wesche's approach. Furthermore, there was some overlap between the categorizations designed by that subject in studies 1 and 3 with the features of meaning and certainty in knowledge appearing systematically in both classifications.

Overall, as our previous studies indicate, learners tend to arrange given words in a number of different ways (quite creatively) in order to express how well they know these words. The 18 categories suggested by the subject in study 3 (Chapter 5) was the greatest number of self-report categories created by a single subject in our research. Furthermore, a large number of different criteria seem to have been chosen as the basis for his complex categorization. This might imply that the arrangement of words in his mental lexicon is more complex and multi-structural than it has been assumed by researchers. Of course, it is not wise to make generalizations on grounds of the data obtained in a single case study. However, since our finding contradicts the existing beliefs in the field we feel this issue needs further exploring.

It also emerged from the last case study that words seem to change their locations within the self-report scales suggested by the learner at different times of testing. In other words, they move from one self-report category to another over a certain period of time. It is not clear at this stage whether words move freely between the different categories or some routes of their movements are more likely than others. Further investigation of the phenomenon of word relocations might provide information on the issue of how foreign language words are developed in the mental lexicon.

Following the findings of our previous experimental studies, a new investigation into the process of self-reporting on the knowledge of words was launched. At this point, we seek to explore some of the issues suggested by the findings reported earlier in this thesis. This time, we were curious to find out how the subject would perform on a set of more challenging words, most of which were believed to be unknown for the subject. The target words for this study were selected from a Chekhov story, the original source printed in Russia. Since the majority of the words were thought to be unfamiliar to the subject, it was interesting to establish whether and how these words would change their status in the subject's mind after the reading.

Objectives

The outlined issues determined the objectives of our current study. They were as follows:

1) to establish whether a large number of self-report categories suggested by the subject in the previous study (18 in total) would be created again on a new set of target words;

2) to compare newly proposed categories with those suggested by the subject in study No 3;

3) to explore whether target words (placed by the subject into particular categories) would change their status and relocate into different categories on the self-rating scale whilst being involved in language activities, such as reading.

Research Question

How will learner classify his own knowledge of words before and after the reading?

6.2. Experiment

6.2.1. Method

6.2.1.1 Target Words

A set of 200 Russian words had been selected for this study. The words had been extracted from the story "Дама с собачкой" (A Lady with a Dog) by Чехов (Chekhov). We chose words most of which we believed would be unknown for the subject. However, in order to sustain the subject's interest in our study a number of easier words had also been included. We took into account the fact that the subject, though being a student in the Advanced class of the Centre for Lifelong Learning at Cardiff University, had attended short (2 hours a week, 24 weeks) Russian language courses. And thus, his vocabulary might not have been as rich as vocabularies of advanced students at Russian language departments who work towards a degree in Russian. We also avoided old fashioned, bookish words which are not in use in contemporary Russian. Each word had been printed on a separate white paper card in black ink; 200 word cards in total.

The target words represented four parts of speech: nouns, verbs, adjectives and adverbs. These words are shown in Table 6-1.

Table 6-1: Target Words

Появиться *to appear*	**Располагать** *to have available*	**Ненависть** *hatred*	**Вихрь** *whirlwind*	**Шорох** *rustle*
Довольный *satisfied*	**Равнодушно** *indifferently*	**Заря** *dawn, sunset*	**Мол** *pier*	**Сытый** *satisfied*
Привыкший*accustomed*	**Приключение** *adventure*	**Чувство** *feeling*	**Значительный** *significant*	**Поделиться** *to share, to divide*
Душно *stuffy*	**Глухо** *deserted*	**Роса** *dew*	**Испугать** *to frighten*	**Побледнеть** *to become pale*
Рост *growth, height*	**Память** *memory*	**Завять** *to fade/ wither*	**Пристань** *harbour*	**Догадываться** *to guess*
Объяснить *to explain*	**Смех** *laughter*	**Пора** *it's time*	**Промелькнуть** *to flash ,to fly by*	**Грустный** *sad*
Сквер *public garden*	**Забавно** *funny*	**Опуститься** *to go down*	**Отчётливо** *distinctively*	**Унизительный** *humiliating*
Наверное *probably*	**Обстановка** *decor, situation*	**Восхищаться** *to admire*	**Хищный** *predatory*	**Дрожать** *to shiver*
Сообразить *to understand* , *to arrange*	**Занять** *to occupy, to borrow*	**Печально** *sadly*	**Особенность** *peculiarity*	**Пьянство** *alcoholism*
Порядочный *decent*	**Цель** *aim (goal), target*	**Жаловаться** *to complain*	**Упрямый** *stubborn*	**Даль** *distance*
Казаться *to seem*	**Однажды** *one day*	**Уважать** *to respect*	**Приставать** *to pester, to stick*	**Чепуха** *neusance*
Сила *force*	**Жалкий** *pitiful, sorry*	**Тревожно** *anxiously*	**Капризный** *capricious*	**Раскаяние** *repentance*
Солидная *solid, respectable*	**Причёска** *hair style*	**Молчание** *silence*	**Обращаться** *to address, to treat*	**Возмущаться** *to be outraged*
Влечь *to draw, to attract*	**Пыль** *dust*	**Ревность** *jealousy*	**Блестеть** *to shine*	**Испытывать** *to test, to feel*
Мыслящая *thinking*	**Походка** *walk/gait*	**Свеча** *candle*	**Растерянность** *confusion*	**Надоесть** *to pester*
Нерешительный *indecisive*	**Пароход** *steamship*	**Страх** *fear*	**Отрывистый** *jerky/ abrupt*	**Сведения** *information, news*
Втайне *secretly*	**Выражение** *expression*	**Гореть** *to burn*	**Некстати** *at the wrong moment*	**Устроить** *to arrange*
Не спеша *unhurriedly*	**Толпа** *crowd*	**Страстно***passionately*	**Пристально** *intently*	**Смутить** *to embarrass/ confuse*
Считать *to count, to consider*	**Нравы** *manners/ways*	**Перестать** *to stop*	**Унылый** *downcast*	**Свидание** *meeting*
Верить *to believe*	**Случай** *case, event*	**Вдруг** *suddenly*	**Влага** *moisture*	**Нищий** *poor*
Недалёкий *not far off*	**Победа** *victory*	**Слёзы** *tears*	**Приветливый** *friendly*	**Освещать** *to lighten*
Умолять *to plead/beg*	**Поворачиваться** *to turn*	**Скрипка** *violin*	**Пугливо** *fearfully*	**Развлекаться** *to amuse o.s.*
Изящный *graceful*	**Взглянуть** *to cast a glance (at)*	**Презирать** *to despise*	**Насмешка** *gibe*	**Трогательный** *touching/ moving*
Грех *sin*	**Пассажиры** *passengers*	**Водопад** *waterfall*	**Духи** *perfume*	**Вынуждать** *to force*
Бояться *to be afraid*	**Мысль** *thought*	**Обмануть** *to deceive*	**Колокол** *bell*	**Оправдание** *justification, excuse*
Успокоиться *to calm down*	**Нарядный** *elegant*	**Впечатления** *impressions*	**Беззаботный** *carefree , careless*	**Дразнить** *to tease*
Изменять *to change*	**Ласкаво***affectionately*	**Служить** *to serve*	**Жадность** *greed*	**Жечь** *to burn*
Мёртвый *dead*	**Молчать** *to be silent*	**Удаваться** *to succeed*	**Добродушный** *good-natured*	**Досада** *annoyance*
Опыт *experience*	**Нюхать** *to smell/sniff*	**Любопытство** *curiosity*	**Лестно** *flatteringly*	**Прятаться** *to hide o.s.*
Волны *waves*	**Движение** *movement*	**Судьба** *fate/ fortune*	**Искренность** *sincerity*	**Торопиться** *to be in a hurry*
Горький *bitter*	**Обнять** *to embrace*	**Безумный** *mad, wild*	**Покрыться** *to cover o.s*	**Качаться** *to swing*
Фонарик *torch*	**Рассвет** *dawn*	**Звонок** *bell*	**Неуместный** *inappropriate*	**Смущение** *confusion*
Общество *society*	**Поцеловать** *to kiss*	**Раздражать** *to irritate /to annoy*	**Ужас** *horror*	**Бестолково** *disorderly*
Церковь *church*	**Очарованный** *fascinated*	**Исчезнуть** *to vanish*	**Обморок** *faint*	**Мрачный** *gloomy*
Скучно *bored*	**Запах** *smell*	**Шутить** *to joke*		
Неподвижно *motionlessly*	**Сказочный** *fairytale*	**Крик** *scream*		
Чувствовать *to feel*	**Желание** *desire/wish*	**Публика** *public*		
Кричать *to shout*	**Сторож** *guard*	**Воспоминание** *recollection*		
Привлекательный *attractive*	**Властный** *imperious*	**Пахнуть** *to smell*		
Покой *rest/ peace*	**Подробности** *detail*	**Следить** *to watch, to follow, to look after*		
Страсть *passion*	**Шевелиться** *to stir*	**Тень** *shadow, shade*		
Метель *blizzard*	**Окурок** *cigar-butt*	**Вечный** *eternal*		
Воскресать *to revive*	**Сутулый** *stooping*	**Настраивать** *to tune*		
Воображение *imagination*		**Кланяться** *to bow*		
Дрянной *worthless*				

6.2.1.2 Participant

The subject in this study was the same learner who took part in our previous case study.

6.2.1.3 Procedure

The procedure consisted of two stages. Each stage included two steps.

Stage 1

Step 1.1) The subject was given the Russian story "A Lady with a Dog" (Дама с собачкой) by Chekhov (Чехов) in English translation to read at home.

Step 2.1) Then, in the classroom, we asked the participant to retell the story in order to check whether he had read it at home. After that, the main part of the study began. Similar to study No 3, the subject received 200 cards in a random pile. Each card contained one target word printed on it. The instruction was as follows: "Arrange the following cards into categories according to how well you know the word on the card". It was important that the subject fully understood the instruction. Our target was to encourage the participant to describe (or express) his knowledge of the given words using self-report categories. The subject did not receive any prompts from the experimenter. No time restriction applied. Having arranged all of the target words into categories, the subject described the created groups of words and explained the reasons for the classification he had produced. His descriptions were recorded.

Stage 2

Step 1.2) The participant was asked to read "Дама с собачкой» by Чехов in the original at home. The target words were not highlighted in the story in any way. The subject was not advised to use reference literature.

Step 2.2) After this task had been completed at home, step 2.1 was repeated in full in the classroom.

The subject's explanations for the created categories were recorded. He received a payment for participation in this study.

6.2.2. Results

The first research objective addresses the issue of quantity of subject's self-report categories in this study. The testee divided the 200 target words into 10 categories at the first stage of the study, and into 19 categories at the second stage. This is illustrated in Tables 6-2 through 6-3 as well as by Figures 6-1 and 6-2.

Table 6-2: Categories suggested by the subject at stage 1.

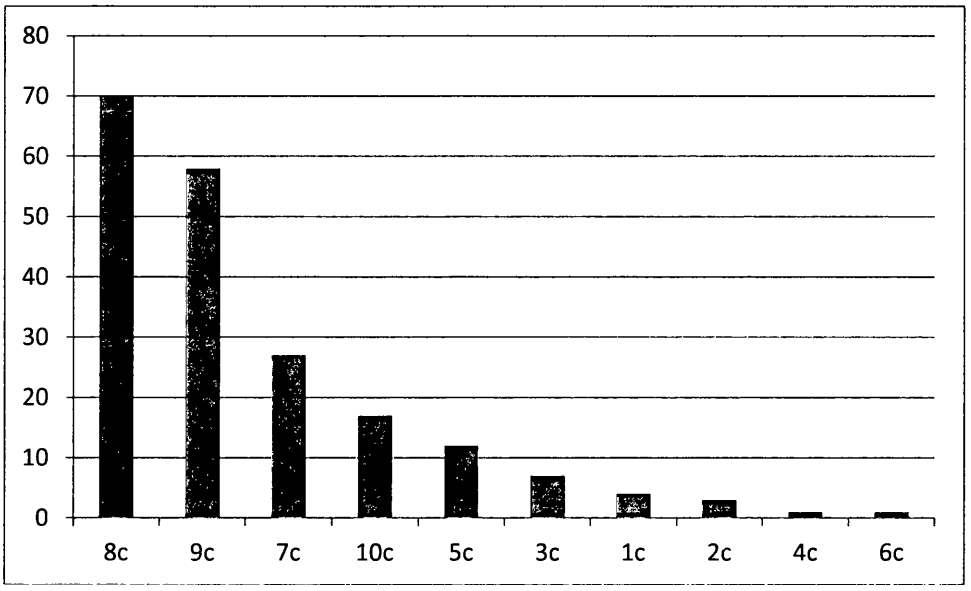
Category Description	No of Words
1. Words which I know the meaning of and feel confident using:	4
2. Words: I am unsure of their meanings:	3
3. Words I recognize, they look familiar but I cannot remember what their meanings are:	7
4. Words which I know the meaning of but I am not confident using them. I think they might be irregular:	1
5. Words: I have not seen them before but I can make a good guess of their meanings based on the components of the words:	12
6. Words I can guess the meaning of because they are similar to words in English:	1
7. Words I do not know their meanings at all but I do recognize they are adjectives:	27
8. Words I do not know their meanings at all but I do recognize they are verbs:	70
9. Words I do not know their meanings at all but I do recognize they are nouns:	58
10. Words I do not know their meanings at all but I do recognize they are adverbs:	17

Note. Self-report categories are listed in the order arranged and presented by the subject.

As seen from the table, at the first stage of this study, the 200 target words were spread among 10 categories created by our subject. This is substantially more than suggested by Paribakht and Wesche. The created categories seem to fall into 3 groups. Group 1 contains categories 1, 2, 4 and 6 with 1 to 4 words. Group 2 includes categories 3, 5 and 10 with 7 to 17 words in the categories. Group 3 contains categories 7, 8 and 9 with the largest number of words: 27 to 70. Categories 4 and 6 seem to have been created around one word only.

It is worthwhile noting that word class information seems to be rather important as a classification criterion here. Clearly, suggested categories 7, 8, 9 and 10 can be combined into one single category: Words I do not know but can recognize their word class. Figure 6-1 also shows how the target words were spread between the created categories.

Figure 6-1: Distribution of words among the suggested categories at stage 1.



Note. 8c, 9c, 7c, 10c ... on the horizontal axis indicate numbers of the created self-report categories. Vertical axis shows No of words allocated under the category.

The figure shows that there are a large number of different categories. Most of the words fall into categories 8, 9 and 7. Categories 4 and 6 contain only one word each.

Stage 2

At the second stage of the study, after the reading, the participant divided the same targeted 200 words into 19 self-report categories. The created categories are illustrated in Table 6-3.

Table 6-3: Categories suggested by the subject at stage 2

Category Description	No of Words
1. Words I know the meanings of. I think they have multiple (or complex) meanings:	4
2. Nouns which I know the meaning of and feel confident using:	12
3. Nouns which I know the meaning of but do not feel totally confident using:	7
4. Adverbs that I know the meaning of:	3
5. Adjectives that I know the meaning of:	5
6. Verbs that I know the meaning of:	3
7. Verbs that I know the meaning of but do not feel as confident using:	1
8. Words: I am unsure of their meanings:	2
9. Words which I can guess the meanings of based on their similarities (morphological and sound) to words I do know:	8
10. Words: I am unsure of their meanings, however I could associate the meaning based on the component within the word (certain parts of the word):	12
11. Words: I am unsure of the exact meaning but recognize parts of the word:	6
12. Adjectives which I know the meaning of but would not feel confident using as I think they might be irregular:	1
13. Words which I recognize based on the fact that they are similar to English word:	1
14. Words which I know the meaning of but I am unsure in which context or how exactly to use them:	2
15. Nouns which I do not recognize:	28
16. Verbs I do not recognize:	37
17. Adverbs that I do not recognize:	3
18. Adjectives I do not recognize:	17
19. Words which look familiar but I cannot recollect their meaning. If I see them in context again I might "re-remember" them:	48

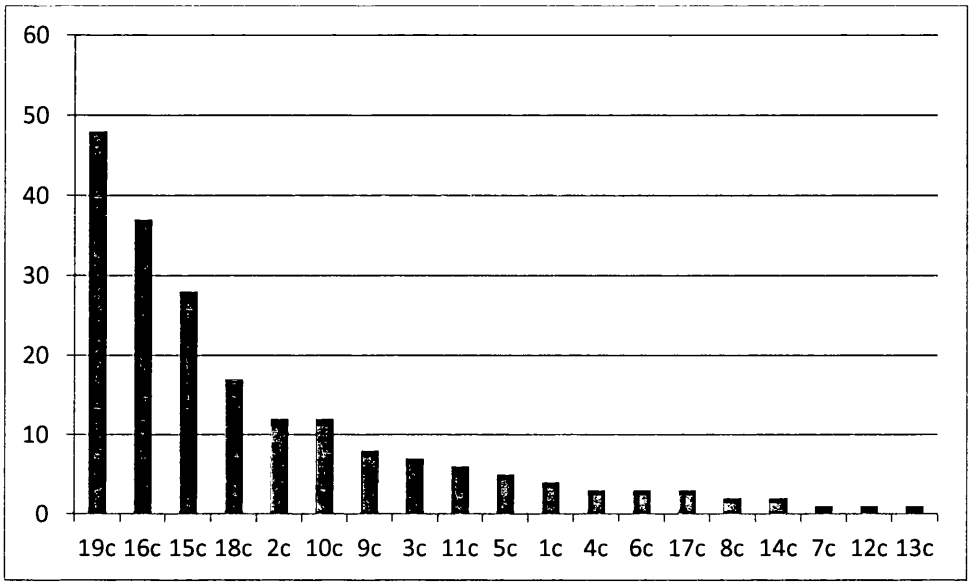
Note. Self-report categories are listed in the order arranged and presented by the subject

As seen from this table, the word class (part of speech) criterion is very distinctive in the second classification as well. Clearly, categories 4, 5 and 6 can be combined into one category: Words I know the meaning of and can recognize their part of speech. Likewise, categories 15, 16, 17 and 18 might be regarded as sub-categories of a single large category: Words I do not know but can recognize their word class.

It should be noted that at the first stage of the study, the subject placed a small number of words in wrong word class groups. For example, he classified some unknown adjectives and adverbs as nouns. However, it did not occur at the second stage of the study.

The learner thoroughly examined each word card and either created a new category for this word or added it to the words in a category already arranged. As seen from the results, there were some one-word categories, for instance Category 7, 12 and 13. This and the time scale (90 minutes) might indicate that the subject did analyse each word in detail, even if its meaning senses were unfamiliar to him. Figure 6-2 shows graphically how the target words are spread between the created categories at the second stage of the study.

Figure 6-2: Distribution of words among the created categories at Stage 2



Note. 19c, 16c, 15c, 18c, 2c ... on the horizontal axis indicate numbers of the created self-report categories. Vertical axis shows amount of words allocated under the categories.

This figure shows an even greater number of different categorizations compared to Figure 6-1. It looks as though all the categories fall into four main blocks. Block 1 contains categories 15, 16, 18 and 19 with more than half the words. Block 2 contains categories 2 and 10 with many words in them. Block 3 holds categories 3, 9 and 11 with several words in each of them. Block 4 contains categories 1, 4, 5, 6, 7, 8, 12, 13, 14 and 17 with only a few words in them.

The results reveal that both categorizations compiled at stages 1 and 2 of this study are complex. In fact, they are considerably more complex than assumed by researchers in the field. The implications of these results will be discussed in the section that follows.

6.3 Discussion

In this chapter, we set out to find an answer to the research question concerning the ways our subject would classify his own knowledge of words before and after the reading. The results confirmed those obtained in the previous studies of this research i.e. the complex nature of the subject's categorizations of his word knowledge. Furthermore, a number of interesting points concerning target word-relocations within the subject's second classification were revealed in this study. In this section, we will highlight the implications of these findings. The section contains two points. First, we will discuss the self-rating categories created by the subject before and after the reading. Then, we will speculate on the ways some target words have moved from their initial categories into different ones suggested in the subject's second classification.

6.3.1 Self-rating categorizations before and after the reading.

The first objective of the current study was to establish the number of self-report categories the subject would create while categorizing his knowledge of the targeted words before and after the reading. At the first stage of the study, the subject arranged the given 200 words in 10 categories. This appears to be a smaller number of categories compared against the results of our previous case study No 3 (18 categories). The reason for that might be the fact that meaning senses of 5 words only (out of 200) were reported as well-known by the subject. Thus, the target words selected for this study proved to be a challenge for the learner. However, despite an obvious lack of meaning knowledge, the subject managed to produce a relatively large number of self-report categories, ten in total. This, once again, indicates that knowledge of meaning senses does not appear to be the only criterion for learners' assessment of their word knowledge. Furthermore, this also suggests that the self-assessment scale introduced by Paribakht and Wesche and built mainly on the assessing learners' meaning knowledge, does not adequately reflect the complexity of the subject's own descriptions of their vocabulary knowledge. Contrary to the VKS approach, the self-report categories formed at this stage indicate that the learner seems to have considered various features (including a variety of word properties) while measuring and rating his own knowledge of words. Table 6-4 illustrates the features considered by the subject while categorizing his word knowledge at the first stage of this study. The table shows how many times the subject referred to a particular feature in his descriptions for the created categories.

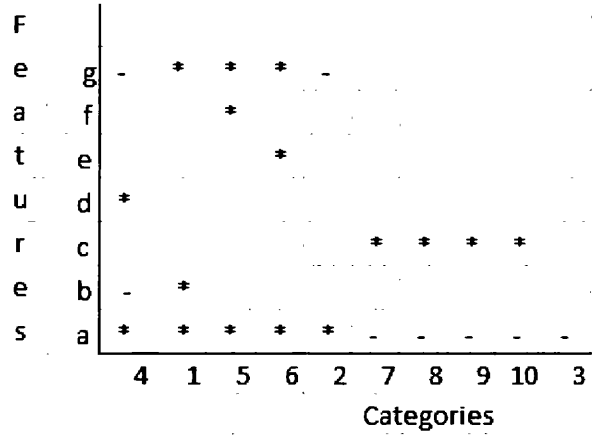
Table 6-4:Criteria/Features referred to in the descriptions for the categories at stage 1.

Exp 6	Meaning		Use		Word Class		Grammar		Sound Word Form		Morpholog Structure		Certainty in Knowledge incl. guess	
Stage 1	+	-	+	-	+	-	+	-	+	-	+	-	+	-
No of referrals	5	5	1	1	4		1		1		1		3	2

+ indicates that the feature was mentioned in a positive context; - shows that the feature was mentioned in a negative context.

This data suggests that the subject does analyze and rate words according to different criteria even if their meanings are unknown to him. Moreover, the subject attempted to figure out what the word might mean by carrying out a thorough analysis of its morphological and phonological structure. The table indicates that on a number of occasions, the subject referred to the aspects within which he was unable to describe the word at that stage. This might imply close links among various properties of a word. This point may also be confirmed by the way various features are distributed among the created categories. Let us take a look at Figure 6-3.

Figure 6-3: Distribution of features among the self-report categories in classification 1



Note. Features shown on the vertical axis: a-meaning; b-use; c-word class; d-grammar info; e-phonetic info; f-morphological info; g-certainty in knowledge

As seen from the figure, the subject's description of each category (apart from Cat 3) in classification 1 involves at least two features. This might suggest that different features, the subject refers to, are related. This, in turn, may imply that a word can be found in the lexicon by its different features. This will be investigated further in our research.

At the second stage of the study, the overall picture of the subject's categorization changed. The changes were both of a quantitative and qualitative nature. The descriptions for the created categories reflect a large number of features. This is shown in table 6-5.

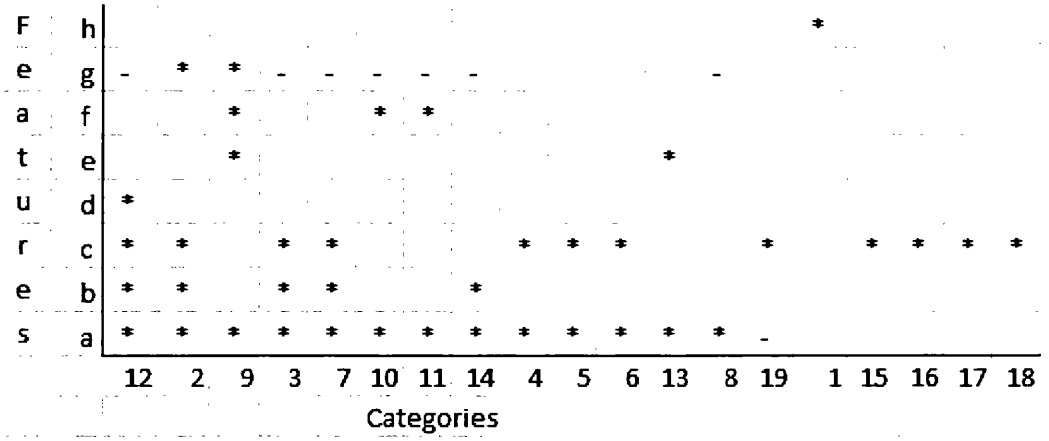
Table 6-5:Criteria/Features referred to in the descriptions for the categories at stage 2.

Exp 6	Meaning		Use		Word Class		Grammar		Sound Word Form		Morph Structure		Certainty		Multiple Meaning	
Stage 2	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
No of referrals	13	1	5		12		1		2		3		2	7	1	

+ indicates that the feature was mentioned in a positive context; - shows that the feature was mentioned in a negative context.

The table indicates that the main features in the subject's descriptions at this stage are meaning knowledge, word class information, the ability to use the word in a sentence as well as certainty in his knowledge. The second largest block of features contains morphological and phonological descriptors. Figure 6-4 illustrates how these features are spread between the self-report categories suggested after the reading.

Figure 6-4: Distribution of features among the self-report categories in classification 2



Note. Features shown on the vertical axis: a-meaning; b-use; c-word class; d-grammar info; e-phonetic info; f-morphological info; g-certainty in knowledge; h-multiple-meaning knowledge

As seen from figure 6-4, in the description to category 12, the subject refers to 5 different features. Categories 2, 9, 3 and 7 reflect 4 various features in their descriptions.

In total, a large number of categories, namely 14, involve two or more features which may be regarded as further evidence of interaction between different features or word properties.

A comparison between the two classification systems (before and after the reading) reveals that after the reading, the subject re-arranged the target words by creating nearly twice as many categories as he had suggested at the first stage of the study: namely, 19 compared to the initial 10. Moreover, the categorization system was completely restructured after the reading: a totally new category – a multi-meaning knowledge – appeared in the second categorization (Table 6-3), whereas a number of existing categories split into smaller word groups. For instance, category 4: "words: known meanings, not confident about use" (Table 6-2) split into categories 3: "nouns", 7: "verbs" and 12: "adjectives" -- known meanings, not confident about use (Table 6-3). At the same time, new categories, for example, categories 4: "adverbs", 5: "adjectives" and 6: "verbs" – that I know the meaning of – were added to the classification at the second stage (Table 6-3).

Another new category (No 9) collaborating morphological and phonological types of word knowledge also emerged in the classification system after the reading. Some of the categories from classification 1 underwent slight alterations mostly in terms of word class marking. For instance, category 1: "known words, confident use" (Table 6-2) changed to category 2: "known nouns, confident use" (Table 6-3). The categories whose descriptions did not change lost a number of words initially allocated under those categories. This will be addressed in the next section.

In respect of the qualitative differences between the two classification systems, it should also be pointed that a completely new feature or a word knowledge aspect: multi-meaning knowledge emerged in the second classification. Furthermore, as seen from figure 6-4 (compared to figure 6-3) the descriptions of the categories created in classification 2 refer to a larger number of features including those mentioned in a negative context. Thus, for example, category 12 refers to 5 different features.

Clearly, judging on the data collected and summarized in the tables and figures presented above, in terms of both quantitative and qualitative characteristics, the second categorization is more complex and richer than the one submitted before the reading. The meaning and word

class types of word knowledge seem to dominate as category descriptors in the second system. We can't note at this stage whether it is a peculiarity of the Russian word knowledge categorization in general or our subject's individual style of classifying Russian words. However, generally speaking, the implication of categories being different is the fact that a learner's categorization system of his word knowledge changes while this knowledge develops, in our case with a lexical input via reading. This point is not taken into account in Paribakht and Wesche's self-assessment methodology which suggests the same fixed categories for any stage or level of knowledge. Furthermore, it might be worthwhile for self-assessment scales to vary on the grounds of how knowledgeable and skillful the learner is.

In the light of these findings, it would be interesting to see how the latest classifications differ from the self-report categorizations suggested by the subject in our previous studies: 1 and 3. This constitutes the second objective of our current study. The quantitative and qualitative characteristics of the word categorizations suggested by the same subject in studies 1, 3 and 4 can be compared on the basis of the summarized data presented in tables 6-6 through 6-7.

Table 6-6:Number of self-report categories suggested by the subject in studies 1, 3 & 4.

	Study 1	Study 3	Study 4/1	Study 4/2
No of Categories	6	18	10	19

As seen from the table, all classification systems created by this subject contain more categories than suggested within the existing approaches to the vocabulary knowledge self-assessment.

The largest numbers of categories were created by the subject in studies 3 and 4/2 which might suggest that the quantitative peculiarities of learner's categorizations are determined by the nature of words being classified. At the same time, the difference in the data submitted in studies 4/1 and 4/2 might indicate how knowledge about a word develops in the subject's lexicon by acquiring new features and properties. Table 6-7 compares the distribution of features among the word categories suggested by the subject in studies 1, 3 and 4.

Table 6-7: Distribution of features among the self-report categories in studies 1, 3 and 4

Features	Exp 1		Exp 3		Exp 4/1		Exp 4/2	
	+	-	+	-	+	-	+	-
Meaning	4	1	9	4	5	5	13	1
Use	1		7	1	1	1	5	
Word Class	1		6	1	4		12	
Grammar	2		8	1	1		1	
Written Form			1					
Sound Form			1		1		2	
Morphology			1		1		3	
Derivatives			1					
MultipleMeaning			1				1	
Four Skills			1					
Collocations			1					
Certainty	4		2	4	3	2	2	7

+ indicates that the feature was mentioned in a positive context; - shows that the feature was mentioned in a negative context.

Table 6-7 shows the qualitative features of the two sets of classifications. A comparison of qualitative characteristics of the subject's categorizations of his word knowledge illustrated in table 6-7 indicates that all of his classification systems are multi-featured and multi-dimensional. Evidently, the richest i.e. referring to the largest number of features, categorization is the classification system submitted by the subject in study 3. The structure of knowledge of better known words seems to be more complex with more links developed between different aspects (types) of word knowledge.

However, despite the fact that the most recent classification lacks a number of word knowledge types reflected in the descriptions for the categories in study 3, it still engages a considerable portion of descriptors. At the same time, differences in the qualitative features of the subject's classification systems (studies 1, 3 and 4) may suggest that the process of word acquisition via reading (when learners encounter more challenging words and possess less knowledge about them) occurs differently from acquiring words through other language activities.

Overall, the learner used 10 different descriptors to characterize the word categories he created within three different studies. These descriptors reflect various aspects of word knowledge as well as learner's certainty in his knowledge. Moreover, each categorization

created by the subject differs from his previous scale/s, both quantitatively and qualitatively. It is worth emphasizing that every time (in study 3 compared to study 1; in study 4/2 compared to study 4/1) while self-rating and describing his knowledge of words, the learner added/ considered new features by referring to new aspects of word knowledge in his descriptions of the proposed categories. Again, this may imply the enormous complexity of the learner's lexicon organization.

In summary, a comparison analysis of the data presented in tables 6-6 and 6-7 indicates that the learner consistently created a fair amount of self-report categories referring to a large variety of features while self-assessing his knowledge of words. This is an important and promising finding as it might indicate that the subject's approach to his word knowledge is considerably more complex and richer than it has previously been assumed in vocabulary acquisition. Though, of course, it is not possible to generalize from the results submitted by one subject, further investigation into this matter is definitely worth pursuing. The importance of this finding lies also in the fact that even if the results obtained so far are not confirmed by further research involving different learners, they (the results) hint at the necessity of addressing the issue of an individual approach in lexical self-assessment methodologies.

6.3.2 Word re-location among the created categories.

Another promising finding of our research is the issue of word re-location. The results of this study indicate that the target words tend to move from one self-rate category to another. Presumably, these re-locations were caused by the involvement of the target words in reading. Though the subject had not been instructed to use reference literature while reading in Russian, he admitted consulting a dictionary on a number of occasions. He also compared the Russian original text with an English translation a number of times trying to understand certain phrases. Thus, the target words used in this study were involved in the process of learning through reading comprehension. Let us take a look at the changes that occurred to the subject's knowledge of the given words. The routes of word re-locations are shown in Table 6-8.

Table 6-8: Re-location of target words after the reading

No of Category Classification 1	Relocated to Category Classification 2	No of Words
3	3; 4; 10; 12	1; 1; 2; 1
5	1; 3; 4; 5; 6; 9	1; 1; 1; 1; 2; 1
7	5; 9; 10; 11	4; 2; 2; 2
8	1; 3; 6; 9; 11; 14; 15	1; 1; 1; 2; 3; 1; 6
9	1; 2; 3; 9; 10; 14	1; 6; 4; 1; 4; 1
10	1; 4; 9; 10	1; 1; 2; 2

As seen from the table, a considerably large number of targeted words moved from the categories in which they had been placed by the subject before the reading. From categories 5, 8 and 9, the targeted words were distributed among the largest number of categories – eleven in total, whereas category 9 lost the largest number of words – seventeen which moved into six different categories within the second classification scale. In summary, the positions of the target words after the reading (at the second stage of the study) are shown in table 6-9.

Table 6-9: Summary of the word re-locations after the reading

Status of words	**Improve**	**Stay the same**	**Regress**
No of words	60	140	0

Note. If a word gained a new feature/s (e.g. moved from "known" to "known with multiple-meanings or derivative forms) it was considered to have improved its position within the classification scale.

This table indicates in summary that a large portion of targeted words were transferred by the learner into different self-report categories within his newly suggested scale. These relocations seem to reflect the development of the subject's knowledge of the given words through reading. Over a quarter of the target words upgraded their location in the learner's self-rating scale after being involved in the subject's reading. This fact never surfaced in other research. However, the importance of this finding is apparent since it demonstrates how a word is acquired.

6.3.3 Summary

Overall, the results of the present study seem to confirm the earlier findings of our research regarding the ways the learner self-rates his knowledge of words. The data collected in this study indicates that the subject's description of his own word knowledge is complex. There are two main reasons for the complexity of his descriptions. They are: the involvement of a large variety of features and interaction between those features. This is illustrated by the tables and figures presented in this study. The implications of these findings are as follows. Clearly, the subject's rating of his own word knowledge is much more complicated and structured than suggested in Paribakht and Wesche's self-assessment methodology. In other words, Paribakht and Wesche's VKS categories are too simple. Considerably more than two features: meaning and use, the VKS is based on, are involved in subject's descriptions of his own word knowledge, namely 10. Furthermore, the relations between the features are more complex than a simple linear transition presented in Paribakht and Wesche's VKS approach. Evidently, complex classifications created by the subject at both stages of this study suggest that a new interpretation of L2 word knowledge self-assessment is needed. Subject's descriptions of his knowledge of words indicate that the self-assessment instrument should be a multiple-featured scale. This scale should reflect on a variety of features to which the subject refers in his own self-report categories.

The second interesting finding of our study is word re-location among the categories suggested by the subject. The patterns of these re-locations will be further examined in studies that follow. Most importantly, we discovered that not only the target words re-locate among the categories but also the list of self-rate categories changes over time. This might imply that the process of acquisition of a certain word may involve re-structuring of the mental lexicon in whole.

The data received would certainly contribute to understanding the structure of people's vocabulary knowledge. However, further investigation into this issue will be required in order to trace the consistency of the data obtained in our research so far. In our next study, we will attempt to find out whether other learners with different language and social background will show similar results whilst assessing their knowledge of the given words. We will also investigate whether subsequent sessions of reading will lead to further changes in subjects' self-rating scales.

6.4 Conclusion

This study addressed some issues that surfaced in the previous chapters, with the hope that the findings would shed some light on the prospects of how learners self-assess their own knowledge of words and how words are acquired, in general.

The study revealed that learner's classification of his lexical knowledge is a lot more complex, varied and multi-dimensional than people have assumed. These findings suggested directions for further research.

In light of the insights gained in this chapter, it seems reasonable to continue investigating the issue of self-rating L2 lexical knowledge. In the next chapters, we will explore the ways of measuring the word knowledge by different learners compared to the classification systems created by our current subject so far. We will also investigate the issue of word re-locations within the self-assessment scales.

CHAPTER SEVEN

Self-Rating of Word-Knowledge: before and after Reading—

--Further Research

Experimental Study 5

7.1. Introduction

In the previous chapters, we attempted to establish how learners of Russian as a foreign language self-rate their knowledge of words. The studies reported earlier in this thesis confirmed that the way our subject self-rates his knowledge of words is different from the vocabulary knowledge self-assessment scales suggested in contemporary methodologies. Furthermore, the previous studies indicated that the subject was consistent in creating more than five self-report categories during each session of self-measuring his vocabulary knowledge. Moreover, in the last studies (namely, 3 and 4) he arranged the target words into 18 and 19 categories which is far more than included in any existing self-assessment instruments for measuring lexical knowledge. In addition, despite some overlap between the classifications suggested by the subject in different studies (1, 3 and 4), each of the created categorizations included a number of criteria or features that were different from those reflected in the previous studies.

Another rather interesting finding of our research are the links or interconnections between various aspects of word knowledge that the learner referred to in his categorizations. It is worth noting that the subject not only considered the features inherent to his knowledge of the word group he described but also some of the features that his knowledge lacked. This finding might shed new light on how different features of word knowledge are acquired.

It was also highlighted in our previous study that a large portion of once classified words moved into different categories after a lexical input through reading. Furthermore, all the words re-located seemed to improve their original positions in the subject's classification system i.e. acquired new features. There appeared to be no cases of regression (i.e. loss of some of the words' properties) recorded. This finding poses a question of conditions and extent of such word re-locations. Inspired by the findings of the previous study, here, we intended to further address this issue: we hypothesised that some portion of the given words

would change their original positions in the classification system suggested by the subject after she encountered these words in a text.

In this study, we continued exploring the issue of self-rating learners' word knowledge. At this point, we intend to further investigate the issues suggested by the findings reported earlier in this thesis. In this chapter, we report on a study that was carried out to further examine the way/s learners measure and categorize their own knowledge of words. Here, we attempt to explore whether level of proficiency would make a difference to the way learners rate their lexical knowledge. Thus in this study, we tested a new subject with a higher command in Russian.

In this study, we also intended to establish whether reading of an original Russian story would have an impact on the subject's categorization of her knowledge of the targeted words. Furthermore, we sought to explore to which extent the original subject's descriptions of her target word knowledge would change (if any changes occurred) after the primary input through reading. Therefore, the objectives of the study were the following:

- to establish how a different student with a higher level of proficiency in the Russian language would measure her knowledge of the target words;
- to compare subject 1' s (the subject we tested in studies 3 and 4) and subject 2's (the subject of our present study) self-report scales before and after the reading;
- to investigate whether the target words would change their positions in the subject's self-rating scale after the reading.

Thus we arranged a set of objectives similar to those formulated in our previous study in order to investigate whether a different learner would demonstrate similar results in self-rating of her knowledge of words.

The research question: Will the new subject produce data similar to the results submitted by our previous testee?

7.2 Study

7.2.1 Method

7.2.1.1. Target Words

The same set of 200 Russian words used in our previous study was administered in the current research. The words had been extracted from the story "Дама с собачкой» (*A Lady with a Dog*) by Чехов (*Chekhov*). Similar to the previous case study, here we assumed that a great majority of the words selected would be unknown for the subject. And again we attempted to avoid old fashioned, bookish words which are not in use in contemporary Russian. Each word was printed in black ink on a separate white card; 200 word cards in total. The target words represented four parts of speech: nouns, verbs, adjectives and adverbs. These words are shown in table 7-1.

Table 7-1: Target Words

Появиться *to appear*	**Располагать** *to have available*	**Ненависть** *hatred*	**Вихрь** *whirlwind*	**Шорох** *rustle*
Довольный *satisfied*	**Равнодушно** *indifferently*	**Заря** *dawn, sunset*	**Мол** *pier*	**Сытый** *satisfied*
Привыкший*accustomed*	**Приключенне** *adventure*	**Чувство** *feeling*	**Значительный** *significant*	**Поделиться** *to share, to divide*
Душно *stuffy*	**Глухо** *deserted*	**Роса** *dew*	**Испугать** *to frighten*	**Побледнеть** *to become pale*
Рост *growth, height*	**Память** *memory*	**Завять** *to fade/ wither*	**Пристань** *harbour*	**Догадываться** *to guess*
Объяснить *to explain*	**Смех** *laughter*	**Пора** *it's time*	**Промелькнуть** *to flash ,to fly by*	**Грустный** *sad*
Сквер *public garden*	**Забавно** *funny*	**Опуститься** *to go down*	**Отчётливо** *distinctively*	**Унизительный** *humiliating*
Наверное *probably*	**Обстановка** *decor, situation*	**Восхищаться** *to admire*	**Хищный** *predatory*	**Дрожать** *to shiver*
Соображать *to understand , to arrange*	**Занять** *to occupy, to borrow*	**Печально** *sadly*	**Особенность** *peculiarity*	**Пьянство** *alcoholism*
Порядочный *decent*	**Цель** *aim (goal), target*	**Жаловаться** *to complain*	**Упрямый** *stubborn*	**Даль** *distance*
Казаться *to seem*	**Однажды** *one day*	**Уважать** *to respect*	**Приставать** *to pester, to stick*	**Чепуха** *neusance*
Сила *force*	**Жалкий** *pitiful, sorry*	**Тревожно** *anxiously*	**Капризный** *capricious*	**Раскаяние** *repentance*
Солидная *solid, respectable*	**Причёска** *hair style*	**Молчание** *silence*	**Обращаться** *to address, to treat*	**Возмущаться** *to be outraged*
Влечь *to draw, to attract*	**Пыль** *dust*	**Ревность** *jealousy*	**Блестеть** *to shine*	**Испытывать** *to test, to feel*
Мыслящая *thinking*	**Походка** *walk/gait*	**Свеча** *candle*	**Растерянность** *confusion*	**Надоесть** *to pester*
Нерешительный *indecisive*	**Пароход** *steamship*	**Страх** *fear*	**Отрывистый** *jerky/ abrupt*	**Сведения** *information, news*
Втайне *secretly*	**Выражение** *expression*	**Гореть** *to burn*	**Некстати** *at the wrong moment*	**Устроить** *to arrange*
Не спеша *unhurriedly*	**Толпа** *crowd*	**Страстно***passionately*	**Пристально** *intently*	**Смутить** *to embarrass/ confuse*
Считать *to count, to consider*	**Нравы** *manners/ways*	**Перестать** *to stop*	**Унылый** *downcast*	**Свидание** *meeting*
Верить *to believe*	**Случай** *case, event*	**Вдруг** *suddenly*	**Влага** *moisture*	**Нищий** *poor*
Недалёкий *not far off*	**Победа** *victory*	**Слёзы** *tears*	**Приветливый** *friendly*	**Освещать** *to lighten*
Умолять *to plead/beg*	**Поворачиваться** *to turn*	**Скрипка** *violin*	**Пугливо** *fearfully*	**Развлекаться** *to amuse o.s.*
Изящный *graceful*	**Взглянуть** *to cast a glance (at)*	**Презирать** *to despise*	**Насмешка** *gibe*	**Трогательный** *touching/ moving*
Грех *sin*	**Пассажиры** *passengers*	**Водопад** *waterfall*	**Духи** *perfume*	**Вынуждать** *to force*
Бояться *to be afraid*	**Мысль** *thought*	**Обмануть** *to deceive*	**Колокол** *bell*	**Оправдание** *justification, excuse*
Успокоиться *to calm down*	**Нарядный** *elegant*	**Впечатления** *impressions*	**Беззаботный** *carefree , careless*	**Дразнить** *to tease*
Изменять *to change*	**Ласкаво***affectionately*	**Служить** *to serve*	**Жадность** *greed*	**Жечь** *to burn*
Мёртвый *dead*	**Молчать** *to be silent*	**Удаваться** *to succeed*	**Добродушный** *good-natured*	**Досада** *annoyance*
Опыт *experience*	**Нюхать** *to smell/sniff*	**Любопытство** *curiosity*	**Лестно** *flatteringly*	**Прятаться** *to hide o.s.*
Волны *waves*	**Движение** *movement*	**Судьба** *fate/ fortune*	**Искреииость** *sincerity*	**Торопнться** *to be in a hurry*
Горький *bitter*	**Обнять** *to embrace*	**Безумный** *mad, wild*	**Покрыться** *to cover o.s*	**Качаться** *to swing*
Фонарик *torch*	**Рассвет** *dawn*	**Звонок** *bell*	**Неуместный** *inappropriate*	**Смущение** *confusion*
Общество *society*	**Поцеловать** *to kiss*	**Раздражать** *to irritate /to annoy*	**Ужас** *horror*	**Бестолково** *disorderly*
Церковь *church*	**Очарованный** *fascinated*	**Исчезнуть** *to vanish*	**Обморок** *faint*	**Мрачный** *gloomy*
Скучно *bored*	**Запах** *smell*	**Шутить** *to joke*		
Неподвижно *motionlessly*	**Сказочный** *fairytale*	**Крик** *scream*		
Чувствовать *to feel*	**Желание** *desire/wish*	**Публика** *public*		
Кричать *to shout*	**Сторож** *guard*	**Воспоминание** *recollection*		
Привлекательный *attractive*	**Властный** *imperious*	**Пахнуть** *to smell*		
Покой *rest/ peace*	**Подробности** *detail*	**Следить** *to watch, to follow, to look after*		
Страсть *passion*	**Шевелиться** *to stir*	**Тень** *shadow, shade*		
Метель *blizzard*	**Окурок** *cigar-butt*	**Вечный** *eternal*		
Воскресать *to revive*	**Сутулый** *stooping*	**Настраивать** *to tune*		
Воображение *imagination*		**Кланяться** *to bow*		
Дрянной *worthless*				

7.2.1.2. Participant

The participant in this study was a French female student in her early twenties (further referred as Subject 2), who had been studying Russian at a University in France for four years. Her native language is French. Russian is her second foreign language after English. She is also fluent in Spanish. Subject 2 was a student of mine at Cardiff as an exchange student between French and British Universities. She had been attending my Advanced Russian at Cardiff for 8 months (24 meetings, 2 hours a week), and demonstrated very good knowledge of Russian vocabulary. Overall, Subject 2 showed a higher level of proficiency in the Russian language than our subject (further referred as subject 1) in studies 3 (chapter 5) and 4 (chapter 6).

7.2.1.3. Procedure

The procedure consisted of two stages. Each stage included two steps.

Stage 1

Step 1a) Our subject was given the story "A Lady with a Dog" (*Дама с собачкой*) by Chekhov (*Чехов*) to read in English translation at home.

Step 1b) Then, in the classroom, we asked the participant to retell the story in order to check whether she had read it at home. After that, the main part of the study began. The subject received 200 cards in a random pile. Each card contained one target word printed on it. The instruction was: "Arrange the following cards into categories according to how well you know the word on the card". We encouraged the participant to describe her knowledge of the given words through self-report categories. The subject did not receive any prompts. No time restriction applied. After the subject had divided all the target words into categories, she explained the reasons for the classification system she had created. The explanation was recorded.

Stage 2

Step 2a) The participant was required to read "Дама с собачкой» by Чехов in the original at home. The target words were not highlighted in the text. The subject was not advised to use reference literature.

Step 2b) After this task had been completed at home, step 1b) was repeated in full in the classroom.

The subject's explanations for the created categories were recorded.

7.2.2 Results

Pursuing the first objective of our study we asked the participant to self-assess her knowledge of the certain words and categorize those words according to how well she thought she knew each of the given words. ***Stage 1:*** At the first stage of the study, the subject divided 200 words into 15 categories. They are shown in Table 7-2.

Table 7-2: Categories suggested by the subject at stage 1

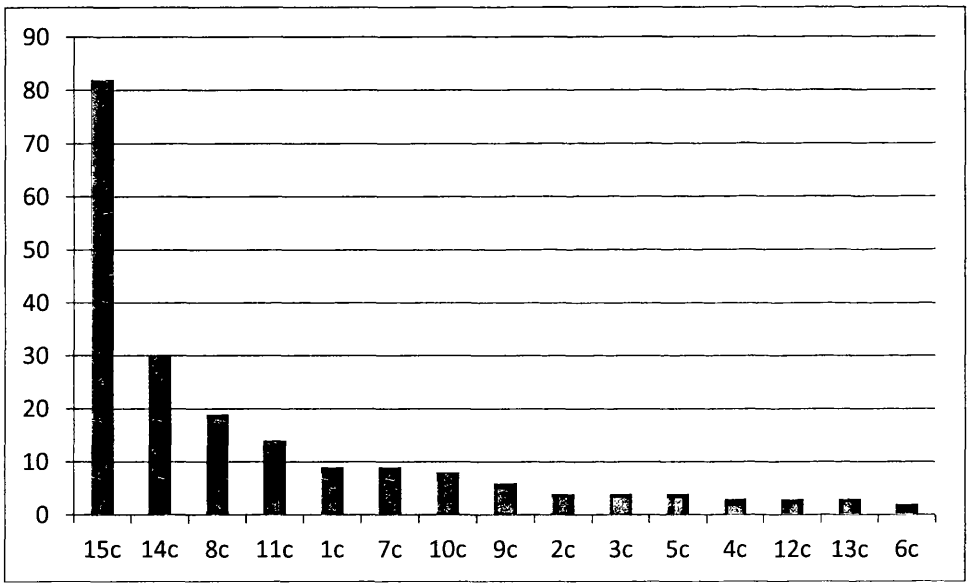
Category Description	No of Words
1: Nouns known, can be used in speech:	9
2: Verbs known well. I can recognize them at the first sight and use them easily (conjugate etc.):	4
3: Adjectives known; can be easily used in speech:	4
4: Adverbs known:	3
5: Verbs known but I need to think how to use them:	4
6: Adjectives. The meaning is known but I am not sure about the use of them (which context, what to describe with them):	2
7: I cannot remember the meaning(s) of these words but I can give some grammatical information about them:	9
8: Nouns -unknown. I can determine the gender; can be declined:	19
9: Nouns -unknown; can be spelled correctly if they are dictated:	6
10: Adjectives- unknown. I do not know the meaning(s) but I can decline them according to a noun:	8
11: Verbs- unknown but can be conjugated:	14
12: Verbs unknown but I think I can guess the meaning:	3
13: Adjectives unknown but I can think about other words (nouns with the same root, for example) when I read them:	3
14: Verbs totally unknown:	30
15: Words which are not familiar at all:	82

Note. Self-report categories are listed in the order arranged and presented by the subject

The table shows that a rather large number of categories, namely 15, were proposed by the

participant at the first stage of the study. And again, the number of self-rate categories suggested by the learner is substantially greater than included in Paribakht and Wesche's Vocabulary Knowledge Scale. The majority of the words fall into categories 8, 11, 14 and 15 – unknown. The second largest group of categories includes categories 1, 7, 9 and 10 – known and partially known. As seen from the table, there are no one word categories in this classification. Yet the smallest category, No 6, includes two target words. The table also indicates that the created classification system strongly signals the criterion of word class belonging. Figure 7-1 shows graphically how the target words were arranged into the suggested categories.

Figure 7-1: Distribution of words among the suggested categories in categorization 1



Note. Horizontal axis shows the created categories: e.g. cat.15, cat.14, cat.8 etc. Vertical axis illustrates the target words.

Figure 7-1 shows the number of words allocated under each category. Most of the words apparently fall into category 15. The next block of categories seems to include categories 14, 8 and 11 with more than 10 words each. The remaining categories: 1, 7, 10, 9, 2, 3, 5, 4, 12, 13 and 6 contain less than 10 target words each.

Stage 2: At the second stage of the study, after the reading, the 200 targeted words were divided into 26 categories. They are presented in Table 7-3.

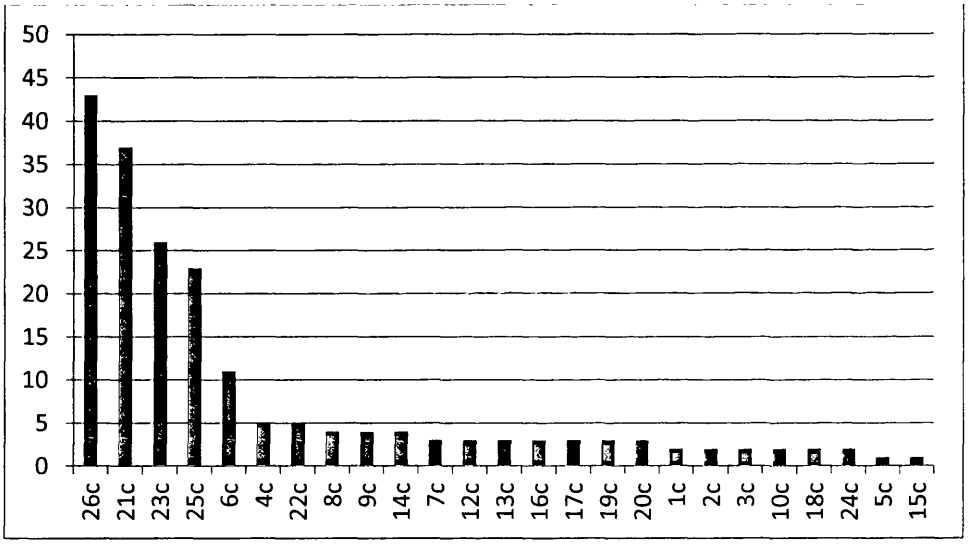
Table 7-3: Categories suggested by the subject at stage 2

Category Description	No of Words
1: Words with 2 meanings that I know:	2
2: Words I know the meaning but I am not able to analyse them grammatically:	2
3: Adverbs I know well:	2
4: Verbs I know well and can use with confidence:	5
5: Verbs I know but I am not confident about using them in speech or writing. However I can recognize them in listening and reading:	1
6: Nouns I know well:	11
7: Words that I can guess the meaning of looking at their morphological roots:	3
8: Words I can guess the meaning thanks to their similarities with French words:	4
9: Verbs I know. I can find the associated perfective or imperfective forms:	4
10: Words meaning of which could be judged on their prefixes (opposites):	2
11: Nouns which I know the meanings of and I can find other words (verbs, adjectives etc) with the same roots:	3
12: Verbs which I know the meanings of and I can find other words (adjectives, nouns etc) with the same roots:	3
13: Adjectives which I know the meanings of and I can find other words (verbs, nouns etc) with the same roots:	3
14: I know these words well; can be used in speech and writing; can be recognized in listening and reading:	4
15: Words which I know the meanings of but I am not sure about the writing (how to spell them properly):	1
16: Words which I know but I am not sure how to decline them because of their ending:	3
17: Adverbs - joining words. I know them and can use them:	3
18: Words with the same root. I am not sure about the meanings:	2
19: Words which look familiar to me but I cannot use them:	3
20: Words I do not know but I would be able to spell them if I hear them:	3
21: Verbs I do not know but I can provide grammar information about them:	37
22: Verbs I do not know and I would not be able to use them grammatically:	5
23: Nouns I do not know. However, I would be able to provide some grammar information:	26
24: Adverbs I do not know:	2
25: Adjectives I do not know but I could use them with a noun I know. I would know how to decline them:	23
26: Words I do not know at all and I cannot provide any information about them:	43

Note. Self-report categories are listed in the order arranged and presented by the subject.

As shown in the table, after the reading, the learner rated her knowledge of words by creating an unusually large number of categories, namely 26. This is considerably more than we expected, basing our prediction on the results obtained so far. The table indicates that there are a lot of categories containing small portions of words, namely 1 to 4. We can also see that the categories containing known or partially known words clearly increased in number. Similar to the first classification, the word class feature seems to be one of the most prominent criteria for this classification. This issue will be addressed in the discussion section. Figure 7-2 illustrates graphically how the words were spread among the created categories at the second stage of the study.

Figure 7-2:Distribution of targeted words among the created categories in classification2



Note. Horizontal axis shows the categories. Vertical axis illustrates the target words.

As seen from Figure 7-2, categories 26 and 21 which contain more than thirty targeted words each can be combined in Block 1. Block 2 would include categories 23 and 25 and 6 with more than 10 words each. Block 3 would contain categories 4, 22, 8 onwards with 5 or fewer target words in them.

Thus, as seen from the tables and figures illustrated above, the number of self-rate categories created in the current study is considerably larger than those suggested in our previous research. The results presented in this section and their implications will be discussed in the section that follows.

7.3 Discussion

In this section, we seek to find an answer to the research question posed in this study: Will the new subject with completely different characteristics produce the data similar to the results submitted by our previous testee? The structure of this section reflects the objectives outlined in the first section of this study. First, we will discuss the ways our new subject measures and ranks her knowledge of the target words. Then, we will compare the classification systems created by this subject (Subject 2) against those suggested in our previous studies by Subject 1. Finally, we will analyse the relocations of words between the self-rate categories within the categorization scales developed by the subject.

In general, the data collected in this study substantiate the findings of our previous research regarding the complexity of the categorization systems produced by a learner while self-assessing their knowledge of words. And again, these findings contradict the existing approaches to self-assessment of vocabulary knowledge. Another finding of our previous research: word relocation among the self-rate categories, was also recorded in the current study. Let us take a look at each of these issues in turn:

7.3.1 Self- categorization of the subject's word knowledge before and after the reading

1. The first objective of our study was to explore how the new subject would classify her word knowledge before and after the reading. The participant of our current study (Subject 2) arranged the given words into the largest number of categories created in our research so far. At the first stage, she managed to produce 15 self-report categories which rose to 26 – at the second stage of the study. Such fruitful classifications totally exceeded our expectations since the previous largest categorization suggested by subject 1 in study 4 (Chapter 6) included 19 categories.

The implications of these findings are rather important. As we noted in our previous studies, notwithstanding their limitations, the results obtained so far clearly indicate that our testees measure their lexical knowledge by creating rather complex self-assessment scales. Such scales are in contrast with the self-measurement instruments developed in the field so far. Moreover, our research is pioneering in exploring the ways learners assess and classify their own knowledge of words.

Comparing the two classification systems suggested by our learner before and after the reading we can note that, similar to the data obtained in our previous study, the second categorization appears to be more complex in terms of its quantitative and qualitative features. Despite a large number of categories created at the first stage of the study, namely 15, only a small portion of words, namely 26, allocated in categories 1 through 6, were declared as known in the first categorization. This means that the remaining 9 groups contain the words that the learner was unable to translate though she still categorized those words according to some other criteria. This, once again, substantiates the necessity of measuring a broad variety of word properties while self-assessing knowledge of a word. And again, this stands in contrast with the current approaches to self-assessment in L2 vocabulary acquisition and testing.

The impressive number of word groups included in the second categorization, namely 26, shows a substantial increase in the categories containing known and well-known words (14 vs. 6) as well as "guessing" categories (3 vs. 1). Thus the original classification becomes more complex since known as well as earlier unknown words are now split into more categories on the grounds of other features (besides the meaning knowledge) being acquired through reading, e.g. multiple meaning. In addition, completely new categories, based on new aspects of word knowledge emerged in the new classification e.g. morphological knowledge: cat.7: "Words that I can guess the meaning of looking at their morphological roots". A comparison between the qualitative characteristics of the two systems can be carried out on data illustrated in table 7-4.

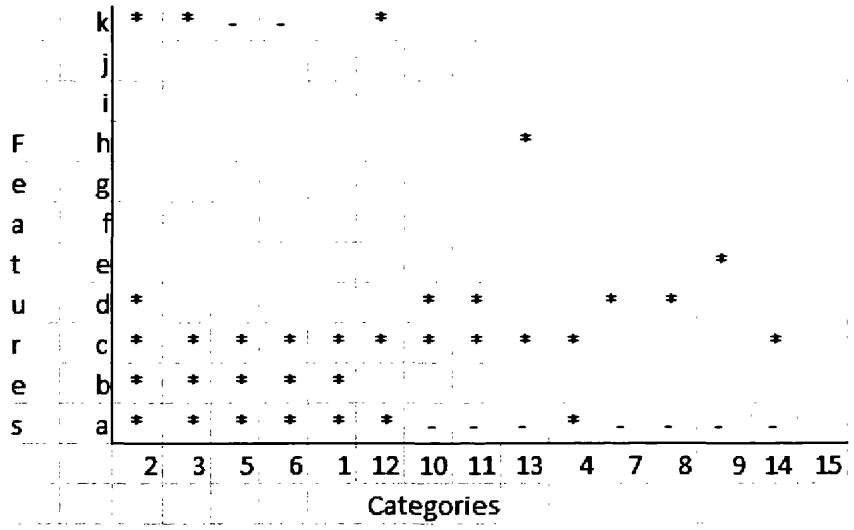
Table 7-4: Distribution of features among the self-report categories at stages 1 and 2

Criteria/Features	Exp 5 stage 1		Exp 5 stage 2	
	+	-	+	-
Meaning	7	7	17	6
Use	5		3	1
Word Class	13		14	
Grammar	5		5	2
Written Form	1		2	
Sound Form			1	
Morphology			3	
Derivatives	1		3	
MultipleMeaning			1	
Four Skills			2	
Certainty incl. guessing	3	2	5	4

+ indicates that the feature was mentioned in a positive context; - shows that the feature was mentioned in a negative context.

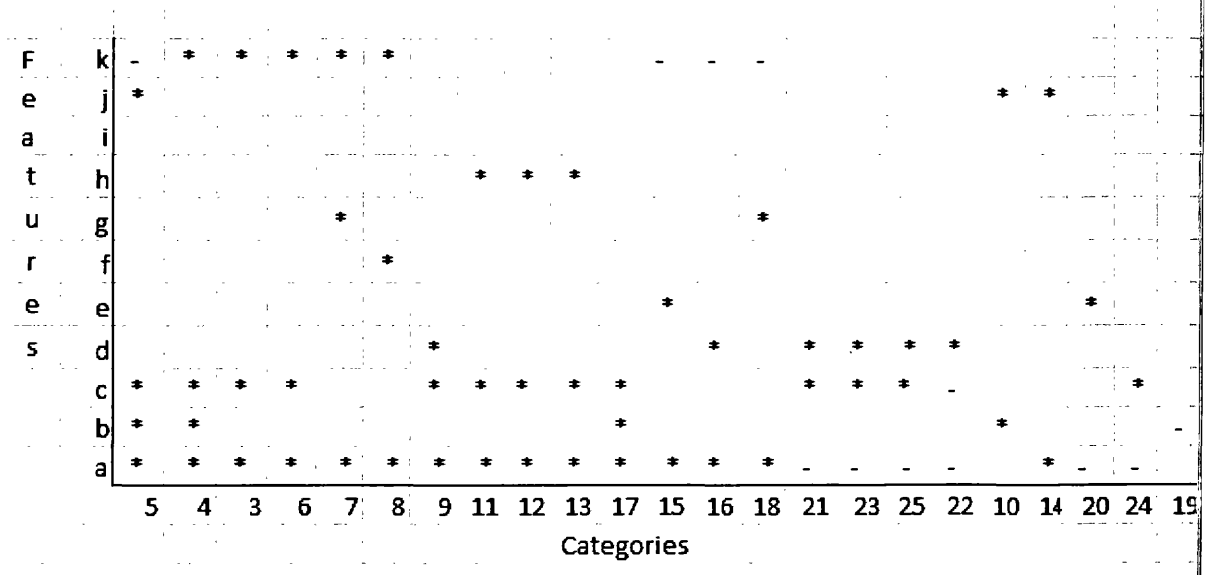
The data shows that the subject of the present study, like our previous participants, referred to a large variety of features while describing her knowledge of the given words, at both stages of this experimental task. At the same time, the data illustrated in the table indicates that at the second stage of the study, the subject considered rather more features (additional four) than she referred to in her first scale. They are: phonological knowledge, multiple meanings, morphological knowledge and productive vs. receptive skills. Clearly, the principle criteria in her categorizations are meaning knowledge and word class belonging. This, once again, confirms the importance of word class factor for L2 learners of Russian. Grammar information, use and certainty in knowledge were also frequently used as category descriptors in both scales. It is instructive to see how the aforementioned features were spread among the created categories. This is shown in Figures 7-3 and 7-4.

Figure 7-3: Distribution of features within each of the suggested categories in classification 1



Note. Features: a-meaning; b-use; c-word class; d-grammar info; e-written form; f-sound for g-morphological info; h-derivation forms ;i-multi-meaning; j-the four skills; k-certainty

Figure 7-4: Distribution of features within each of the suggested categories in classification 2



Note. Features: a-meaning; b-use; c-word class; d-grammar info; e-written form; f-sound form; g-morphological info; h-derivation forms ;i-multi-meaning; j-the four skills, k-certainty

When the two figures are compared the results reveal similarities between the two scales. Indeed, descriptions for most categories in both classification scales include more than two features. The maximum number of descriptors used to characterize one category (cat. 2/ clas.1 and cat. 5/ clas.2) in each scale was 5. Furthermore, a large proportion of the category descriptions in both systems are based on three and more features which were used both in positive and negative (i.e. absence of feature) context. These results are promising since they support the idea of interaction between various features developed from our previous data. Such links, despite their obvious importance, were not detected in other research on L2 lexical self-assessment.

Overall, the current data supports the findings of our previous studies concerning the high level of complexity of subjects' categorizations of their lexical knowledge. It might be worth-while pointing out that Paribakht and Wesche's methodology actually prevents this type of data from emerging. It not only assumes that five categories are sufficient, but also assumes that the categories are optimal and doesn't allow for finer categorisation appearing as a result of reading.

7.3.2 Comparative analysis of the categorizations submitted by subject 1 and subject 2 before and after the reading

The second objective of the present study was to carry out a comparison between the self-rate classification systems created by two different learners in studies 4 and 5. In order to fulfill this task, we attempt to compare quantitative and qualitative features of the two sets of self-rank categorizations. The summarized quantitative data is presented in table 7-5.

Table 7-5: Number of self-report categories suggested by the two subjects

	Study 4/1 (Subj 1)	Study 4/2 (Subj 1)	Study 5/1 (Subj 2)	Study 5/2 (Subj 2)
No of Categories	10	19	15	26

The table indicates that the second set of categorizations, namely 5/1 and 5/2, compiled by Subject 2, contains considerably more categories than the classification systems produced by Subject 1 in the previous study. This might suggest that Subject 2 measures her knowledge of

words in a more complex, probably more detailed way than Subject 1. This may also reflect Subject 2's individual style of classifying her vocabulary knowledge. The former point will be further explored by comparing the qualitative features of the subjects' classification systems. In regards to the latter idea, it does not seem possible to further investigate that assumption on a limited number of subjects.

In general, the data indicates a consistent increase in amount of categories after the reading. This may suggest that known words are categorized in a more detailed way than unfamiliar ones.

The findings of the present study confirm that a new lexical self-assessment methodology should be developed. This methodology would benefit from considering a variety of features revealed in the self-assessment systems built by the learners in our research. Such summarized features are shown in table 5-6 which also illustrates the distribution of these features among the self-rank categories in both rounds, before and after the reading, in studies 4 and 5.

Table 7-6: Criteria/Feature distribution among the self-report categories in studies 4 and 5

Features	Exp 4/1		Exp 4/2		Exp 5/1		Exp 5/2	
	+	-	+	-	+	-	+	-
Meaning	5	5	13	1	7	7	17	6
Use	1	1	5		5		3	1
Word Class	4		12		13		14	
Grammar	1		1		5		5	2
Written Form					1		2	
Sound Form	1		2				1	
Morphology	1		3				3	
Derivatives					1		3	
Multiple Meaning			1				1	
Four Skills							2	
Certainty incl. guessing	3	2	2	7	3	2	5	4

+ indicates that the feature was mentioned in a positive context; - shows that the feature was mentioned in a negative context.

A comparison of the presented data reveals some similarities as well as differences between the subjects' classifications. Regarding the similarities, it is noteworthy that both subjects used a rather large number of classification criteria in both rounds of the relevant studies. The main features referred to by both subjects are meaning knowledge, word class belonging and use. Both subjects indicated their certainty in knowledge as well as broadly using guessing. In general, they share five descriptors in their first classifications and eight descriptors in the second systems. As seen from the table, both learners address a number of features in a negative context which implies that they self-assess not only the existing properties of their word knowledge but also non-existing features. This once again indicates that Paribakht and Wesche's scheme is inadequately detailed. From the present evidence though obtained in case studies, we can conclude the necessity of a multiple- feature self-assessment scale to measure a larger set of features highlighted by our subjects. The obvious conclusion here is that Wesche and Paribakht's model fails to register the enormous complexity of self-assessment categorizations that is emerging from these detailed case studies.

It is also worth while noting that in both sets, the second categorizations, produced after the reading, appear to be more complex than those created at the first stage. This complexity relates to both quantitative and qualitative features of the second categorizations and, presumably, implies the developing complexity of the knowledge of certain words.

Despite such similarities between the two sets of classifications, Subject 2's categorizations, produced before and after the reading, are, clearly, richer in terms of descriptors used by the learners. In fact, she considered eleven different features while describing her target word knowledge after the reading compared to the eight used by Subject 1. These differences relate to spelling knowledge, knowledge of derivation forms as well as the four language skills. It is also worth noting that, though both subjects used the equal amount of descriptors, namely 7, in their original (stage 1) categorizations, the differences between them lie in the types of knowledge they referred to. Thus, besides the common descriptors, Subject 1 engaged phonological and morphological aspects of word knowledge whereas Subject 2 used spelling and derivation kinds of lexical knowledge. Furthermore, as seen from table 7-6, Subject 2 refers to the grammar feature considerably more than the other participant. Again, this might be explained by the learners' individual styles of word knowledge self-rating which, in turn, might reflect their individual ways of vocabulary acquisition.

The implication of the sets of categorizations being different might be that Subject 2 acquires new words and develops her existing vocabulary knowledge in a more complex and structured way compared to Subject 1. This finding appears to be in contrast with general assumptions in the field that the process of vocabulary acquisition is the same in learners at different levels of knowledge. Our argument is that a single set of categories within a self-assessment scale might not be the best way of describing vocabulary knowledge. Different learners might need different frameworks. However, at the same time, we need to bear in mind that some learners may categorize certain words as special on the grounds other than their level of knowledge: e.g. an unusual letter sequence. This suggests the importance of learners' detailed explanations of the classification criteria they use.

7.3.3 The issue of word re-locations within the self-rate categorization after the reading

The third objective of this study addresses the matter of word re-locations. This issue emerged in the previous study as a rather important finding of our research. In our earlier experimental tasks, we were primarily concerned with the way/s learners self-assesses their L2 word knowledge. However, the data received in study 4 (chapter 6) triggered the idea of further exploring the issue of word movements among the self-report categories. The previous study revealed that 60 targeted words improved their original positions within Subject 1's classification systems after the reading. In light of this finding we were curious to see what kind of changes in the positions of targeted words would occur after the reading in the present study. The evidence we obtained here indicates that a rather large number of target words, namely 97, re-located into different categories after the reading. This is considerably more than recorded previously. Table 7-7 illustrates the summarized results of these relocations.

Table 7-7: Summary of word re-locations after the reading

Status of Words	Improve	Stay the same	Regress
No of Words	94	103	3

As illustrated in the table, 94 target words upgraded (i.e. moved into the categories with a larger number of features attached to them) their initial categories after the reading. There is no evidence of strong regression in the subject's target word knowledge, for instance: moves from "known" to "unknown". However, at the second stage of the study, three words re-

located from the categories: "unknown meanings but can provide some grammar information" (Classification 1) to "totally unknown" (Classification 2).

In general, this data confirms our earlier finding that learners acquire L2 vocabulary knowledge in a different way. The re-location of 94 words compared to 60 in the previous study, out of the same 200 targeted words, suggests that the processes of lexical acquisition occur at a different rate in learners at different levels. However, it should be noted that this finding requires further exploration on a larger set of data. Table 7-8 shows the word re-locations identified in this study in more detail.

Table 7-8: Word re-locations after the reading

No of Category Classification 1	Relocated to Category Classification 2	No of Words
3	1; 14	1; 2
5	4; 9; 12	1; 2; 1
6	14	2
7	3; 6; 26	1; 1; 2
8	6; 26	1; 1
9	2; 8; 11	1; 1; 1
11	4	2
12	12	2
13	13	2
14	5; 21	1; 25
15	1; 2; 3; 6; 7; 8; 9; 10; 11; 13; 16; 18; 19; 20; 23; 25	1; 1; 1; 3; 3; 3; 1; 2; 1; 1; 2; 1; 1; 1; 10; 15

The data shown in the table indicates that encountering target words in a text led to significant shifts of vocabulary items among self-report categories in the subject's self-rating scale. As seen from the table the largest number of movements of words occurred from category 15, i.e. 47 words re-located into 16 different categories within the second classification. It is apparent

that the gains in the subject's vocabulary knowledge were achieved due to the primary lexical input through reading. The nature of these relocations is of particular interest. Different words initially located in the same category seem to move in different directions after the reading. Let's take a look at Figure 7-5.

Figure 7-5: Word-relocations from category 9 in classification 1

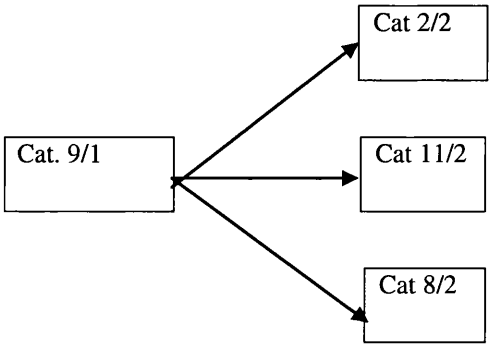


Figure 7-5 illustrates the word relocations from category 9 "unknown nouns" after the reading. As seen from this figure, some words moved straight into category 2 "known well" in classification 2. Others moved into different categories while the learner acquired other types of information about the words: Cat 8: "phonological knowledge" and Cat 11: "derivation forms". At this stage, we cannot yet determine the main patterns of word movements. This needs to be traced in a repeated longitudinal study we will discuss in the next chapter.

The data seem to indicate that the vocabulary acquisition process does not always occur as a linear transition from "known" to "unknown". In fact, in many cases, it appears to be more complex and varied than commonly assumed in the field. This data poses a question: "In which direction will the target words move from there?" This will be investigated in our next study.

In summary, this study showed evidence that the initial categorization was re-structured and new self-report categories emerged after the reading activity had taken place. The elicited data clearly indicate substantial shifts of lexical items within the suggested categories. The question is: "What can this be attributed to?" This may simply be due to the fact that learners re-assess their vocabulary knowledge from scratch after a certain vocabulary input has

occurred. However, at this stage, we need to admit that further research into this issue is needed. The question that must be asked at this point is: Will subsequent sessions of reading lead to even greater changes in the subject's categorization of their knowledge of the target words? This is the issue to be investigated in our next study.

7.4 Conclusions

In this study, we continued investigating the way/s L2 learners measure and categorize their own knowledge of words. We found that the learner with a better command of Russian classified her lexical knowledge in a different, more complex way than the participant of our previous studies. These findings suggest that further investigation of this issue might provide evidence on individual approaches in lexical self-assessment.

The findings have also provided confirmation regarding changes in the positions of words within classification systems after reading. In light of these findings, we decided to test these results on a larger set of reading sessions. Thus we intend to carry out a repeated longitudinal study in order to investigate some of the aforementioned points.

In the next study, we will investigate whether another learner at a rather higher level of L2 knowledge would rate her knowledge of words in a more complex way than our previous subjects. We will further explore the issue of word-relocations: we will attempt to establish whether subsequent sessions of reading will lead to further changes in the subject's self-rating scales.

CHAPTER EIGHT

Self-Rating of Word-Knowledge: before and after Reading— a Repeated Longitudinal Study

Experimental Study 6

8.1 Introduction

In the previous study, we examined how a learner with a higher (compared to our previous subject) proficiency in Russian, a French native speaker, would self-classify her knowledge of words. It emerged that the classification systems produced by that participant (tested in our previous study) were more complex quantitatively (i.e. the number of the suggested categories) than those created by the previous testee. Furthermore, despite some overlap in the criteria used by subject 1 (study 4) and subject 2 (study 5), the latter's categorizations proved to be also far richer in terms of their qualitative characteristics (i.e. the features or word properties the learner referred to). Furthermore, the data collected in our earlier research revealed that the second categorizations created by the subjects after the reading differed considerably from those submitted initially, at the first stages of the relevant studies. In light of this finding, logic suggests that further categorizations that will be produced after subsequent sessions of reading might reflect even greater changes (gains or losses) in learners' knowledge of particular words. This rather important issue has never been investigated via self-assessment before and will require longitudinal self-measurements.

In our previous study, we also investigated whether the positions of words within the subject's categorizations would change after a lexical input through reading. The results indicated that a reading session did alter the locations of certain targeted words. Moreover, the amount of words re-located from their initial categories appeared to be rather large. Ninety four targeted words moved into different categories after the reading. At this point, we were curious to establish whether further exposure to the targeted words via successful sessions of reading would lead to further re-locations among those words. For this purpose, a repeated longitudinal study was administered as the next step in our research.

It is also worth noting that the previous study revealed 3 (out of 94) instances of slight regression in the subject's word knowledge. These examples, though few in number, posed an

intriguing question: How common are cases of regression within the process of lexical acquisition? In order to find an answer to this question we needed to trace possible instances of regress in the learner's word knowledge by means of repeated self-assessments. This was another reason for switching to a longitudinal study.

Thus in this chapter, we will report on a repeated longitudinal study undertaken to investigate the points outlined above. Briefly, these points are: the issue of self-categorization and the issue of word re-locations between the self-classification systems. We intended to explore whether another subject with even higher command in Russian (than the previous two testees) would measure her knowledge of 200 Russian words by arranging those words into a complex classification system. On the grounds of our previous findings, we assumed that the participant of our current study would classify her knowledge of words in a more complex and varied way than the previous testees. We will also report on changes in the positions of words in the learner's classifications. By addressing these points our purpose was to further explore (during a longer period of repeated self- observations) issues that emerged in our earlier studies as well as to check the important findings of our research.

Thus the objectives of the current study were as follows:

1. To investigate how a different student with a higher level of proficiency in Russian than our previous subjects will describe her knowledge of the given words.
2. To establish whether the subject's categorization of her word knowledge will change after each of the subsequent sessions of reading.
3. To measure quantitative and qualitative changes in the subject's word knowledge after each reading session.
4. To explore movements of words among suggested self-classification scales.

The research question was:

If a learner is regularly exposed to certain words, for instance through reading, will the description of their own knowledge of those words change?

8.2 Study

8.2.1 Method

8.2.1.1Target Words

A new set of 200 Russian words was selected for this study. The words had been extracted from another story by Chekhov "Крыжовник». Taking into consideration the subject's high command in Russian we chose words most of which we believed would be unfamiliar to the subject. However, in order to sustain the subject's interest in our study a number of easier words were also included. We also avoided old fashioned, bookish words which are not in use in contemporary Russian. Each word was printed on a separate white paper card in black ink; 200 word cards in total. The target words represented four parts of speech: nouns, verbs, adjectives and adverbs. These words are shown in Table 8-1.

Table 8-1: Target Words

Обложить *to surround*	**Раздумье** *in thought*	**Воображаемый** *imaginary*	**Случай** *case, event*	**Лицемерие** *hypocrisy*
Скучно *to be born*	**Дворянство** *nobility*	**Виноват** *guilty*	**Жёстко** *hard (adv)*	**Когти** *claws*
Нависать *to overhang*	**Хвост** *tail*	**Чудак** *strange man*	**Вредный** *harmful*	**Беда** *trouble*
Поле *field*	**Смерть** *death*	**Наглое** *impudent*	**Бельё** *linen, underwear*	**Постепенно** *gradually*
Ветеринарный *veterinary*	**Долг** *debt*	**Купец** *merchant*	**Очевидно** *obviously*	**Потери** *losses*
Утомиться *to become tired*	**Сочувствовать** *to sympathise*	**Стучать** *to knock*	**Враньё** *lie*	**Доказательства** *proofs*
Бесконечный *endless*	**Стеречь** *to guard*	**Кровь** *blood*	**Достичь** *to achieve, to reach*	**Мелкие** *small, shallow , fine*
Мельницы *mills*	**Борьба** *fight*	**Ошибаться** *to make mistakes*	**Тишина** *silence*	**Ссылаться** *to refer*
Село *village*	**Ловить** *to try to catch*	**Беспокоиться** *to worry*	**Цель** *aim, target*	**Слегка** *slightly*
Тянуться *to stretch, to extend*	**Шум** *noise*	**Печалиться** *to be sad*	**Спокойствие** *quiet, order*	**Естественный** *natural*
Задумчивый *thoughtful*	**Поймать** *to catch*	**Пропасть** *to disappear, to be lost*	**Судьба** *fate*	**Благополучно** *safely, happily*
Кроткий *meek*	**Эгоизм** *selfishness*	**Проведать** *to come to see, to find out*	**Громко** *loudly*	**Законность** *lawfulness*
Проникнут *imbued*	**Стая** *flock, pack*	**Заботнться** *to take care (of)*	**Доволен** *pleased*	**Охота** *hunt , wish*
Сарай *shed*	**Лень** *laziness*	**Изгородь** *hedge*	**Возмутиться** *to be outraged*	**Явление** *phenomenon*
Староста *head, boss*	**Носиться** *it's in the air, wear*	**Душа** *soul*	**Отчаяние** *despair*	**Управлять** *to manage, to control*
Собираться *to gather together*	**Подвиг** *heroic deed*	**Грустный** *sad*	**Кладбище** *cemetery*	**Перескочить** *to jump over*
Вздохнуть *to sigh*	**Тосковать** *to long (for)*	**Важность** *importance*	**Подавляющий** *suppressing, depressing*	**Необходимо** *necessarily*
Предвидеть *to foresee*	**Простор** *space*	**Мысль** *thought*	**Покойник** *the deceased*	**Невыносимо** *unbearably*
Лить *to pour*	**Желание** *desire/wish*	**Пьяный** *drunk*	**Плотина** *dam*	**Образование** *education*
Исчезать *to vanish*	**Свойства** *attributes*	**Робкий** *shy*	**Протестовать** *to protest*	**Угнетать** *to suppress*
Крыжовник *gooseberries*	**Особенность** *peculiarity*	**Кланяться** *to bow*	**Невежество** *rudeness*	**Сердито** *angrily*
Холм *hill*	**Запереть** *to lock*	**Прежний** *former/ ex*	**Порядок** *order,way*	**Зрелнще** *show*
Луга *meadows*	**Угол** *corner, angle*	**Перемена** *break, change*	**Бедность** *poverty*	**Осуществляться** *to come true*
Ивы *willows*	**Привыкиуть** *to get used to*	**Баня** *Russian bath*	**Бремя** *burden*	**Ненавидеть** *to hate*
Громадное *huge*	**Удобства** *conveniences*	**Судиться** *to sue*	**Умоляющий** *pleading*	**Пахнуть** *to smell*
Ползущий *crawling*	**Простить** *to forgive*	**Полезный** *useful*	**Проститься** *to say good-bye*	**Удовлетворить** *to satisfy*
Гусеница *caterpillar*	**Нищий** *poor man*	**Обижаться** *to take offence*	**Бодр** *cheerful*	**Сумерки** *twilight*
Природа *nature*	**Укрыться** *to hide, to cover o.s.*	**Обращаться** *to address*	**Украшения** *ornaments, jewellery*	**Скорбить** *to grieve*
Казаться *to seem*	**Жадничать** *to be greedy*	**Развивать** *to develop*	**Уставать** *to get tired*	**Изящный** *elegant*
Дух *spirit*	**Лавочка** *bench*	**Улыбка** *smile*	**Резные** *carved*	**Раздражаться** *to get irritated*
Усыплять *to put to sleep*	**Праздник** *holiday*	**Сытость** *satiety*	**Смысл** *meaning*	**Хозяйство***housekeeping, farm*
Бедняга *poor thing*	**Запах** *smell*	**Слеза** *tear*	**Распятие** *cross*	**Гореть** *to burn*
Трубочка *tube/pipe*	**Прятать** *to hide*	**Праздность** *idleness*	**Слоновая** *elephantine*	**Умно** *cleverly*
Всякие *any*	**Совет** *advice, council*	**Молча** *silently*	**Жалкий** *pitiful*	**Голос** *voice*
Чехол *cover, case*	**Копить** *to accumulate/to store*	**Самомнение** *conceit*	**Постель** *bed*	**Вникать** *to understand*
Рамы *frames*	**Радость** *joy*	**Волнение** *worry*	**Ступенька** *step*	**Стремиться** *to rush, to strive(for), to want to go*
Грешный *sinful*	**Вдова** *widower*	**Взгляд** *look, view*	**Признаваться** *to confess*	**Вдоволь** *in abundance*
Чин *rank*	**Объявление** *advert*	**Ягода** *berry*	**Дно** *bottom*	**Кисло** *sourly*
Воля *will, freedom*	**Впроголодь** *half-starving*	**Рисовать** *to draw, to depict*	**Телега** *cart*	**Крестьянские** *peasant (adj)*
Труп *corpse*	**Торжество** *celebration, triumph*		**Умиление** *emotion*	**Наказание** *punishment*
Истина *truth*				

8.2.1.2 Participant

In this study, we tested a new subject, a young professional who had recently graduated from the Russian department of Nottingham University, UK. Out of all the subjects tested in our research, this female student possessed the highest proficiency in the Russian language. A native English speaker, she is also fluent in French (her third language after English and Russian). She had also been attending my Advanced Russian at Cardiff University for 8 months prior to the assessment. In regards to her vocabulary knowledge, the subject persistently demonstrated an excellent command in Russian vocabulary for the level she was attending.

8.2.1.3 Procedure

The procedure of this repeated longitudinal study consisted of five stages. Each stage included two steps. There was a one week interval between each stage.

Stage 1

Step 1. The subject was offered the story "Gooseberries" (Крыжовник) by Chekhov (Чехов) to read in English translation at home.

Step 2. Then, in the classroom, as in studies No 3, 4 and 5, the subject received 200 cards in a random pile. Each card contained one target word printed on it. The instruction was as follows: "Arrange the following cards into categories according to how well you know the word on the card". The subject did not receive any prompts from the experimenter. No time restriction applied. Having arranged all the target words into categories, the subject explained her reasons for the classification produced. The explanation was recorded.

Stage 2

Step 1. The participant was asked to read "Крыжовник» by Чехов in the original at home. The target words were not highlighted in the story in any way. The subject was not advised to use reference literature.

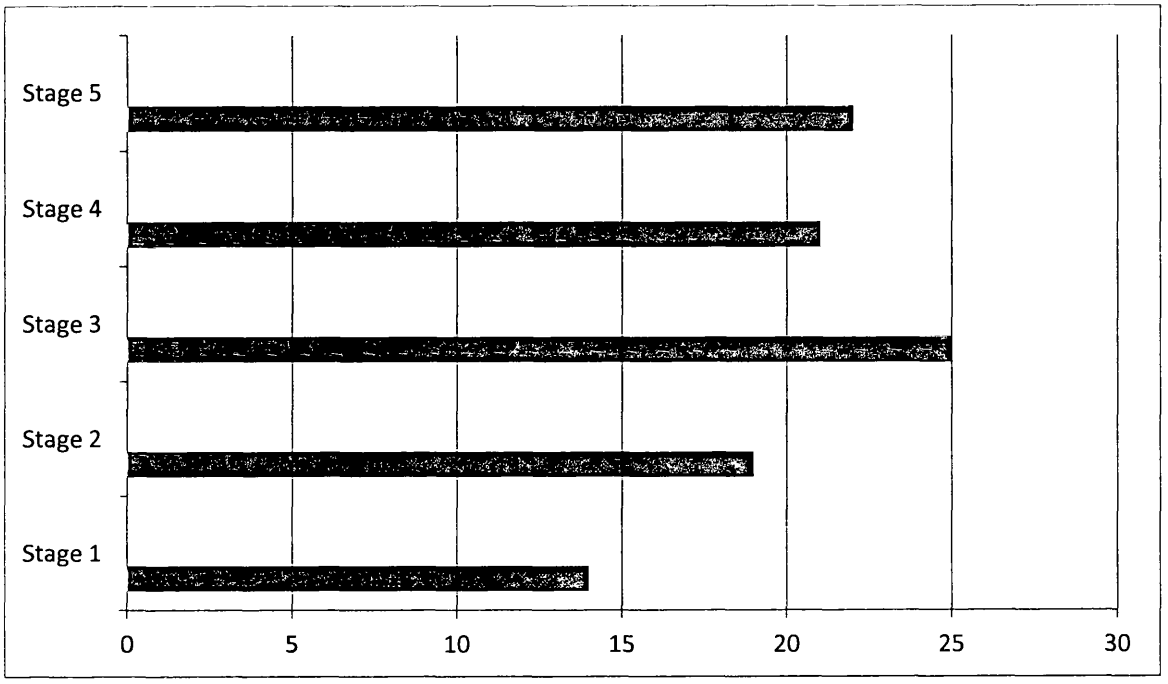
Step 2. After this task had been completed, step 2 of the first stage was repeated in full in the classroom.

Stages 3, 4 and 5 were the same as stage 2.

8.2.2 Results

The first three objectives of the present study seek to explore the ways of self-rating the targeted word knowledge by the learner before and after the reading. Pursuing these objectives we asked the subject to measure and categorize her knowledge of 200 given words five times in this study. The total results of all five categorizations are shown in figure 8-1.

Figure 8-1: Number of the created categories at stages 1 through 5



Note. Horizontal axis shows the categories. Vertical axis indicates the stages of the study.

As seen from this figure, a large number of self-rating categories was submitted at each stage of the study. The largest number of categories, namely 25, was recorded at the third stage of the current study. This is five times as many as proposed by Paribakht and Wesche. This rather substantial list of categories suggested by the learner (together with the other large sets of categories submitted by her at the other stages of this study) indicates the complexity of the learner's self-assessment process and challenges the existing approaches to L2 vocabulary self-assessment. Let us take a look at each of the submitted classifications in turn. The first categorization of the 200 targeted words is introduced below.

Stage 1- Categorization 1: At the first stage of the study, the subject described her knowledge of the given 200 words by dividing them into 14 categories. Her first classification is illustrated in table 8-2.

Table 8-2: Categories suggested by the subject at stage 1

No of Category	Description of the Category	No of Words
1	Verbs I am confident about; can use them in conversation and writing:	6
2	Verbs I can understand but not confident to use:	3
3	Adjectives I am confident to use in writing and speech:	3
4	Adverbs I recognize and would use in conversation:	4
5	Nouns I know well and would use in speech and writing:	25
6	Words I am not confident about but can guess meaning from similar roots (parts of the words) I know:	19
7	Nouns that I know and can transform into adjectives or other parts of speech:	2
8	Words I have usually seen in another grammatical form:	2
9	Words I can provide more than one meaning:	3
10	Words that look familiar but I cannot translate them:	6
11	Verbs I do not know:	42
12	Adjectives I do not recognize:	22
13	Nouns that I do not recognize:	53
14	Adverbs I do not know:	10

Note. Self-report categories are listed in the order arranged and presented by the subject

The table indicates that at the first stage of the study, the learner produced a rather complex classification. Categories 5, 6, 13, 11, 12 and 13 contain a rather large number of words each: 19 or more. The remaining categories hold 10 or fewer words each. As seen from the table, seven different categories which embrace forty six words in total, were declared as known, whereas only 4 categories containing 127 words were described as unknown. Categories 1, 3 and 5 can be combined into one larger block: Words known well, can recognize the word class and use in speech and writing. Likewise, the last 4 categories: 11, 12, 13 and 14 can be considered as sub-groups of a larger category: Unknown meanings but can name the word class. This indicates that the word class feature is a crucial criterion for the subject's

categorization which matches the data submitted by our previous participants. However, at the same time, it should be noted that the subject did make mistakes, though few in number, in determining the word class by classifying nouns as adjectives, verbs and even adverbs. This point will be further addressed in the discussion section.

Stage 2-Classification 2: After the reading, the subject re-arranged her initial categorization. This time, she divided 200 words into 19 categories. This is shown in table 8-3.

Table 8-3: Categories suggested by the subject at stage 2

No of Category	Description of the Category	No of Words
1	Verbs that I know well and I knew them prior to the reading of the text:	8
2	Nouns that I know well and I knew them prior to the reading of the text:	15
3	Nouns that I could recognize prior to the reading of the text but I would not be confident using them as I am not sure about their grammar:	8
4	Adverbs that I recognize and can use in speech and writing:	3
5	Adjectives that I recognize and knew before the reading:	4
6	Adjectives I know since having read the text. I might have recognized some of them before but not properly:	5
7	Nouns/Words (the subject combined two groups together) that I know more or less since having read the text. Some of them I could recognize before but I was not sure, some of them are completely new words:	31
8	Verbs that I can recognize since having read the story. I would more likely recognize them in writing when I do my reading rather than use them	7
9	Adverbs that I can recognize since having read the text. I would not be confident to use them in speech myself:	2
10	Words which look similar to English words -- easy to recognize:	3
11	Words which I know more than one meaning of:	4
12	Words that I can guess the meaning of judging on the parts of these words which I know:	5
13	The same as cat 12 but I am not as certain that I can guess the right meaning -- some parts of these words look familiar:	7
14	Words that I know and I know different forms (word classes) of these words:	9
15	Nouns that I do not know. There is nothing in them that could help me guess what they mean:	39
16	Words I do not know but unlike cat 15, I can take a guess as some parts of these words look familiar to me:	12
17	Words that I do not know but they look familiar and I might remember what they mean when I see the text again:	8
18	Adjectives that I do not recognize:	9
19	Verbs that I do not recognize:	21

The table shows that after the reading, the 200 targeted words were split among newly created 19 categories: that is five categories more than she submitted at the first stage. Categories 7, 15 and 19 with the largest portion of words, 91 in total, constitute the main block of the system. Categories 2 and 16 fall into the second largest block of the subject's categorization. The remaining fourteen categories contain only a few words (2 to 9) each. Categories 1 through 11 and category 14 which contain in total 99 words may be regarded as known: that is five categories more (or 53 words more) than stated in the previous categorization. At the same time, only two categories: 18 and 19 containing 30 words between them were described as totally unknown. Let us take a look how the two sets of categories (Stage 1 and Stage 2) are related. This is shown in table 8-4.

Table 8-4: Transformation of the subject's classification at stage 2

Classification 1	Classification 2
1/1	1
2/1	
3/1	5
4/1	4 + use in writing
5/1	2
6/1	12, 13, 16
7/1	14
8/1	
9/1	11
10/1	17
11/1	8, 19
12/1	6, 7, 18
13/1	7, 15
14/1	7, 9
	3
	10

Though it does not seem possible to detect the exact routes of the transformation of the subject's initial categorization, the changes that the first classification underwent appear to be as follows: categories 1/1, 3/1, 5/1, 7/1, 9/1, 10/1 remained the same, yet the numbering was changed to 1, 5, 2, 14, 11 and 17 respectively. Category 4/1 was altered from: "known, can use in speech" to "known, can use in speech and writing". Categories 6/1, 11/1, 12/1, 13/1 and 14/1 split (e.g. 6/1 split into 12, 13 and 16). Categories 3 and 10 were added to the system whereas categories 2 and 8 seem to have disappeared. The alterations as above, once again, demonstrate the complexity of the process of L2 vocabulary acquisition and challenge the existing approaches to lexical self-assessment.

After another session of reading, the second categorization underwent a new transformation.

Stage 3-Classification 3: At the third stage of the study, after the second session of reading, the subject re-categorized her knowledge of the 200 targeted words by arranging them into 25 categories. Her new categorization is shown in table 8-5.

Table 8-5: Categories suggested by the subject at stage 3, after the second reading

No of Category	Description of the category	No of words
1	Verbs I know well. I am confident to use them in speech, can recognize them in listening and reading:	5
2	Verbs that I know the meanings of but I am less confident using them in speech but I would understand them in listening or reading:	4
3	Verbs that I recognize but I would not be confident using them in speech or writing but if I see or hear them in context I would understand... I would not be certain about their grammar characteristics either:	6
4	Verbs: I am not sure of their meanings but in context I would understand what they mean I can provide some grammar information on these words:	5
5	Words: I recognize some parts of these words and I can guess their meanings based on those parts:	10
6	Words I do not know but I might be able to guess in the context:	10
7	Words that look similar to the English words – easy to guess what they mean, can use, confident about the grammar:	3
8	Nouns that I do not know well. I will understand what they mean in the context but would not be able to use them in speech or writing:	7
9	Nouns I know well, can use in speech and writing; will understand in reading or listening. I am confident about their grammar, etc.:	26
10	Nouns that I know but I would not be as confident using them in speech. However, I could use them in writing. I will understand them in listening and reading:	9
11	Nouns I know the meanings of. They are new to me. I would not use them in speech, might not understand them in listening but I will understand them well in reading:	2
12	Adverbs I know well and I am confident to use them:	1
13	Adverbs I recognize in reading or listening but I would not be confident using them:	2

14	Adverb I think I know the meaning of, would understand in the reading context, but might not in listening, would not use in speech:	1
15	Adverb I think I know the meaning of, would understand in the reading context, but might not in listening, would not use in speech:	7
16	Adjectives I know but would not be confident using in speech. I would understand them in the written form. I can provide grammar info and write them down:	6
17	Adjectives I know well and I am confident using them:	6
18	Words – new to me. Now I know what they mean and can recognize them in context:	7
19	Words that I know more than one meaning of:	4
20	Words I do not know the meaning of. I recognize them as seen before but can't remember:	8
21	Adverbs I do not recognize but might recollect in context:	2
22	Adjectives I do not know the meaning of but can provide some grammar information:	5
23	Nouns I do not know the meaning of but can provide some grammar information:	33
24	Verbs I do not know the meanings of but can provide some grammar information:	25
25	Words I do not know at all:	6

Note. Self-report categories are listed in the order arranged and presented by the subject.

The table indicates a new increase in the self-report categories: 25 compared to 14 and 19 submitted at stages 1 and 2 respectively. On the quantitative basis, the categories suggested by the subject at this stage appear to fall into two blocks. Block 1 contains categories 9, 23 and 24 with the largest portions of words in each of these categories (26, 33 and 25 words respectively). The remaining categories that could be placed in block 2 include 10 or fewer words each.

The "known categories" are 1-3, 7 and 9-19 with 89 words which is ten words fewer than in the previous classification and therefore, this might imply some cases of regression. As seen from the table, categories 21 through 24 can be combined into one larger category: Words I do not know the meaning of but can recognize their word class and provide some grammar information. The relationships between classifications 2 and 3 are illustrated in table 8-6.

Table 8-6: Transformation of the subject's categorization at stage 3

Classification 2	Classification 3
1/2	1
2/2	9
3/2	8
4/2	12
5/2	17
6/2	17
7/2	18
8/2	2, 3(+ grammar)
9/2	13
10/2	7
11/2	19
12/2	5
13/2	5
14/2	15
15/2	23 (+grammar), 25
16/2	5
17/2	6, 20
18/2	22 (+ grammar), 25
19/2	24 (+grammar)
	4; 10; 11; 14; 16; 21

Evidently, the second classification underwent significant transformation after a concurrent session of reading. As noted earlier, these alterations can be interpreted in different ways. Our attempt to analyse how the latest sets of categories are related is shown in table 10.6: Categories 1/2, 2/2, 3/2, 4/2, 7/2, 9/2, 10/2, 11/2, and 14/2 seem to stay in the new classification, though they are now numbered differently: 1, 9, 8, 12, 18, 13, 7, 19 and 15 respectively. Categories 5/2 and 6/2 are combined into one category --17. Likewise, categories 12/2, 13/2 and 16/2 now constitute one category --5. In contrast, categories: 8/2, 15/2 and 18/2 appear to have split into two categories each. One category within each of these newly created pairs includes a new criterion: grammar feature. Furthermore, category 19/2 seems to have been slightly altered to accommodate a grammar criterion.

Categories 4, 10, 11, 14, 16 and 21 appear to be newly created within the third classification system. Qualitative characteristics of the created categorizations will be further addressed in the discussion section.

Stage 4- Classification 4: After the third reading session, the subject, once again, re-arranged the list of her self-report categories. She reported some changes in her knowledge of the 200 targeted words. The new categorization shown in table 8-7 consists of 21 groups.

Table 8-7: Categories suggested by the subject after the third reading

Category Description	No of Words
1: Nouns I am most confident. I would use them in speech and be able to provide some grammar:	22
2: Nouns I know but I am less confident using them in speech but would be able to recognize them in reading and listening. Able to provide grammar information:	15
3: Nouns, I understand the meanings of and would be able to provide grammar information:	8
4: Nouns I know the meanings of and would be able to provide some grammar info. However, I would understand them best in a written text rather than speaking or listening:	10
5: Words that I am not confident of the meanings but would be able to guess what they mean from parts of the words (+ even better in the context):	14
6: Words I know well, can recognize easily because they resemble English words:	3
7: Adjectives I know well:	6
8: Adjectives I know the meanings of but I would be less confident about their grammar:	8
9: Adjectives I know the meanings of but I would not use them in speech. OK with spelling:	2
10: Nouns I know more than one meaning of:	5
11: Words I know the meanings of and can provide other morphological forms:	8
12: Adverbs I know well and I am confident to use them:	6
13: Adverbs I am not confident about but I would take a guess at their meaning:	3
14: Verbs I know well:	9
15: Verbs I know the meaning of but I am not confident using them and I am not confident about their grammar:	19
16: Verbs I know the meaning of but I would have greater difficulty conjugating them (grammar):	5
17: Adverbs I do not know:	2
18: Word I can guess its meaning judging on the sound of the word:	1
19: Adjectives I do not know:	7
20: Verbs I do not know:	19
21: Nouns I do not know:	28

As seen from the table, the subject divided the 200 target words into 21 categories which is slightly less than at the previous stage (25 categories). However, at the same time, this is still substantially more than she submitted at stage 1 (14 categories) and stage 2 (19 categories). The largest categories: 1, 20 and 21which contain more than 20 words each, fall into block 1. Categories 2, 4, 5, 15 with the medium amount of words 10-20 may be regarded as

the second largest block of the system compared to the rest of the categories which contain less than 10 words each.

The table also indicates an increase in the number of words allocated under the" known" categories (i.e. categories 1-4, 6-12 and 14-16) –126 words compared to 89 at the previous stage. "Unknown" categories 17, 19, 20 and 21 can be combined into a larger group: Words I do not know but can recognize the word class.

Clearly, categorization 4 is different from the classification systems introduced by the subject earlier. Let us see how the recent set of categories developed and how it is linked to Classification 3. This is shown in table 8-8.

Table 8-8: Transformation of the subject's categorization at stage 4

Classification 3	Classification 4
1/3	14
2/3	
3/3	15, 16
4/3	
5/3	5
6/3	5
7/3	6
8/3	
9/3	1
10/3	2
11/3	4 (+grammar)
12/3	12
13/3	
14/3	13
15/3	11
16/3	9 (-grammar)
17/3	7
18/3	
9/3	10
20/3	
21/3	17 (-wish to recall
22/3	19 (-grammar)
23/3	21
24/3	20 (-grammar)
25/3	
	3, 8, 18

The table shows that categorization 3 was considerably transformed. Though a number of categories seem to remain the same (e. g. 1/3, 7/3, 9/3, 10/3, 12/3, 14/3, 15/3, yet numbered differently) considerable changes did occur: category 11/3 received a grammar descriptor, whereas categories 16/3, 22/3, 24/3 lost it; category 23/3 changed from "unknown adverbs might recollect in context" to "unknown adverbs" or was replaced by the latter. Categories 5/3 and 6/3 seem to have been combined into one category –5. On the contrary, category 3/3 split into two categories -- 15 and 16. A large number of categories (e. g. 2/3, 4/3, 8/3, 13/3...) appear to have shrunk while three new categories: 3, 8, and 18 were added to the system. This analysis reveals the complexity of modifications which occurred to classification 3 after the reading. This will be further discussed in the next section.

Stage 5- Classification 5: After the last session of reading, the picture of the subject's word knowledge changed again. This time she described her knowledge of the 200 given words by re-arranging them into 22 self-report categories illustrated in table 8-9.

Table 8-9: Categories suggested by the subject after the third reading

Category Description	No of Words Allocated
1: Nouns I know. I am confident to use them in speech and writing. I would understand these words in listening/reading and be able to spell them:	19
2: Nouns I know the meanings of. I would use them in speech but would be slightly less confident than with cat.1:	22
3: Nouns I know the meaning but not confident with the grammar. I would still use them in speech. I would understand these words in reading and listening:	12
4: Words that I know and can provide some other morphological forms of:	16
5: Adverbs. I know what they mean and can use confidently in speech, understand in reading and listening:	4
6: Adverbs. I know what they mean. I would understand them best in reading:	3
7: Words I can provide more than one meaning:	5
8: I knew different Russian words for these phenomena/objects before I started reading. Synonyms:	5
9: I found out that they are words with similar meanings (both are target words):	2
10: I know the meanings of these words. But it also possible to guess the meanings based on the individual parts of these words:	7
11: Adjectives that I know the meaning of and confident to use in speech, understand in reading and listening:	8
12: Adjectives I know the meaning but less confident to use in speech. I would understand them in a written text:	7
13: Recognizable as they are similar to English words:	3
14: I am not sure of the meanings of these words but would be able to take a guess using parts of the word or the context of a sentence:	9
15: Verbs I know the meaning and would feel confident using in speech, writing and listening:	11
16: Verbs I know what they mean, can use in speech but not as confident as the previous category. Understand in reading and listening:	17
17: Verbs. I know their meanings but not confident with their grammar. I would understand them in reading:	11
18: Words – I do not know what they mean and would find difficult to spell them:	2
19: Words-I know the meaning but might find difficult to spell:	4
20: Words which look familiar. I can't translate them but maybe in the context I would remember them:	12
21: I can't translate these words but I know which part of speech they belong to:	18
22: Words that I do not know at all:	3

Note. Self-report categories are listed in the order arranged and presented by the subject.

Judging by the descriptions of the categories presented in the table, the majority of these groups, namely 17, fall into a large block of "known" categories. This can be compared against the other block of "unknown" categories which include only five remaining groups: 14, 18, 20, 21 and 22.

If categorization 5 is compared against the previous classification it is readily apparent that the two systems are rather similar in terms of the quantity of the submitted categories. In light of this observation it is interesting to establish how the two sets of categories are related. This is shown in table 8-10.

Table 8-10: Transformation of the subject's categorization at stage 5

Categorization 4	Categorization 5
1/4	1(+writing, spell)
2/4	2 (-grammar)
3/4	3
4/4	3
5/4	14, 10 (+meaning –context guess)
6/4	13
7/4	11
8/4	12 (+grammar)
9/4	12 (-spelling)
10/4	7
11/4	4
12/4	5
13/4	
14/4	15
15/4	17,16 (-grammar)
16/4	17
17/4	18, 20, 21, 22
18/4	
19/4	18, 20, 21, 22
20/4	18, 20, 21, 22
21/4	18, 20, 21, 22
	6, 8, 9, 19

Again, the transformation of categorization 4 into classification 5 is complex and multi-directional. As shown in the table, the previous categorization underwent a large variety of alterations, such as: categories 1/4, 2/4 transformed into categories 1 and 2 having lost some features and acquired others; categories 3/4 and 4/4 combined into category 3; categories 8/4 and 9/4 also joined together creating category 12 while losing and acquiring some properties; category 5/4 and 15/4 seem to have split into two categories each, having lost or acquired features within some of the newly created categories; categories 17/4, 19/4, 20/4 and 21/4 appeared to have considerably changed to form categories 18, 20, 21 and 22. Furthermore, categories 13/4 and 18/4 seem to have disappeared from the set, whereas a number of completely new categories, namely 6, 8, 9 and 19, were added to it. Finally, categories 6/4, 7/4, 10/4, 11/4, 12/4, 14/4 and 16/4 appear to be the same, but differently numbered.

In summary, having compared the quantitative characteristics of the five classification systems submitted by the subject at the five subsequent stages of the study, we emphasize the complexity and variability of each of the created systems. It is also important to point out that the process of transformation of an earlier suggested categorization into a new system proved to be rather complex and multi-dimensional, i.e. included a variety of operations, such as: extending, shrinking, forming totally new categories, disappearing and popping up again within a new system. These findings hint at the complexity of L2 vocabulary acquisition and stand in contrast to the current vocabulary self-assessment methodologies. The qualitative features of the suggested categorizations will be analyzed in the section that follows.

8.3 Discussion

The third objective of this study is to assess the quantitative and qualitative characteristics of the classification systems created by the subject in this study. The first part of this objective was achieved in the previous section, whereas the second task will be considered in this section. The fourth objective of the study seeks to identify possible re-locations of words among the created categories in each of the five systems submitted. In an attempt to fulfill these tasks, first, in this section, we will analyze the qualitative characteristics of the five self-report systems created by the learner. Second, we will discuss whether any re-locations of target words occurred between the categories that were created by the subject at the subsequent stages of the study.

8.3.1 Qualitative features of the created classification systems

Following the quantitative analysis of the five categorizations carried out in the previous section, we now seek to explore how these five sets of categories differ in terms of their qualitative characteristics. Having analyzed the descriptions of the created categories within each of the suggested categorizations, we identified the features referred to in these descriptions. They are illustrated in table 8-11.

Table 8-11: Features used to describe self-report categories created by the subject within classifications 1 through 5

Features	Classif 1 positive	Classif 1 negative	Classif 2 positive	Classif 2 negative	Classif 3 positive	Classif 3 negative	Classif 4 positive	Classif 4 negative	Classif 5 positive	Classif 5 negative
Meaning	5	5	12	5	17	6	15	4	14	3
Use	5	-	3		10	3	4	1	7	
Word Class	10		12		17		19		11	
Grammar			1		8		6	1	2	
Written Form					1		1		1	2
Sound Form			1		1		2		1	
Morphol incl guess	1		3		1		1		2	
Derivatives	2		1		1		1		1	
Multiple-meaning	1		1		1		1		1	
4 Skills	3		2		9		2		7	
Synonyms									2	
Context guessing					1		1		1	
Certainty	3	2	4	3	5	8	5	5	4	6

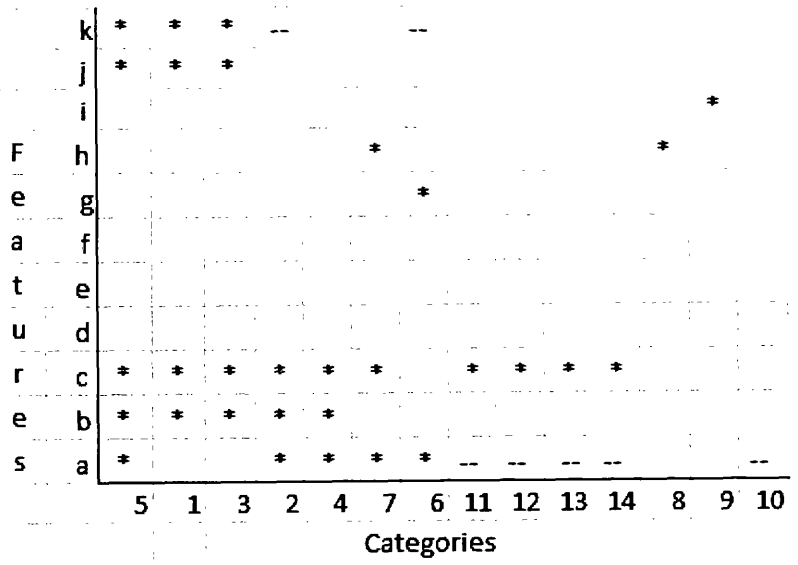
Table 8-11 indicates that the subject used a large variety of descriptors (the largest ever recorded in our research) to characterize her knowledge of the created categories at each stage of this study. Clearly, the main descriptors here are: meaning knowledge, word class belonging, use and certainty in knowledge. Grammar and four skill references would probably constitute the second largest group of features used by the learner to describe the categories she created. Two classification criteria: context guessing and synonyms emerged for the first time in our entire research.

The table also indicates a consistent increase in the number of descriptors (or features) that the subject refers to while describing her knowledge of words, i.e. eight - in the first categorization, ten –in the second, twelve – in the third and fourth, and thirteen – in the final classification. This might illustrate how the knowledge of certain words becomes more complex and varied while moving through the stages of vocabulary acquisition.

It is also worth emphasizing that this learner, similar to our previous participants, used some descriptors in negative context which suggests that the subject considered a large variety of features while categorizing her knowledge of words. This, in turn, supports our point regarding interaction between different features or properties of a word. Further evidence in favour of the existence of those links is presented in figures below. Figures 8-2 through 8-5

illustrate how the features or classification criteria were spread among the suggested categories within classifications 1 through 5.

Figure 8-2: Distribution of features among the suggested categories in classification 1

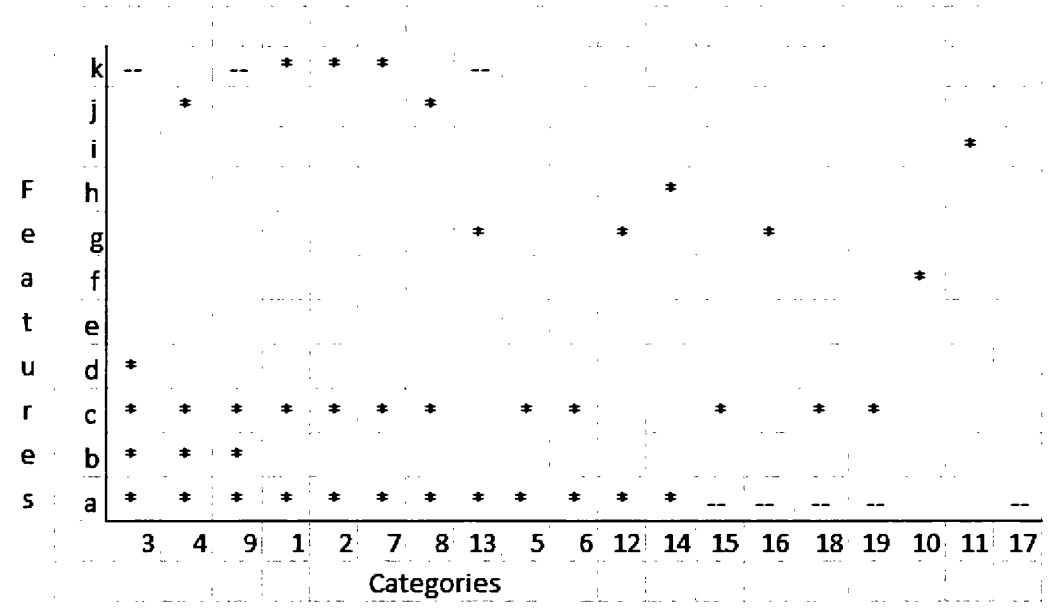


Note. Features: a-meaning; b-use; c-word class; d-grammar info; e-written form; f-sound for g-morphological info; h-derivation forms ;i-multi-meaning; j-the four skills; k- certainty

As seen from this figure, the majority of descriptions for the created categories refer to a number of different features. Moreover, category 5 is described by involving the largest (compared to the other categories in the same classification) number of features, namely 5. Furthermore, categories 1, 3 and 2 are characterized via four descriptors. Descriptions for categories 4, 7 and 6 are based on three different criteria whereas categories 11 through 14 are described by referring to two different features. Thus only the three remaining categories of the created system: 8, 9 and 10 refer to one feature in their descriptions.

We were curious to see whether the number of features reflected in each of the descriptions for the newly created categories in categorization 2 produced after the reading would change. Figure 8-3 illustrates the distribution of features among the submitted categories at the second stage of this study.

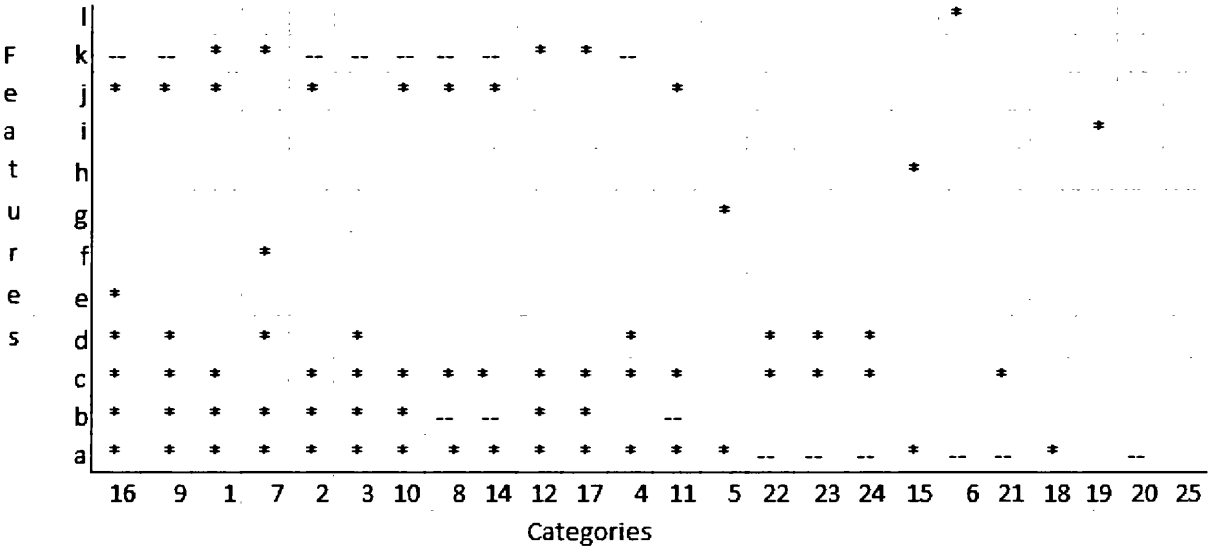
Figure 8-3: Distribution of features among the suggested categories in classification 2



Note. Features: a-meaning; b-use; c-word class; d-grammar info; e-written form; f-sound form; g-morphological info; h-derivation forms ;i-multi-meaning; j-the four skills; k- certainty

A comparison between figures 8-2 and 8-3 indicates that in classification 2, the number of categories described by using three and two features considerably increased: 5 vs. 3 and 8 vs. 4 respectively. In light of this finding we hypothesized that the number of descriptors used to characterize each of the categories in a new categorization 3 submitted after another session of reading would increase again. This is shown in figure 8-4.

Figure 8-4: Distribution of features among the suggested categories in classification 3

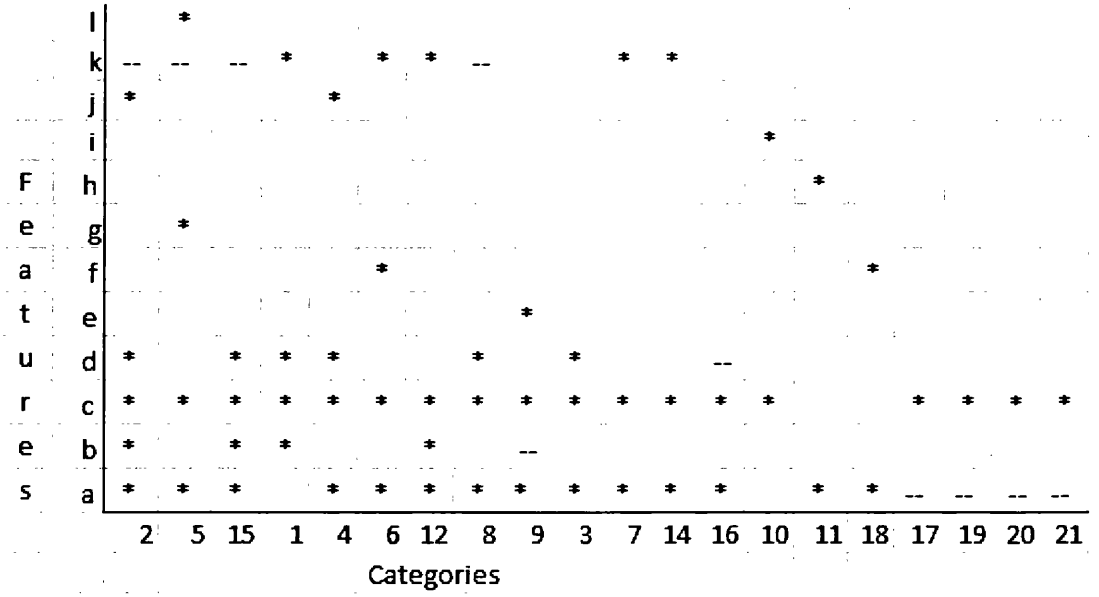


Note. Features: a-meaning; b-use; c-word class; d-grammar info; e-written form; f-sound form; g-morphological info; h-derivation forms ;i-multi-meaning; j-the four skills; k- certainty; l-context guess

A comparison between figures 8-3 and 8-4 reveals that, indeed, the number of features referred to in each of the newly suggested descriptions increased. As seen from figure 8-4, in classification 3, a new descriptor (l: context guess) was added to the list: 12 different features in total were employed to describe this set of categories. Most importantly, category 16 was characterized through the largest number of different descriptors, namely 7. That was the largest number of descriptors used to characterize a single category submitted in our research.

Furthermore, the second largest number of features referred to within one description was six (category 9). In addition, another seven categories in this system, namely 1, 7, 2, 3, 10, 8 and 14, were described via five different features. Thus the data analyzed so far seem to indicate that repeated encountering of certain words leads to the acquisition of new features attached to these words. At this point, we seek to explore this assumption by analyzing the descriptions of some other categories submitted at the later stages of this study. Let us take a look at figure 8-5.

Figure 8-5: Distribution of features among the suggested categories in classification 4

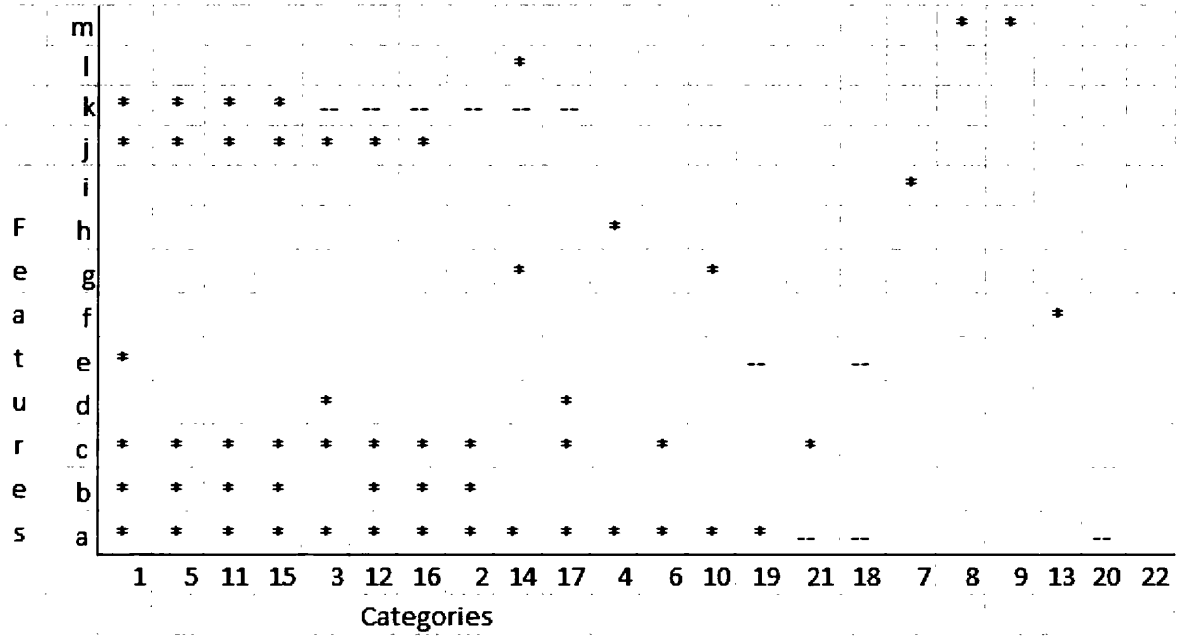


Note. Features: a-meaning; b-use; c-word class; d-grammar info; e-written form; f-sound fo g-morphological info; h-derivation forms ;i-multi-meaning; j-the four skills; k-certainty; l-context guess

Comparing figures 8-4 and 8-5 we note a slight decrease in the number of descriptors used to characterize a single category in classification 4. Nevertheless, the descriptions for this set of

categories can still be regarded as multi-featured, i.e. they refer to rather large number of features. Six different features were referred to in the description of category 2; five descriptors were used to describe categories 5 and 15; four features were mentioned in the descriptions of the further six categories in the figure, namely 1, 4, 6, 12, 8 and 9; three features were referred to in the descriptions for the subsequent four categories in the figure. It should also be noted that the minimum number of features referred to within a description of a single category was two. For the purposes of further analysis we need to compare these results with the data shown in figure 8-6.

Figure 8-6: Distribution of features among the suggested categories in classification 5



Note. Features: a-meaning; b-use; c-word class; d-grammar info; e-written form; f-sound form; g-morphological info; h-derivation forms ;i-multi-meaning; j-the four skills; k-certainty; l-context guess, m-synonyms

When two figures (8-5 and 8-6) are compared, the results reveal some differences between these figures. First of all, a completely new descriptor (m: synonyms) was added to the list at this final stage of the study taking the total number of descriptors used by this subject to 13. In terms of the number of features referred to in a single category description, we can note that five descriptors were used six times (categories 5, 11, 15, 3, 12 and 16) compared to two categories in the previous classification. On this basis, we can conclude that in general, despite an apparent increase in one-feature descriptions, the final categorization seems to be richer than the categorization created at the previous stage of the study.

In summary, we can conclude that the five sets of categories created by the learner in the present study hint at the consistent and enormous complexity of the subject's vocabulary systems. Overall, the subject of the current study submitted the most complex categorizations of word knowledge ever created in our research. That was achieved both in terms of their quantitative (i.e. consistently large numbers of categories in the sets: 19, 25, 21, 22) and qualitative (i.e. the number of features referred to, namely 13 as well as the distribution of these features among the categories) characteristics (compared to 11 descriptors used in the previous study). Though it does not seem wise to generalize the data obtained in case studies, there are still good grounds for saying that our subjects acquire lexical knowledge in different, yet rather complex ways. Furthermore, the complexity seems to increase in the classifications produced by the learners at higher levels of knowledge. This indicates that learners might require individual self-assessment scales to accurately measure their knowledge of words. There are also good grounds for saying that the existing models of L2 word knowledge self-assessment scales fail to register the enormous complexity and richness that are emerging from this detailed longitudinal case study.

8.3.2 The issue of word re-locations

The fourth objective of our study was to explore the re-location of words among the categories after each of the subsequent stages of reading. We switched to a repeated longitudinal study in attempt to establish the most common patterns of these relocations. This will be discussed further in this section. But first let us take a look at the general picture of the re-locations of targeted words after the four concurrent reading sessions. This is shown in table 8-12.

Table 8-12: Summary of the word re-locations among the suggested categories at stages 2 through 5

Status of words	Improve	Stay the same	Regress
No of words Stage 2	75	120	5
No of words Stage 3	28	162	10
No of words Stage 4	64	125	11
No of words Stage 5	53	132	15

As seen from this table, the largest portion of words, namely 80, moved into different

categories within newly created classification 2 after the first session of reading. The second largest re-location, namely 75 words, was recorded at the fourth stage of the study. The majority of these movements can be regarded as cases of progression since the words re-located into the categories with rather more features attached to them. However, there were some cases of regression as well. As seen from the table, a number of words moved backwards within each of the newly suggested classification systems i.e. lost some of the features they were early described with. These relatively large numbers (compared to the results of our previous studies) of regress cases may be regarded as an important finding of our research. This has never been identified by other researchers in the field. The question that might be asked at this stage relates to the causes of such regression. One possible explanation we suggested is that the learner slightly overestimated her knowledge of the certain words at the previous stage/s of the study. However, later on, her perception might have been corrected by the reading.

Another explanation may relate to the fact that the learner did not initially consider a certain feature that she took into account later, for example a move from "known" to "unsure about the grammar" or a move from "known" to "not very confident". In general, the whole system seems to be unstable, with progression and regression taking place simultaneously. Let us take a look at the data collected in more detail. Table 8-13 illustrates cases of word re-location at certain stages of this study.

Table 8-13: Word re-locations that occurred at stages 2 through 5

No of Cat Classif 1	No of Cat Classif 2	No of Cat Classif 2	No of Cat Classif 3	No of Cat Classif 3	No of Cat Classif 4	No of Cat Classif 4	No of Cat Classif 5
1/1	8-1; 11-1	1/2	2-1; 3-3; 8-1	1/3		1/4	2-2; 3-3
2/1	1-2	2/2		2/3	14-2	2/4	1-3
3/1		3/2	10-1; 19-1	3/3	14-1	3/4	1-2; 2-5
4/1	9-1	4/2		4/3	15-5	4/4	2-3; 3-3
5/1	7-3; 11-1	5/2		5/3	1-1; 4-1; 5-1; 7-1; 8-1; 11-1; 12-1; 14-1; 15-1	5/4	4-2; 6-2; 10-2; 15-1; 22-1
6/1	1-1; 2-2; 6-2; 7-3; 14-6	6/2	17-3	6/3	2-2; 3-2; 15-1; 16-1	6/4	
7/1		7/2	5-2; 6-1; 9-3; 15-3; 16-2; 23-1	7/3		7/4	19-1
8/1		8/2	1-1; 2-2; 4-1	8/3	1-2; 4-3; 8-1; 16-1	8/4	4-1; 14-1
9/1	2-1	9/2		9/3	2-5; 4-1	9/4	11-1; 19-1
10/1	5-1; 7-4	10/2		10/3	1-2; 8-1; 16-1	10/4	
11/1	1-2; 8-6; 13-1; 16-6	11/2		11/3		11/4	
12/1	6-3; 7-3; 12-1; 13-2; 16-4	12/2		12/3		12/4	
13/1	7-14; 10-1; 12-1; 13-2; 16-1	13/2		13/3	12-2	13/4	5-1; 6-1
14/1	7-1; 9-1; 13-1; 16-1	14/2	5-1; 23-1	14/3	13-1	14/4	16-3
		15/2	6-3; 14-1; 18-1	15/3	12-1	15/4	15-4
		16/2	4-1; 16-1	16/3	7-2	16/4	16-1
		17/2	4-3	17/3	8-3	17/4	12-2
		18/2		18/3	3-2; 4-3	18/4	
		19/2	6-1	19/3		19/4	11-3; 12-1
				20/3	5-2; 12-1; 15-3	20/4	14-1; 16-5
				21/3		21/4	2-6; 3-1; 8-2; 10-1; 19-2
				22/3	8-1		
				23/3	3-3; 4-1; 5-4; 13-1		
				24/3	5-3; 16-1		
				25/3	5-1		

Note. 1/1 – category 1 within classification 1; 15/3 – category 15 within classification 3, etc. 8-1 –one word moved to category 8; 19-2 – two words moved to category 19, etc.

For example, the first row of the table indicates the following. After the first session of reading, one word moved from category 1 (classification 1) into category 8 (classification 2) and one word moved from category 1 (classification 1) into category 11 (classification 2). After the second session of reading, one word moved from category 1 (classification 2) into category 2 (classification 3); three words moved from category 1 (classification 2) into category 3 (classification 3) and one word moved from category 1 (classification 2) into category 8 (classification 3). After the third session of reading, there was no word re-location from category 1 (classification 3). And finally, after the fourth session of reading, two words moved from category 1 (classification 4) into category 2 (classification 5) and three words moved from category 1 (classification 4) into category 3 (classification 5).

Let us take a look at these re-locations at each of the stages that they occurred, in detail. At the second stage of the study, after the first reading session, 80 words changed their initial positions in classification 1. Most of these changes, namely 75, indicate gains in the subject's word knowledge since the words moved into the categories with more features attached to them. However, 5 words did regress compared to their positions in the first categorization i.e. lost some of the features attributed to them earlier. In terms of the nature of these re-locations, a large variety of patterns of word movements emerged at this stage (e.g. from *receptive* to *productive knowledge*; from *morphological guessing* to *known;* from *look familiar, can't translate* to *receptively known* or *known*). However, the main pattern of word re-locations here seems to be movements from "totally unknown" to "known more or less", for example 14 words re-located from category 13/1 "nouns I do not recognize" to category 7/2 "known more or less". Furthermore, the second largest re-location, namely six words, occurred from category 11/1 "unknown verbs" into category 16/2 "prepared to take a morphological guess". Another six words moved from category 6/1 "prepared to take a morphological guess" into category 14 "known with derivatives". In regards to the cases of regression, the main pattern of word re-locations appears to be a change from "known productively or well known" to "known receptively or less confident in my knowledge".

It is also noteworthy that the largest distribution of words (i.e. re-location into a large number of different categories) occurred from categories 6/1, 11/1, 12/1, 13/1 and 14/1.

At the third stage of the study, after the second reading, 38 words (that is 42 words less than at the previous stage) changed their locations while moving into newly created classification 3. This relatively low figure might be explained by the fact that the subject, according to her own confession, did not read the story as thoroughly as she did the first time. This might also explain a slight increase in regress figures in classification 3. In respect of the patterns of word re-locations at this stage, we admit that we found it difficult to establish the most common patterns due to a rather high variation in the directions in which the certain targeted words moved. The largest distribution of words occurred from category 7/2: twelve words moved into six different categories within classification 3.

At the fourth stage of the study, after the third session of reading, 75 words (including 11 cases of regression) relocated into different categories within categorization 4. The most common patterns of word re-locations at this stage seem to be moves from the categories "words which are not known well" or "not sure about the meaning, prepared to take a guess" to the categories "known well", "known but not confident using in speech", "known but might have problems with grammar or spelling". For cases of regression, the main pattern appears to be a move from "known well" (cat17/3) to "unsure about the grammar" (8/4) or a change from "well known" (cat 9/3) to "less confident using in speech" (cat 2/4 and 4/4).

The largest distribution of words occurred from category 5/3: nine words re-located from this category into 9 different categories within classification 4.

At the final stage of the study, after the fourth reading session, 68 words re-located into different groups within classification 5. The most common patterns of word re-locations at this final stage seem to be a move from "unsure of the meaning, prepared to take a guess" (e.g. cat 5/4) to "known" (cat: 4/5, 6/5, 10/5), a move from "known" (e.g. cat 3/4) to well known (cat 2/5) as well as from "unknown" (e.g. cat 21/4) to "known" (cat 2/5, 3/5, 8/5). In respect of the recorded cases of regression, it should be noted that they are mostly related to the subject's certainty in her knowledge and do not really imply serious backsliding.

The data collected in the current repeated longitudinal study hints at the immense complexity of the L2 vocabulary acquisition process. As seen from the table, each lexical input via reading led to a significant re-shift of the subject's categorization. Furthermore, the study revealed a rather large range of patterns of word re-locations after each session of reading.

This variety of patterns may be explained by the subject's individual style of acquiring new vocabulary as well as by the specificity of the targeted words. Let us take a look how certain words moved around in the learner's categorizations. For instance, the word **"Кисло" (Sourly)** moved through the following stages:

Stage 1: Category 14 (Adverb I do not know) --- Stage 2: Category 7 (Word I know more or less since I have read the story) --- Stage 3: Category 15 (Words I recognize and can provide their morphological forms)—Stage 4: Category 11 (Words I know the meanings of and can provide other morphological forms for these words)—Stage 5: stayed the same.

The target word **"Объявление" (Advertisement)** seems to have chosen a slightly different route of development in the subject's categorizations. *It was totally unknown at stage 1, became known receptively at stage 2, stayed the same (known receptively) at stage 3 and developed to productively known ("confident to use in speech") at stages 4 and 5.*

The word **"Сумерки" (Twilight)** followed its own pattern of changes in the learner's classifications:
Word that looks familiar but I cannot translate it (stage 1) – Noun I know more or less since I read the text (stage 2) – Noun I know but would not be confident using it in speech (stage 3) – Noun I know, can use in speech but I am not confident (stage 4).

The word **"Гусеница" (Caterpillar)** progressed in the subject's lexicon by passing the following stages:

Noun I do not know (Stage 1) --- Noun I do not know (Stage 2) --- Can guess in context (Stage 3) --- Known receptively (Stage 4) --- Known productively (Stage 5).

Similarly, the word **"Уставать" (To get tired)** moved through a number of different phases in the learner's categorizations:

Verb I do not know (Stage 1) – Looks familiar, might be able to remember in context (Stage 2) – Not sure of the meaning but would understand in context (Stage 3) – Known receptively (Stage 4) – Known productively (Stage 5).

Some words moved back and forth within the subject's classification systems, for example the word **"Хвост" (Tail)** passed the following stages:

Nouns that I do not recognize (Stage 1) – Nouns that I know more or less...(Stage 2) – Nouns I do not know the meaning of but can provide some grammar information (Stage 3) – Nouns I understand the meanings of and would be able to provide some grammar information (Stage 4) -- Nouns I know the meanings of. I would use them in speech... (Stage 5)

However, at the same time, there were also cases when a totally unknown word moved straight into a category "known well" (e.g. six words moved from category 21/4 into category 2/5 at the final stage of the study).

Such examples indicate that different words do follow different routes or patterns of development or regression within the L2 vocabulary acquisition. These patterns may considerably vary in terms of stages the word passes through. The importance of this finding is apparent since it demonstrates how a particular word develops in the learner's lexicon. Furthermore, consistent relocations of the targeted words indicate that each reading session facilitates new changes in the learner's perception of her word knowledge. Most importantly, this process is described by the learner herself as a view from inside the system (i.e. through the learner's own perception).

8.4 Conclusion

In this chapter, we reported on the results of the repeated longitudinal study which was undertaken to further explore the issue of learners' categorizations of their own word knowledge. The results of the study indicated that the subject was consistent in creating rather complex classifications while self-assessing her knowledge of the certain words at each of the five stages of the present study. Furthermore, the general picture that emerged from this study is that the current subject produced a more complex classification system than those we noted earlier in our research. This indicates that self-assessment methodologies in L2 vocabulary acquisition should take into consideration learners' individual styles of acquiring and assessing their lexical knowledge.

The study also revealed that the subsequent sessions of reading promoted significant changes in the subject's self-rating system. Moreover, the obtained data demonstrated a large diversity in patterns of word re-locations after each of the subsequent sessions of reading. The whole system proved to be rather dynamic with progression and regression occurring simultaneously

which also hints at enormous complexity of L2 lexical acquisition.

Thus the findings of our research suggest that L2 vocabularies may be a lot less stable than some of the earlier discussions in the literature imply. The implications of this idea as well as some other interesting points that emerged from our research will be discussed in more detail in the next chapter.

CHAPTER 9

DISCUSSION

9.1 Introduction

In the previous chapters of this research, we reported on several experimental studies that were undertaken to explore how L2 learners self-assess their knowledge of words. The results of these studies came as a surprise: not only did they demonstrate that the learners were generally very good at measuring and classifying their knowledge of words but they also showed an enormous complexity and variability in the subjects' categorization systems. This complexity proved to be both of a quantitative (i.e. the number of self-rate categories created by the subjects) and qualitative (i.e. the number of features referred to as well as the distribution of these features among the categories) nature. The data also indicated that the categorization systems submitted in our research noticeably differed in terms of their quantitative and qualitative features. We cannot definitely ascertain whether these differences are conditioned by learners' levels of knowledge or their individual styles of lexical self-assessment. Nevertheless, it follows from our research that learners' individual characteristics should be taken into account by self-assessment methodologies. The results of the repeated longitudinal study presented in the previous chapter indicated that the immense complexity of the learner's categorization system did not seem to be incidental since it was repeated at each stage of the study. These findings seem to radically change the whole view on the process of L2 vocabulary knowledge self-assessment. Though further research on larger numbers of subjects would be beneficial in terms of further testing the findings obtained in this research, it is clear that a new approach to the lexical knowledge self-assessment is required.

Another discovery of our research is extensive re-locations of words among the categories created by the learners within their classifications of own lexical knowledge. After each session of reading taking place in our studies, a learner's classification system was significantly re-shifted and altered. This was best demonstrated by the results of the longitudinal study presented in chapter 8. The categorization system created by the subject in that study was rather unstable with the words moving backwards and forwards among the changing categories. Again, that was not revealed by other research in the field. These issues as well as some other interesting findings will be further addressed in our discussion section.

Generally speaking, the work reported in the previous chapters has raised a number of important issues that have not been picked up in research on vocabulary self-assessment. The issues that will be discussed in the chapter that follows are:

- The importance of partial mastery
- How learners deal with words they cannot translate
- The instability of the classifications: re-categorization
- The instability of the classifications: re-location of words

Let us take a look at each of these points in turn.

9.2 Discussion of the outlined issues

9.2.1 The importance of partial mastery

In this section, we will discuss the issue of self-testing learners' partial knowledge of words. We will argue for the importance of assessing a large variety of features of L2 lexical knowledge while self-measuring knowledge of words.

The results of our experimental research provided evidence in support of the view that knowledge of a word's meaning is only one factor among many that make up what the learner knows about a word. This could be illustrated by the ways that learners describe and categorize their knowledge of words. Our study shows that learners take into consideration a large number of aspects when they attempt to evaluate their knowledge of words. Self-report categories suggested by the subjects while assessing their own knowledge of words include major types of word knowledge, such as meaning senses, use in speech and writing, word class, derivation forms, morphological knowledge, grammar features, written and sound word forms and others which supports Nation's concept of vocabulary depth (2001).

In fact, in our research, we were curious to establish how learners themselves understand "mastery of the word". In their interpretation, as reflected in the descriptions for suggested categories, this seems to imply a wide range of features. This is clearly illustrated by all classification systems submitted in this research. Moreover, the categorization that includes

the largest number of different features, namely 13, was created at the final stage of the longitudinal study reported in the previous chapter. The list of descriptors used by that subject contains the following features:

a) meaning knowledge;
b) use in a sentence;
c) word class/part of speech knowledge;
d) grammar information;
e) morphological knowledge;
f) knowledge of derivation forms;
g) spelling knowledge;
h) phonological knowledge;
i) differentiation between the four language skills;
j) multiple-meaning knowledge;
k) ability to guess in context;
l) knowledge of Russian synonyms;
m) certainty in knowledge.

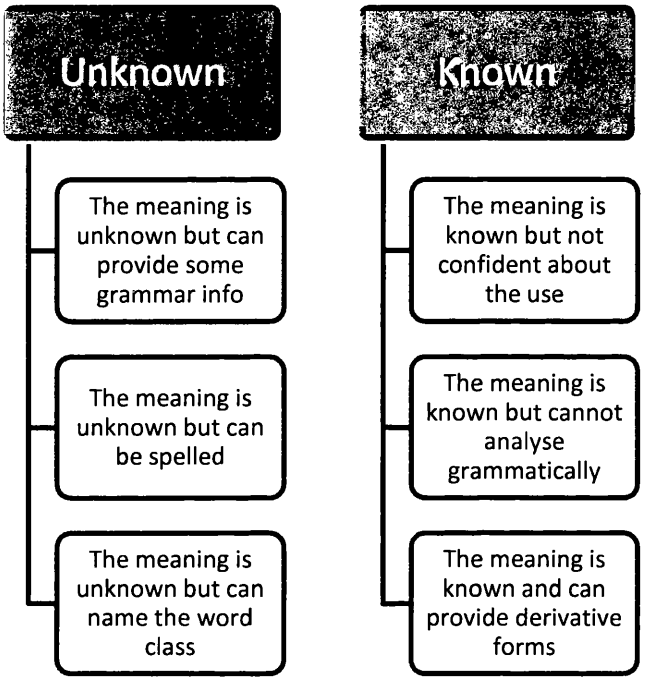
Clearly, in this categorization, the phenomenon "word knowledge" incorporates a rather large number of features. Of course, these features vary from subject to subject and from stage to stage. The point we are making here is that our subjects considered all or most of these features or some other word properties while assessing how well they knew the word. Furthermore, in instances, when learners did not know what the word meant, they were still assessing their knowledge of some other features. For example, they reported that they would be able to spell the word and/or determine its word class belonging and/or describe its grammar features and/or provide its derivatives. Hence, in regards to meaning senses, it should be emphasized that even though meaning knowledge was reported among the most important features for our subjects, they also used a large number of other descriptors. This is contrary to Paribakht and Wesche's self –assessment concept which focuses on the knowledge of a meaning sense and use in a sentence. Furthermore, none of the existing self-assessment methodologies takes into consideration the immense complexity of lexical classification systems created by the learners in our research.

With regards to the word forms (phonological, morphological, grammar), they seem to have been referred to in descriptions for a large number of categories within the subjects' self-rating scales. Interestingly, the learners were assessing possible difficulties: phonological, morphological and grammar which they may experience while using the word in speech and writing and/or encountering it in reading or listening. For example, among the most common comments were the following: "I know the meaning of the word but I would not be confident using it as I am not sure about its grammar" and "I will recognize the word in text but I would not be able to use it myself" or "I think I know the meaning of this word, I would understand it in the reading context but might not in listening, I would not use it in speech". Such examples indicate that a large variety of combinations of different features were used by the learners to describe their categories. Moreover, these features were mentioned both in positive and negative context which again, indicates that subjects used a comprehensive (i.e. broad, embracing various features) approach while measuring their knowledge of words.

However, a lack or a shortage in knowledge of the word forms, as well as the word's meaning/s, still did not appear to stop the learners' claims that a word was known to some extent. For instance, in most cases, they were able to name the word class and argued that they would be able to guess in context. This proves, again, that learners consider a large variety of word features while creating their complex classification systems. Simply put, our subjects did not just choose between "known" or "unknown", they chose between "totally unknown" and "known or unknown to a certain degree". In other words, there seems to be different ways or levels of "unknown". If a word's meaning sense is unknown, the word must not immediately be labeled as totally unfamiliar to the learner. Furthermore, the cases where the student is able to demonstrate some knowledge in relation to other word knowledge aspects (other than meaning) should be treated as "unknown to a certain degree". At the same time, cases where a learner is able to translate/understand the word should not be automatically accepted as known. If a subject struggles to provide any further information on the word or use it correctly in speech and/or writing, the word should be taken as "known to a certain degree". This speculation is based on the data provided by our subjects and challenges the existing approaches to the self-assessment of lexical knowledge.

Let us take a look at Figure 9-1. This figure presents examples of different levels of "unknown" and "known".

Figure 9-1: Example levels of unknown and known



Note. This is only an example taken from the subject's descriptions of the self-report categories she created in study 5 (chapter 7). A large number of other levels are possible.

As seen from Figure 9-1 different levels of "unknown" and "known" can and should be identified when the depth of learners' word knowledge is self-assessed. The "unknown word" issue will be further considered in the next section.

Overall, the self-report categorization indicated that learners consider different types of word knowledge as well as different levels of certainty in that knowledge. In other words, the word familiarity scales constructed by the subjects are typically based on two dimensions: areas of word knowledge as well as certainty in that knowledge. Generally speaking, this is a new interpretation of self-assessment of word knowledge based on learners' self-perception of their own vocabulary knowledge which has not been investigated in other research. This raises the issue of an adequate self-assessment instrument for measuring learners' lexical knowledge. Those suggested so far including Paribakht and Wesche's VKS, however, do not

provide sufficient facilities for taking accurate measurements of this knowledge. In other words, the existing self-assessment methodologies do not seem to reflect the complexity of subjects' own descriptions of their word knowledge. These descriptions, however, hint at the multi-constituent structure of learners' vocabularies which should be considered within the structure and scoring system of a vocabulary self-assessment scale. This may be achieved by means of a multiple-feature self-assessment instrument which would encourage and aid learners in measuring different aspects of their word knowledge. This is a new approach to the self-assessment of learners' word knowledge based on the data collected in our research. The question here is why these points were not taken into consideration by Paribakht and Wesche?

9.2.2 How learners deal with words they cannot translate

Another issue that emerged from our research and briefly outlined in the previous section, is the issue of how learners deal with the words that they cannot translate. Whereas Paribakht and Wesche simply categorize any unknown word into a single category of "unknown", the evidence gathered in our research shows that learners have different ways of not knowing a word. In other words, the research highlighted how lexical items with unfamiliar meanings are categorized in learners' self-assessment scales. Our subjects seemed to have a number of different approaches to this type of words. Their ways of dealing with unfamiliar lexical items appear to be as follows:

1) a learner admits that the word is totally unfamiliar. In cases like this, our subjects demonstrated two different approaches: a) under the first approach, they stated that they did not see this word before and were unable to provide any information about this word. Consequently, the word was allocated into the category "totally unknown"; b) within the second approach, a learner used a number of features/word properties in a negative context to emphasise particular aspects within a general picture of not knowing the word, e.g. *"Words – I do not know what they mean and would find difficult to spell them"*. This suggests that the subject did not simply disregard the word as "unknown" but emphasized and described certain aspects of her unfamiliarity in relation to that word.

2) The second main scenario of dealing with unknown words begins when a learner

acknowledges that they do not know what the word means. However, they provide (or claim that they are able to provide) some information about the word, for instance, its grammar features or word class belonging (see Figure 9-1), e.g. "*Nouns I do not know the meaning of but can provide some grammar information*" and "*I can't translate these words but I know which part of speech they belong to*". Furthermore, the issue of word class belonging should be addressed separately. Whereas grammatical, for example, features of an unknown word can be identified in English (in many instances) when the word's meaning is not known, the word class belonging is not strongly marked in the English language. On the contrary, in Russian, in most cases, a word's word class or part of speech can be identified from its form. Furthermore, this information may assist in acquiring other properties of the word, grammar for instance. Although there is a possibility of misjudging the word class of some of the Russian words (due to deceptive endings) especially if the word's meaning is unknown (which was revealed in our studies), this aspect appeared to be one of the main classification criteria for unknown words.

3) The third way of dealing with unfamiliar words is demonstrated in cases when a learner states that the word looks familiar, however they cannot remember what it means and they are prepared to guess. Under these circumstances four further scenarios are possible:
3a) the learner makes a guess at the word's meaning based on the morphological structure of the word;
3b) the learner makes a guess on the word's meaning based on the phonological structure of the word;
3c) the learner makes a guess on the meaning based on the phonological similarities with the words they know;
3d) the learner attempts to find words with the same root morphemes (i.e. symphorms) within the set of the target words in order to guess their meanings.

With regards to the outlined guesses, it is necessary to note the following. As Huckin and Bloch (1993) point out there is always a possibility of "deceptive transparency" or "mistaken ID". This seems to have been demonstrated by some of the created categories. In fact, some of the target words did look or sound as if they provided clues to their meaning. For instance, the Russian noun **изгородь** looks as if it consisted of **из +город +ь** and meant something linked to the word **город "city, town".** However, the actual meaning of the word **изгородь** is

"hedge". An example of phonologically "mistaken ID" might be illustrated by the word **труп**. The English translation for this word is "**corpse**". However, phonological associations might lead the learner to a translation guess "**troops**". Overall, although some cases of "mistaken ID" did take place in the learners' self-assessment, it is important to emphasise the following points. All subjects demonstrated (in their category descriptions) that they were willing to take a guess at the word's meaning. Moreover, they seem to have created separate categories for different kinds of guesses (for example: phonological, morphological, context, etc). In addition, the data obtained in study 6 might indicate that guessing was significantly enhanced by subsequent sessions of reading since the number of "guessing" categories was increased at the consequent stages of the experiment. Thus, guessing might suggest that some sort of relations: morphological, phonological etc. between new and known words are becoming established in the learner's lexicon after reading.

Furthermore, all of our studies revealed that a large number of words moved from "guessing" categories into "known" categories, at the subsequent stages of the word categorization. This might suggest that guessing did promote the acquisition of some of the targeted words.

4) Another way of dealing with unknown words is shown in cases when a learner admits that they do not know the meaning of the word and though they think it looks familiar, the learner is not attempting to make a guess. They simply state that they are unable to translate the word. This might suggest that the target word which appeared to be new for the subject left only a light trace in their memory. Since no links to other words or parts of the words have been created, the learner may not have been able or prepared to guess.

The implications of these findings are as follows. As seen from the research, "unfamiliar" lexical units should not all be considered as equally unknown. The self-assessment reports show that learners themselves differentiate between "unknown" words. Clearly, there are different levels or stages of unfamiliarity. Let us take a look at Figure 9-1 again. The descriptions of the subjects' word knowledge (in the left column) which were predominantly recorded at the initial stage/s of the assessments (before the reading session/s started) indicate that learners do not automatically reject the words they cannot translate. On the contrary, they attempt to analyze these words and draw as many clues as they possibly can in order to assist themselves in their understanding. This word analysis might promote further acquisition

of the target words when a learner encounters them in a reading text. Since some analysis of the word had been carried out by a learner during the initial vocabulary self-assessment/s, the image of that word may have been improved or enhanced in the learner's mind after the reading.

The issue of different levels of "unknowing" a word never surfaced in the previous research on lexical self/-assessment. However, the importance of self-/measuring to which degree the word is unknown seems to be apparent since it constitutes an essential part in determining the depth of the knowledge of that particular word. In other words, this is the other side of the issue of partial knowledge of a word which was brought up by Paribakht and Wesche (1992). However, despite the speculation on the issue of measuring learners' partial vocabulary knowledge, this idea was not realized in Paribakht and Wesche's self-assessment approach. Furthermore, the number of self-rating categories included by the authors in their scale is insufficient for detecting minor changes in the developing knowledge of a word.

9.2.3 Re-categorization of self-rating classifications

In this section, we will discuss the issue of creating new sets of self-report categories after the vocabulary input through reading. The data recorded in our studies shows that learners generally created a very large number of categories to describe their knowledge of the given words. Moreover, the classification systems originally produced by the learners were significantly re-structured at each stage of the studies: new categories were added, some of the existing groups were split or altered and some simply vanished. In general, each classification system created by our subjects was complex, varied and multi-featured.

As noted in the previous section, a wide set of criteria was considered by each learner. These criteria normally fell into the two dimensions: familiarity and confidence. The former represents word properties or aspects of knowledge, while the latter shows a degree of certainty in this knowledge. It is likely that the detailed descriptions of the categories submitted within each of the categorization systems, indicate that learners store a great amount of information about each word in their memory. Once encountered in a text and retrieved from the memory, the word might pull out the features attached to it. These sets of features may direct the learner while they are arranging or re-arranging words into the categories within their classification systems. Thus, re-categorization of words is likely to be

caused by acquiring, losing or simply considering some new features that were not taken into account at the previous stages.

The processes of re-categorization seem to vary considerably from study to study and from subject to subject. This differentiation might reflect the individual perception of the word knowledge by our subjects as well as their individual routes of acquiring this knowledge. In other words, while assessing their lexical knowledge, different people might make different judgments in terms of the importance of the word features they consider. The priorities of some features over others, referred to by learners while rating their knowledge of the target words, might also indicate which features are acquired first. With regards to the Russian vocabulary acquired by a foreign learner, such features as: word class knowledge, semantic knowledge, use in a sentence, grammar information and morphological knowledge appear to be learned (and assessed) first. However, it should be noted that learners' priorities appear to change after reading which is reflected in significant alterations of their classification systems. Thus, as indicated by the results of this research, all our subjects altered their initial sets of self-rating categories after vocabulary input through reading. Moreover, the subject of our repeated longitudinal study (chapter 8) devised five different systems including the richest (i.e. the most multi-featured) classification of our entire research. Thus, at different stages, learners might have a different hierarchy of the features with regard to a certain word. Furthermore, as noted above, this seems to explain why our learners consistently change the classification systems created by them at the previous stages.

In general, every word seems to be associated with specific features at a particular stage of a learner's categorization. This assumption is supported by the fact that, describing their knowledge of a particular word, learners tended to refer to certain features even if they indicated the absence of knowledge in regards to that particular feature. For instance: "The meaning is known but I am not sure how to use the word (or unsure about its grammar, etc)". Furthermore, the learners seemed to create a category over a word rather than simply placing words into already prepared categories. This might suggest that each target word underwent a separate analysis in terms of its features and properties. Now and again, the subjects were thinking aloud describing a target word. They repeated a word a number of times in an attempt to determine its features, assess their knowledge of the word against those features and, subsequently, either create a new category for the word or allocate it into one of the

existing categories of the similar (according to the chosen criteria) words. This procedure was repeated consistently whenever the learners carried out the classification tasks.

Analyzing the data collected in this research, study 6 in particular, we note that sets of categories suggested by the learners, might expand at some stage, then shrink, then expand again. This can be seen from Table 9-1 (study 6).

Table 9-1: Self-categorization of the target word knowledge in study 6 (Chapter 8).

Exp. Stages	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
No of Cat	14	19	25	21	22

Evidently, some categories disappear from the self-rating scale when the target words become better known. New categories are created when the information about the properties of the words expands. For example, comparing categorizations created by the learner at stages 1 and 2 in study 6 (Tables 8-2 and 8-3) it can be seen that the added categories express the subject's willingness and readiness to take different types of guesses in regards to the words most of which were characterized as unknown at stage 1. However, eventually, the number of "guessing" categories within the new categorizations suggested at the subsequent stages of the study reduces. This occurs due to the relocations of words into the various "known" categories or sometimes back to "unknown" (cases of regression).

With regards to the expanded categorizations, as we noted earlier, the continuous development of knowledge of particular words seems to have led to new features/ descriptors to be chosen to form a new self-rating category of words. Thus, for instance, the entirely new criterion that emerged at the final stage of experiment 6 (Chapter 8): *the ability to find Russian synonyms for the new words acquired through reading* formed two new categories within the subject's final categorization.

In respect of the qualitative changes in submitted categories, it is fair to note that it was generally hard to distinguish between newly created categories and those which were altered. Let us take a look, for instance, at the following descriptions for the created categories of words: "Nouns that I do not recognize" (classification 1, study 6, chapter 8) compared to: "Nouns that I do not know. There is nothing in them that could help me guess what they

mean" (classification 2, study 6, chapter 8) and the next category in the same classification: "Words I do not know but unlike the previous category, I can take a guess as some parts of these words look familiar to me" (classification 2, study 6, chapter 8). In case of a slight differentiation in the description of the category towards its upgrading (as a new feature was considered by the subject), we can speculate about qualitative changes in these descriptions. These changes, in turn, might hint at the alterations occurring in the learner's lexicon to accommodate the new lexical items being acquired. If this is the case, it would be worth arguing for the necessity of learners' regular self-measurement of their knowledge of the words which are being learned. This might highlight the areas or word knowledge aspects which are required to be corrected or enhanced which, in turn, may accelerate the process of word acquisition in general.

In summary, most importantly, the diversity and immense complexity of word categorizations created by the learners in our research might suggest that a word can be identified in the mental lexicon by its different features or properties. The connections between various features of the word might persistently become stronger (or in rare cases --weaker) with every new vocabulary input. Furthermore, evidently, new features become added to the knowledge of a word (e.g. different types of word knowledge). At the same time, some of the existing properties of the word might become dull or completely vanish (e.g. the ability to promote various kinds of guesses). This is reflected in quantitative and qualitative changes in the categories within the self-ranking scales which occurred over a period of time (Studies 1 and 3) as well as after each of the reading sessions (Studies 4, 5 and 6).

Overall, our findings suggest that the processes of L2 vocabulary self-assessment and lexical acquisition in general are enormously complex and multi-directional. This feature has been seriously underestimated by other scholars in the field.

9.2.4 Instability of the categorizations: re-location of words

In the previous section, we discussed one of the major findings of our research: re-categorization of self-rating systems created by the learners. We noted that every reading session seemed to promote certain alterations in the structure of learners' categorizations of words. This might indicate certain changes in the subjects' knowledge of words or their perception of this knowledge. In this section, we will discuss another rather important finding

of our studies: re-location of words among the self-rating categories suggested by the subjects at different stages of this research. Indeed, the data obtained in our research shows considerable shifts of the targeted words. Evidently, every single reading input led to some kind of changes in the positions of certain words within the self-assessment categories. This may indicate how a particular word moves (up- or down-grades) in the structure of the learner's vocabulary. The most intensive relocations of words were recorded in studies 5 and 6. In study five, for example, 97 words in total (out of 200 words assessed) relocated (94 upgraded, 3 recessed) into different categories after just a single reading. In study six, 80 words (out of 200 words assessed) re-located from their original categories after the first reading at the second stage of the study, 38 words moved at the third stage after the second reading, 75 words – at the fourth stage and 68 words moved into different categories at stage 5.

This continuous shift of words among the changing categories might imply that words are in constant rotation in the lexicon. Some of them might mature after having acquired new properties or features, others get dull and eventually fade out (see cases of progression and regression in our studies: chapters 6, 7 and 8). Furthermore, the numerous word relocations, traced in our research, seem to reflect the process of word assimilation and development in the learners' lexicon. The relocation of a word among categories created by a learner appears to suggest that a word moves through different phases during the process of its development in the learner's mind. These phases seem to vary considerably from word to word. Some examples of word relocations are illustrated in Figure 9-1 and Figure 9-2.

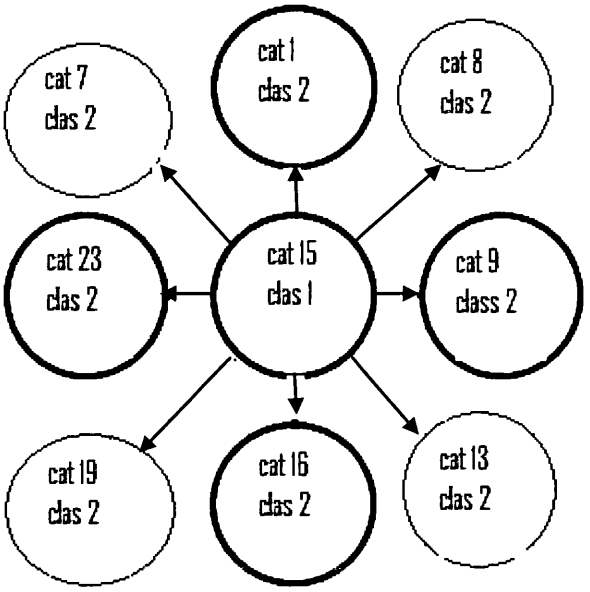
The repetitive exposure to certain words via successive sessions of reading appears to considerably accelerate the process of movement of certain words through stages. Though on most occasions, a word moved to a different, higher level of knowledge, cases of regression were also recorded. As noted above, different words seem to follow different routes while being assimilated in learners' lexicons. Repeated reading sessions revealed certain stages along these routes of development of the targeted words. Let us take a look at a number of examples from study 6 (chapter 8). The Russian word "**Гусеница**" (caterpillar) was progressing in the subject's self-rating scales by moving through the following stages: **Noun I do not know (Stage 1) --- Noun I do not know (Stage 2) --- Can guess in context (Stage 3) --- Known receptively (Stage 4) --- Known productively (Stage 5).** The word

"**Объявление**" (advertisement) followed its own pattern of development passing the following stages: **Noun that I do not recognize (Stage 1) -- Word that I know more or less since having read the text (Stage 2) -- Noun I will understand what it means in context but I would not be able to use it in speech or writing (Stage 3) -- Noun I am very confident about. I would use it in speech and would be able to provide some grammar information (Stage 4) -- Noun I know. I am confident to use it in speech and writing. I would understand this word in listening/reading and would be able to spell it (Stage 5).**

Examples like these seem to indicate that certain words develop through a number of stages. The examples also show that a word might pick up new properties or features at each phase it goes through. For instance, judging on the subject's descriptions of her knowledge of the word "**Объявление**" (*advertisement*) at each stage of the experiment, this word acquired a large number of different features while progressing in the subject's lexicon. These features/descriptors appear to be the following: *unknown – known more or less – receptively known – well known (meaning, use and grammar knowledge are used as descriptors) –well known (productive and receptive skills as well as spelling knowledge are used as descriptors).*

Furthermore, the data indicate that a large number of target words acquired diffuse (or spread) among a large number of different categories rather than simply turn from unknown into known (Paribakht and Wesche's framework) (see the tables in studies 4, 5 and 6). However, this is not always the case. Some words appear to skip stages and move straight from "unknown" to "known". This suggests that the way/s the word develops in the learner's mind might depend on the specificity of the word. This is illustrated in Figure 9-2.

Figure 9-2: Example of some of the word relocations after reading (Exp 5)



This map shows the routes of relocation of some of the targeted words from initial category No15 "totally unknown" in classification 1 after the reading (study 5, chapter 7). As seen from this figure, a number of words moved straight into category 1 (classification 2): "known well". Some other words moved into category 7: "can guess the meaning on the grounds of the morphological structure of the word". Others moved into category 8: "can guess the meaning on the basis of associative knowledge". Others relocated into category 16: "known meanings, unsure about the grammar" or category 23: "Unknown nouns but can provide some grammar information". This is only one example of a large number of various patterns of relocations of words revealed in our research. This data illustrates the immense diversity of the L2 vocabulary acquisition and totally contradicts Paribakht and Wesche's concept on how word knowledge develops. According to them, the knowledge of any word is acquired and develops through a simple linear transition only:

Unknown – Known meaning – Can use in a sentence

Besides the specificity of the word itself, the ways in which the word develops in the learner's mind also seems to depend on the learner's individual abilities to acquire a certain word. Let us see how the same words are acquired by different learners. This can be seen from the patterns of word re-locations within categorizations created by different subjects (Table 9-2).

Table 9-2: Patterns of re-location of seven words after one session of reading

Words	Categorization created by Subject 1	Categorization created by Subject 2
Жаловаться To complain	Unknown -- Unsure of the exact meaning	Unknown -- Known
Опуститься To go down	Unknown -- Can take a guess based on similarities	Unknown -- Known
Общество Society	Unknown -- Known	Unknown -- Known well
Приключение Adventure	Unknown -- Can take a morph. guess	Unknown -- Known well
Желание Desire, wish	Unknown -- Known but not confident using	Unknown -- Known well
Молчание Silence	Can't remember -- Known but use is a problem	Unknown -- Can take a morph. guess
Капризный Capricious	Unknown -- Can take a guess based on similarities	Unknown -- Can take a morph. guess

The seven words are taken from the list of 200 target words used in studies 4 and 5. Subject 1 is our participant in study 4 (Chapter 6). Subject 2 is our testee in study 5 (Chapter 7). Both subjects were exposed to the same words and read the same story containing those words.

As indicated in the table, the same sample words progress differently in the lexicons of two different subjects. Subject 2 seems to acquire new words more readily than subject 1 who apparently requires further exposure to the same new words in order to acquire those words. This might be explained by the differences in the individual lexicon organizations as well as cognitive processes of these individuals, in general. It also seems that qualitative changes in the word positions within the learner's lexicon depend on the level of language proficiency of that particular learner: subject 2 is a far more experienced Russian language student than subject 1. Furthermore, the quantitative changes occurred in the subjects' categorizations after the reading also seem to confirm this point. This is shown in Table 9-3.

Table 9-3: Number of re-located words as reported by Subject 1 (Study 4) vs. Subject 2 (Study 5)

Subjects	No of Words that Progressed	No of Words that Stayed The Same	No of Words that Regressed
Subject 1	60	140	0
Subject 2	94	103	3

As shown in the table, subject 2 reported a considerably greater number of words improve their initial positions after a vocabulary input through reading.

Overall, the data presented in tables 9-2 and 9-3 seem to indicate that the number of words which relocated from their initial categories and the nature of these relocations are determined by the learner's experience. This seems to contradict the general assumption in the field that learners acquire words in the same way. Furthermore, this finding might suggest that, in regards to higher level learners, vocabulary acquisition is carried out at a higher rate.

In summary, based on our data, we can conclude that the process of word acquisition can be multi-directional. This seems to be determined by the specificity of a word as well as the learning peculiarities and experience of a learner. The patterns of movements of the words: **"Гусеница"** *(caterpillar)* and **"Объявление"** *(advertisement)* introduced earlier in this section indicate how totally unknown words develop in the mental lexicon by acquiring new features after each reading session. As the patterns illustrate, each of these words follows its own route on the way from "totally unknown" to "known". Multiple stages they pass may indicate the complexity of the process of vocabulary acquisition in many instances.

9.3 General discussion.

The complexity of learners' own categorizations and directions of further research

As stated in the introduction to this thesis, our research is very exploratory. Initially, we intended to investigate whether Russian language learners would be able to adequately assess their own knowledge of words. The results of the pilot study indicated that our informants did self-evaluate their knowledge of words accurately. However, exploring this issue, we came across rather interesting data in regards to the way/s learners describe their knowledge of the given words. This posed a number of questions in respect of the existing self-evaluation instruments of learners' own word knowledge. From that point, we continued to explore the issues suggested by the findings of our pilot study: the ways learners self-rate their knowledge of words. We decided to explore this issue in detail due to the fact that the data obtained in our pilot study appeared to be considerably richer than the possibilities of lexical self-assessment scales provided within the existing self-assessment methodologies in L2 vocabulary testing.

We challenged the existing methodologies of self-assessment of word knowledge, primarily Paribakht and Wesche's VKS, arguing that they do not reflect the complexity of learners' own descriptions. We ran a series of six main case studies (two small group studies followed by four single subject studies) each broadly investigating how L2 learners self-measure and rate their knowledge of words. While pursuing this goal we asked the subjects to categorize their knowledge of certain Russian words by arranging these words into groups and labeling these groups with their own category descriptions of how well they knew those words. We hypothesised that a number of self-report categories in which the learners would allocate the targeted words might be greater than a number of stages suggested in the existing self-assessment instruments for self-measuring learners' word knowledge. However, the actual results of our experimental studies exceeded our expectations. The subjects created considerably more word categories than we expected.

Inspired by these findings we took a step further. We decided to test whether directed exposure to certain words via reading would influence the way L2 learners categorize their knowledge of those words. Furthermore, having recorded some cases of word re-locations between categories in classification systems created at a different time of testing (study 3), we were curious to establish whether reading interventions would also impact on movements of words in subsequently created categorizations. This was investigated in chapters 4 through 6 in detail.

Originally, we assumed that learners would categorize their word knowledge by creating 6-8 groups of words which is greater than considered by Paribakht and Wesche. In fact, our subjects used up to 26 self-report categories to distinguish between various levels of their knowledge of the given words. This is five times as many as reflected in Paribakht and Wesche's VKS which is claimed to be the most elaborate (out of the current self-assessment instruments in the field) and most widely used self-assessment methodology for measuring the depth of learners' knowledge of words. Moreover, it should be emphasized that a large number of categories suggested by our subjects are based on the features or word properties that have never been included or even referred to within the existing self-assessment methodologies.

The repeated longitudinal study carried out in our research (Study 6, chapter 8) revealed the immense complexity of the process of re-categorization that occurred within the learners'

classification systems. Our longitudinal study also showed an enormous diversity of the word-relocation patterns recorded after sessions of reading. This unexpected data obtained in our research indicates that a new repeated longitudinal study would be required in order to check the findings reported in the present research. This new repeated longitudinal study should focus on exploring the processes of self-measuring and self-classifying L2 vocabulary knowledge during a considerably longer period of time (e.g. 1 to 2 years). It would be of special interest to investigate quantitative and qualitative changes which might occur in learners' categorization systems over this period of time. It would also be beneficial for future research to examine possible cases of regression in learners' lexical knowledge as well as to further explore the common routes of re-locations of words.

However, talking about longitudinal studies we should also be aware of a possible effect of repeated testing. For instance, in our repeated longitudinal study which provided rather interesting data in regards to the continuously changing categorization systems suggested by our learner, there is always the possibility of experimental effects for a repeat test taker. This implies the following. Although the subject remained unaware of the aim of our study, she may have come to a certain conclusion in respect of the experimental goals and procedure. This is illustrated by the descriptions for some of the categories suggested by the informant after the reading ("Words that I know well and I knew them prior to the reading of the text"). Furthermore, a number of categories repeated in all or most of the subsequent categorization systems might imply that the subject simply re-used some of the categories that she remembered from the previous testing.

The results of these studies indicated that the way our subjects measure their own knowledge of words is considerably more extensive, complex, varied and multi-featured than we expected. Furthermore, the categorization systems built by our learners appear to be rather unstable with lexical items persistently relocating between categories as learners' knowledge of the given words changes.

The implications of these findings seem to be as follows:

1)The data obtained in our research indicate that self-evaluation of L2 lexical knowledge might not be carried out in the way the existing self-assessment methodologies assume. The Vocabulary Knowledge Scale by Paribakht and Wesche is the main tool that has been

used (or referred to) for self-assessment of L2 vocabulary knowledge. Paribakht and Wesche are core references in this area. Most researchers take their VKS methodology for granted. This thesis, however, does question Paribakht and Wesche's VKS. We argue that the VKS methodology makes wrong assumptions. Paribakht and Wesche suggest only a few self-assessment categories. They assume that they know what self-assessment categories are. They assume that all L2 learners categorize words in the same way. Thus the VKS self-report categories are assumed to be the same for all learners. Paribakht and Wesche also assume that the categories are fixed (or stable). The thesis shows that none of these assumptions hold up. Where does it leave us? It severely restricts the value of the VKS and the area of its application. Basically, the VKS is unable to provide full comprehensive information on how well a particular word is known (i.e. it does not measure the knowledge of a word in depth). Our research shows that it does not appear to be going to work when learners are placed in fixed and limited self-assessment conditions. Furthermore, the fact that words behave in different ways while being re-categorized by the subjects indicates that it does not seem to be possible to fit the words tested into the limited pre-arranged self-report categories.

Overall, we argue that the VKS methodology has a rather limited use and can be regarded as an instrument plainly for evaluating breadth (or size) of word knowledge.

2)The evidence gathered in our studies and discussed within the issues 9.2.1 through 9.2.4 in the previous section indicate that the process of development of L2 word knowledge might not occur the way everybody in the field assumes. The data allow us to assume that there is a possible link between the way/s L2 learners self-measure and categorize their knowledge of words and the processes of lexical acquisition in their mental lexicons. The major evidence in favor of this assumption is the fact that rather complex and multi-featured categorization systems created by L2 learners appear to be very unstable i.e. constantly changing as learners' knowledge of the words was changing. If categorization systems built by L2 learners when they evaluate their knowledge of words mirror the processes of vocabulary acquisition that occur in their lexicon this would imply that self-assessment methodology can be used to explore the structure of the lexicon (similarly to word association tests used to investigate lexical networks by analyzing L2 learners' word association behavior).

However, from another perspective, the extremely complex categorization systems created in

our studies may reflect our subjects' individual styles of categorizing and their characterological traits (creativity, for instance). It might also be determined by the level of meta-linguistic knowledge possessed by a learner tested. In other words, meta-linguistic knowledge of our subjects may be considered as a possible factor of their abilities to evaluate their knowledge of words. It has been found that learners' L2 proficiency is correlated with their levels of meta-linguistic knowledge (Roehr, 2007). Moreover, meta-linguistic knowledge and language proficiency are assumed to constitute two separate factors of linguistic ability. Our study shows a relation between language proficiency and the ways learners assess their knowledge of words. This implies that learners at low levels of meta-linguistic knowledge may measure and classify their knowledge of words differently from those at higher levels. The informants in our studies were at an intermediate or even advanced level of Russian language proficiency. This may suggest that their L2 language analytic ability as well as good knowledge of meta-language (linguistic terms, for instance) had some impact on the way they measured and categorized their knowledge of words. In this sense, our informants did have knowledge and instruments at their disposal to accurately describe their L2 lexical knowledge. However, as noted earlier, other groups of L2 learners, for instance children or those who have forgotten all the meta-language they ever knew, might describe their lexical knowledge as fully and accurately as our informants did.

In this respect, it might be worth considering (for further research into this issue) the use of think-aloud protocols. This may assist in understanding how different groups of L2 learners gradually (step by step) build categorization systems of their lexical knowledge. Also, think-aloud protocols might provide further evidence in relation to possible links between learners' categorization processes and the organization of their L2 mental lexicons.

.

Overall, although the findings of our studies require further investigating on a considerably larger number of subjects over a significantly longer period of time it is readily apparent that these findings indicate a new approach to self-assessment in L2 vocabulary testing as well as a new view on the processes of L2 vocabulary acquisition in general. The data obtained in this research is rather important for developing a more effective and fine-grained lexical self-assessment instrument. The diversity of the categorization systems suggested by different learners at a different time of testing suggests that different people might acquire L2 lexical knowledge in rather different ways (i.e. different models of vocabulary acquisition). This in

turn suggests that different learners may require different types of self-assessment scales (to match their individual SLVA models) to adequately measure their lexical knowledge. The question is: What will we need to include in a L2 self-assessment vocabulary test? The answer to this question is not easy. As shown in this thesis our research raised a number of issues (summarized in the current chapter 9) which need to be addressed and overcome before we can go any further. Why do these issues matter? The data obtained in this research posed a question whether the L2 vocabulary acquisition occurs in a standard route (from unknown to known) or whether there is a variety of different scenarios of acquiring a word.
Since the evidence collected in this research highlights the enormous complexity of the process of L2 lexical acquisition we cannot provide a certain answer to this question yet. The next step along the route towards a new methodology of self-evaluating L2 lexical knowledge would be resolving the issues outlined and discussed in this chapter.

9.4 Conclusions

In this chapter, we considered a number of rather important issues that have not been picked up in research on vocabulary self-assessment. We attempted to show the importance of our findings for L2 lexical self-/assessment as well as vocabulary acquisition in general. We argued that the existing self-assessment methodologies do not reflect the complexity of subjects' own descriptions of their word knowledge, neither do they take into account the complexity of re-structuring already created classification systems including continuous word re-locations within these systems. We also attempted to prove that in light of these findings, a new self-assessment methodology based on an individual approach would be required to embrace an enormous complexity of learners' own classification systems. We emphasized that future studies of this type should be carried out as repeated longitudinal studies and involve a larger number of subjects.

CHAPTER 10

Conclusion

In this thesis, we were exploring how learners of Russian as a foreign language self-assess their knowledge of words. We asked a set of simple questions such as:

- Which criteria are considered by learners when they assess their own knowledge of words?
- Do learners entirely focus on their meaning knowledge (Paribakht and Wesche's approach) or do they look into the other aspects of word knowledge?
- How do different degrees of certainty about the knowledge of a target word (which received a lot of attention in Paribakht and Wesche's scale) manifest themselves in self-assessment categories created by learners?

As usual with theses, these questions turned out to be more complex than we had expected. In our case this problem was exacerbated by the fact that we were unable to run traditional large scale group experiments with our subjects. Nevertheless, the work reported in this thesis has raised a number of significant issues concerning the way that learners of Russian self-categorize their knowledge of Russian words. The main findings of our research are:

- Learners measure and classify their knowledge of words in a rather complex and varied way. A large variety of features are considered and chosen by learners as criteria for their categorizations. These features are referred to both in positive and negative context which suggests that learners carry out a rather thorough analysis of their knowledge of words.
- Descriptors of lexical knowledge used by learners to categorize their knowledge of words fall into two main groups: familiarity and certainty in knowledge. The familiarity group contains a wide range of word knowledge aspects referred to by a learner while classifying their knowledge of words. The certainty group embraces a variety of degrees and levels of confidence in knowledge including wiliness to guess.
- The research revealed a rather large distribution of features among the created categories in learners' classification systems with up to seven different descriptors used to characterize a single category. This might indicate that different properties of a word are interrelated in the lexicon.

- The categorization systems suggested by different learners vary considerably in terms of their quantitative and qualitative characteristics. This hints at the necessity of an individual approach within L2 lexical self-assessment methodologies.
- The classification systems created by learners seem to be very unstable with re-categorization taking place at each stage of testing: new categories become added to the system, some of the existing categories split, become altered or disappear from the system.
- The research indicated continuous re-location of words among the suggested categories within learners' classification systems. The patterns of these movements considerably vary from subject to subject and from word to word.

Anyone who is familiar with the main stream in L2 vocabulary acquisition will recognize that this list contains a number of features which are not normally present in research on lexical self-assessment. The implication of this is that we may need to re-assess the current models of vocabulary acquisition. Unfortunately, a major re-assessment of this sort is beyond the scope of a single thesis.

We hope that this thesis will inspire other researchers to look again at some of the issues raised in this work and test the existing L2 self-assessment approaches for other unusual languages.

APPENDIX

Pilot Study

Self-Assessment vs. Levels Test Assessment

A.1 Introduction

This is a brief report on our study which investigates the issue of L2 lexical self-assessment. In light of the controversial results reported by different scholars (and presented in chapter 1 of this thesis) regarding the validity and reliability of L2 self-assessment we were curious to explore how learners of Russian would assess their own knowledge of words. Following the encouraging findings of the research on self-assessment of L2 learning abilities by Oscarson, 1978; Hargan, 1994; Birckbichler et al., 1993; Deville and Deville, 1999; Brantmeier, 2006; Paribakht and Wesche, 1993-1997; Wan-a-rom, 2010 and others, we carried out this study in order to check whether self-assessment of vocabulary knowledge can be regarded as a reliable alternative to traditional assessment. Thus in the current study, we hypothesised that self-assessment can be used as an indicator of lexical knowledge.

Thus the study reported here was questioning the reliability of self-assessment of L2 lexical knowledge as a measuring instrument. Since it did not appear to be possible (within the scope of this thesis) to test L2 learners' ability of assessing their own knowledge of various aspects of lexical knowledge, we decided to focus on the meaning knowledge. We also intended to investigate how learners themselves would describe their knowledge of words.

Hence, the main goal of the first study was to establish whether learners of Russian would be able to assess their meaning knowledge of the target Russian words. It was accompanied by the second target of this study which was an investigation of learners' approach to their knowledge of the words that were reported by them as familiar. We also hoped that this information supplied by the learners themselves, might shed light on the issue of their perception of their meaning knowledge. The research question posed in this study was as follows.

Research Question:

Will learners of Russian be able to accurately evaluate their own knowledge of words?

A.2 Study

A.2.1 Method

A.2.1.1 Target words

The targeted words were selected from the Russian word frequency list (Vakar, 1966; also checked with Аграев et. al., 1977). We took 15 words from 2k, 15 words from 3k and 15 words from 5k. The total number of the Russian target words was 45: 15 verbs, 15 nouns, 11 adjectives and 4 adverbs. Each word was typed onto a separate card. Hence, there were 45 cards in total. The target words were as follows: *мальчик (boy), нельзя (it is not allowed), нужно (it is necessary), народ (people), дедушка (grandfather), завтра (tomorrow), умный (clever), столовая (dining room, self-service canteen), сильный (strong, severe), повторить (to repeat, to revise), начинать (to begin), предложить (to propose, to offer), приятный (pleasant), горячий (hot), низкий (low, short), белый (white), гость (guest), имя (name), тётя (aunt), уезжать (to leave), спрашивать (to ask), звонить (to ring), вернуться (to return), учиться (to study), послать, мало, час, комната, дорогой, красивый, забыть, надеяться, расти, завидовать, следовать, получать, встреча, успех, порядок, опыт, огонь, больной, деревня, счастливый* and *чистый.*

A.2.1.2 Participants

The participants of the current study were 32 adults which had been studying Russian as a foreign language at Cardiff University. Students' proficiency levels ranged from beginners - Year 1 (15 people) through lower intermediate – Year 2 (8 people) to upper intermediate - Year 3 (9 people). The subjects represented different cultures and backgrounds. Though most of the participants were British, with English as their first language, Italian, Polish, French and Portuguese were also represented as native languages.

A.2.1.3 Procedure

The study consisted of two stages. At the first stage, the subjects were asked to indicate which of the target words were familiar (in terms of their meaning) to them. The participants were required to choose the words they knew the meaning(s) of, from a pile of 45 cards with the target Russian words written on them (one word per card). They were also required to compile the heading for the group of the chosen words (i.e. describe their knowledge of the selected words). The instruction was: "Select the words that you know. Write the heading for the group of words you have chosen". Each participant was awarded one point for each word

indicated by them as known. Subjects were not restricted in time. They were not allowed to consult dictionaries or discuss the target words.

The self-assessment was followed by the teacher's assessment at the second stage of the study. We used the Levels Test format to evaluate the subjects' actual word knowledge. This test was administered after the self-assessment task had been completed. The participants were asked to complete the modified Levels Test based on the same 45 targeted words and arranged as a multiple-choice test. Subjects were required to choose the Russian equivalent for each of the three English words (i.e. match words with their translations). The instruction was: "Next to each English translation, write the number of the Russian word it corresponds to". The following example was offered to illustrate the task.

Example: 1 человек
2 гостиница
3 друг
4 игра
5 вместе
6 отчество

together **5**
patronymic **6**
game **4**

Taking this example as one item, the whole test consisted of 15 items of this sort, each testing three target words. There was no time restriction. Subjects were not allowed to discuss the task or consult reference literature. Each correct answer was awarded 1 point (i.e. one point for each word that was correctly matched with its translation).

A.2.2 Results

The research question asked in this study along with the main goal of the study address learners' abilities to accurately assess their knowledge of words. Pursuing this target we asked the participants to assess their own knowledge of 45 targeted words. The summarized results for the three groups of subjects are shown in table A-1.

Table A-1: Mean scores of the self-assessment vs. teacher's assessment

Group	SA2k	L2k	SA3k	L3k	SA5k	L5k	SATot	LTot
Yr 1 n=15	7.00	7.00	8.93	8.47	3.47	3.33	19.40	18.80
Yr 2 n=8	9.50	9.25	11.75	11.63	4.00	3.50	25.25	24.38
Yr 3 n=9	11.56	11.33	12.33	12.11	8.22	8.33	32.11	31.78

Note. SA -- self-assessment, L – Levels Test, T – total

Table A-2: Total Mean scores for self-assessment, assessment and difference

Group	SA	SD	L	SD	Difference
Yr 1	19.40	5.29	18.80	5.56	.60
Yr 2	25.25	7.27	24.38	8.05	.87
Yr 3	32.11	10.90	31.78	11.60	.33

Note. SA -- self-assessment, L – Levels Test, SD – Standard Deviation

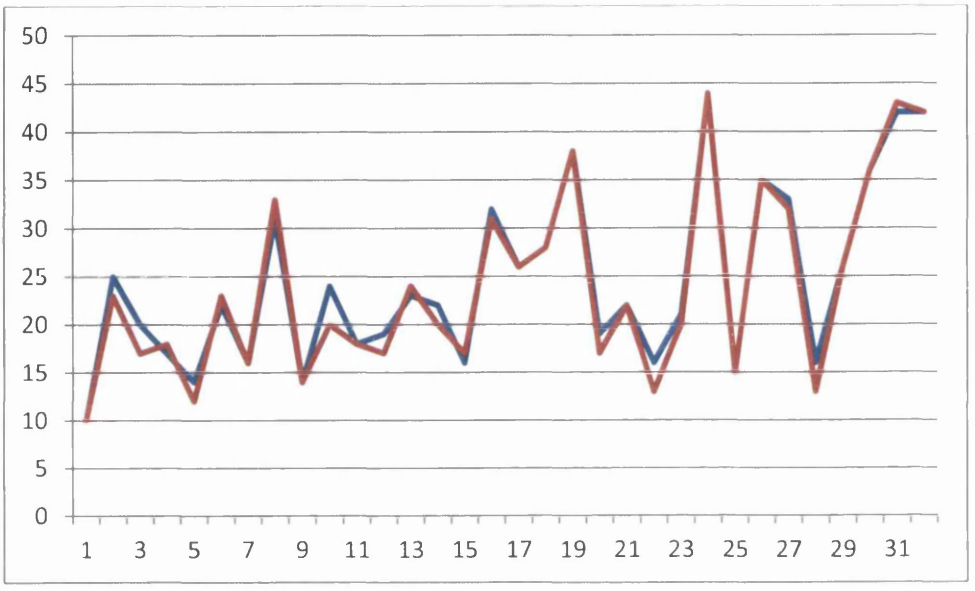
The data show that subjects were generally very good at evaluating their own knowledge of the given words. It is seen from the table that the SA and L values are very close. However, scores for the Level 3k words are considerably higher than we might have expected, and the data do not show the monotonic decline over frequency. It is also obvious that the scores drop off dramatically for Level 5 words. The standard deviations indicate that there are some differences (variation) in the data submitted within each of the three groups of subjects. Furthermore, as seen from the table this difference increases with the level of proficiency.

Pursuing the main goal of our study, we compared the results of the self-assessment against the Levels test data. In order to carry out this comparison, Pearson's correlation test and paired T-test were administered in this study. We did not run ANOVA in this study due to the small number of subjects.

The tables clearly indicate that the mean numbers of the self-assessment are very close to the mean numbers of the actual assessment. This was confirmed by the statistical analysis:

Pearson's correlation coefficient **r** = .989 which suggests a very strong positive relationship. The correlation was statistically significant: p<.001. The results of the paired T-test were as follows: t = 2.343; df = 31; st. er. of dif. = .253. A good match between the subjects' self-assessment and assessment scores is also illustrated by figure A-1.

Figure A-1: Self-assessment vs. Assessment



These results will be further discussed in the section that follows.

With regards to the second target of this study -- learners' self-perception of the known words, the subjects provided the following descriptions of their knowledge.

Subjects' labels for the group of words they selected as known:

"Familiar" (3 submitted) or "very familiar with" (1);

"I think I know these words" (3 submitted);

"I know these words" (9 submitted);

"I know what these mean";

"Words I know";

"These words I believe to be nouns..., verbs..., adjectives...etc";

"Known with confidence";

"Words I have come across before, I know or feel like I know...";

"I am pretty sure I know these words";

"Words which are very familiar and I could translate correctly";
"Know (very) well" (2 submitted);
"I recognize these and know their meanings";
"Recognize these words" or "I recognize these words without doubt";
"I know exactly what these words mean and how to use them";
"I know the meaning and spelling of these words";
"I know the meaning of these words"; or "I know what these words mean";
"Words I know in order in which I learnt them".
This will be discussed in the next section.

A.3 Discussion

Addressing the two goals of our study, in this section, we will discuss the following two issues. The first issue is the reliability of the lexical self-assessment we conducted in this study. The second issue is the ways that students labeled the groups of words they selected as known.

A.3.1. Self-assessment vs. assessment

Pursuing the first goal of the study we asked the participants to assess their own knowledge of the 45 given words. They were required to choose the words (out of the 45 targeted words offered for the assessment) they knew. Though the general results of the Levels test, in terms of demonstrated meaning knowledge, appear to be a little disappointing (especially for Level 5 words), they generally matched the data received from the self-assessment task.

Tables A-1 and A-2 indicate how self-assessment results related to the results of the actual assessment. As seen from the tables, on most occasions, subjects successfully assessed their knowledge (meaning knowledge) of the given words. Even visual inspection of the means presented in tables A-1 and A-2 shows a close match between the self-assessment and assessment mean scores. Figure A-1 also illustrates an obvious correspondence between the self-assessment and assessment curves. Furthermore, this is supported by the statistical analysis. A strong positive relationship between the self-assessment and actual (via a modified version of the Levels Test) assessment of the subjects' lexical knowledge is

confirmed by the results of Pearson's correlation test ($r = .989$; $p < .001$) and paired T-test ($t = 2.343$; $df = 31$; st. er. of dif. $= .253$).

Thus, learners of our study appeared to be able to accurately evaluate their own knowledge of words. This suggests (though we are avoiding generalization) that this type of assessment might compliment or even replace traditional assessment in many instances: subject to further research.

A.3.2 Labeling the group of selected words

Pursuing the second target of this study the subjects were asked to label the group of words they selected as known in the self-assessment task. Though the instruction required all the learners to choose the words they knew, suggested headings for the groups of words they selected appeared to be very different. In general, all the labels submitted may be classified according to the following two criteria: familiarity (i.e. meaning knowledge in connection with other aspects (types) of word knowledge) and confidence (i.e. degrees of certainty in the meaning knowledge).

With regards to the first criterion, headings created by the subjects seem to suggest that for these learners, knowledge of words might imply more than just knowledge of their meaning. Furthermore, labels like these: "I know exactly what these words mean and how to use them", "I know the meaning and spelling of these words" or "These words I believe to be nouns..., verbs..., adjectives...etc" might indicate that learners evaluate their meaning knowledge in close connection with other aspects (or types) of word knowledge. This might also suggest the existence of links among different kinds of word knowledge.

As for the second criterion, judging on the subjects' label descriptions, the following degrees of certainty in the subjects' knowledge may be distinguished (in descending order):

1) known very well or known with confidence;
2) familiar (known);
3) I think I know;
4) I feel like I know...

This shows that even "known words" could be graded at different levels depending on

learners' perception of their knowledge. Furthermore, the fact that most subjects did not blindly copy the wording of the instruction "the words you know", but altered it in order to reflect their perception of "known words" might be considered as an argument in favour of exploring how L2 learners themselves measure their own word knowledge.

Generally speaking, the current study revealed that while assessing their knowledge of given words, our learners tended to consider a wide range of various features including different aspects of word knowledge, such as spelling, word class belonging and use in speech/writing. This is what one could expect to be included in self-assessment scales evaluating learners' knowledge of words.

A.4 Conclusion

A comparison of the results obtained in this study has revealed the apparent match between the subjects' self-assessment of their word knowledge and their actual performance on the Levels Test. Although we cannot generalize the results obtained in a single study we found that measuring learners' L2 vocabulary knowledge by means of self-assessment might be a good alternative to traditional evaluating instruments.

This study also revealed that learners tend to consider different features while describing their knowledge of words. This raised a number of new issues in relation to the way/s that L2 learners measure and categorize their knowledge of words. That was investigated in the main part of our thesis.

BIBLIOGRAPHY

Achara, W. 1980

Self-assessment in English skills by undergraduate and graduate students in Thai Universities. In J. Read (Ed.) *Directions in language testing*. Singapore: Singapore University Press, 240-60.

Anderson, P. 1982

Self-esteem in the foreign language: A preliminary investigation. *Foreign Language Annals*, 15, 109-114.

Bachman, L., and A. Palmer. 1981

The construct validity of the FSI Oral Proficiency Interview. *Language Learning* 31, 67-86.

Barrow, J., Nakanishi, Y., and H. Ishino. 1999

Assessing Japanese college students' vocabulary knowledge with a self- checking familiarity survey. *System* 27, 223-247.

Beck, I., Perfetti, C., and M. McKeown. 1982

Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 71(1), 506-521.

Birckbichler, D., Corl, K., and C. Deville. 1993

The dynamics of language program testing: implications for articulation and program revision. The Dynamics of Language Program Direction. Heinle & Heinle, Boston.

Blanche, P., and B. Merino. 1989

Self-assessment of foreign language skills: implications for teachers and researchers. *Language Learning* 39, 313-40.

Blanche, P. 1988

Self-assessment of foreign language skills: Implications for teachers and researchers. *RELC Journal*, 19 (1), 75-93.

Boud, D. 1995

Enhancing learning through self-assessment. London, Philadelphia: Kogan Page.

Brantmeier, C. 2005

Non-linguistic variables in advanced L2 reading: learner's self-assessment and enjoyment. *Foreign Language Annals* 38 (4), 493-503.

Brantmeier, C. 2006

Advanced L2 learners and reading placement: Self-assessment, computer-based testing, and subsequent performance. *System*, 34 (1), 15-35.

Buck, G. 1992

Listening comprehension: construct validity and trait characteristics. *Language Learning,* 42, 313-57.

***Collins Dictionary and Thesaurus*. 2000**

Harper Collins Publishers. Glasgow: Omnia Books Ltd.

Cronbach, L. J. 1942

Measuring knowledge of precise word meaning. *Journal of Educational Research*, 36, 528-534.

Daller, H., Milton, J. & Treffers-Daller, J. (eds). 2007

Modelling and assessing vocabulary knowledge. Cambridge: CUP.

Dale, E. 1965

Vocabulary measurement: Techniques and major findings. *Elementary English*, 42, 895-901.

Deville, M., and C. Deville. 1999

Computer adaptive testing in second language contexts. *Annual Review of Applied Linguistics* 19, 273-299.

Dickinson, L. 1987

Self-instruction in language learning. London: CUP.

Duff, P. A. 2007

Case study research in applied linguistics. Lawrence Erlbaum Associates, New York, London.

Ekbatani, G. 2000

Moving toward learner-directed assessment. In G. Ekbatani and H. Pierson (Eds.) *Learner-directed assessment in ESL.* New Jersey, 1-11.

Evans, M., Markowitz, J., Athlswede, T., and K. Rossi. 1986

Digging in a dictionary: Building a relational lexicon to support natural language processing applications. Paper presented at the Computers in Language Research Conference, Urbana, IL.

Evers, R. 1981

Self-evaluation of proficiency levels in French as a second language. Paper presented at the 6th International Congress of Applied Linguistics, Lund, Sweden.

Ferguson, N. 1978

Self-assessment of listening comprehension. *International Review of Applied Linguistics* 16, 149-56.

Ferris, D. 1982

The influence of continuous self-evaluation of speaking skills on teaching methodology. In D. Coste (Ed.) *Contributions à une renovation de l'apprentissage et de l'enseignement des langues.* Strasbourg, France.

Ferris, D. 1983

The influence of the continuous self-evaluation of oral skills on language learning methodology. In D. Coste (Ed.) *Contributions to a renewal of language learning and teaching: some current work in Europe.* Strasbourg, France.

Fok, A. C. Y.Y. 1981

Reliability of student self-assessment. Hong Kong: Hong Kong University Language Centre.

Folse, K. 2006

The Effect of Type of Written Exercise on L2 Vocabulary Retention. *TESOL Quarterly* 40, 2, 273-294.

Fulcher, G. 1988

Lexis and reality in oral evaluation. Revised and expanded version of a paper presented at the annual meeting of the International association of Teachers of English as a Foreign Language, Edinburgh, Scotland.

Gall, M.D, Borg, W.R., & Gall, J.P. 2003

*Educational research: an introduction (7th edition).*White Plains, NY: Longman.

Gass, S. 1989

Second language vocabulary acquisition. *Annual Review of Applied Linguistics*, 9, 92-106.

Hargan, N. 1994

Learner autonomy by remote control. *System* 22 (4), 455-462.

Heilenman, K. 1991

Self-assessment and placement: a review of the issues. In Teschner, R (Ed.), *Assessing Foreign Language Proficiency of Undergraduates*, AAUSC Issues in Language Program Direction. Heinle & Heinle, Boston, 93-114.

Henriksen, B. 1999

Three dimensions of vocabulary development. *Studies in Second Language Acquisition 21, 303-317.*

Holec, H. 1979

Autonomy and foreign language learning. Oxford: Pergamon Press.

Hornby, A. 1978

Oxford student's dictionary of current English. Oxford: OUP.

Janssen-van Dieten, A-M. 1989

The development of a test of Dutch as a second language: the validity of self-assessment by inexperienced subjects. *Language Testing* 6, 30-46.

Krausert, S. 1991

Determining the usefulness of self-assessment of foreign language skills: post-secondary ESL students' placement contribution. Ph.D. Diss., University of Southern California.

Laufer, B., and Y. Yano. 2001

Understanding unfamiliar words in a text: do L2 learners understand how much they don't understand? *Reading in a Foreign Language* 13, 2, 549-566.

Le Blanc, R., and G. Pinchaud. 1985

Self-assessment as a second language placement instrument. *TESOL Quarterly* 19, 673-87.

Lewkowicz, J., and J. Moon. 1985

Evaluation: a way of involving the learner. In Lancaster practical papers in English language education, Volume 6: Evaluation, J. Alderson (Ed.), Oxford: Pergamon Press, 45-80.

McNamara, M., and D. Deane. 1995

Self-assessment activities toward autonomy in language learning. *TESOL Quarterly*, 5, 18-23

Meara, P.M. 1989

Matrix models of vocabulary acquisition. *AILA Review*, 6, 66-74.

Meara, P.M. 2006

Emergent properties of multilingual lexicons. *Applied Linguistics* 27, 4, 620-644.

Meara, P.M., and B. Buxton. 1987

An alternative to multiple choice vocabulary tests. *Language Testing*, 4, 142-154.

Meara, P.M., and G. Jones. 1988

Vocabulary size as a placement indicator. In P. Grunwell (Ed.), *Applied linguistics in society* (pp. 80-87). London: CILT.

Meara, P.M., and G. Jones. 1990

Eurocentres Vocabulary Size Test: User's Guide. Zurich: Eurocentres.

Meara, P.M., and J. Milton. 2003

X_Lex, The Swansea Levels Test. Newbury: Express.

Meara, P.M., & I. Rodriguez Sanchez. 2001

A methodology for evaluating the effectiveness of vocabulary treatments. In: M Bax, and J-W Zwart (eds.), *Reflections on Language and Language Learning*. Amsterdam: Benjamins. 267-278.

Meara, P.M., & B. Wolter. 2004

V Links: Beyond Vocabulary Depth. *Angles on the English Speaking World* 4, 85-97.

Milton, J. 2009

Measuring second language vocabulary acquisition. Bristol: Multilingual Matters.

Nation, I.S.P. 1990

Teaching and learning vocabulary. Boston: Heinle & Heinle.

Nation, I.S.P. 2001

Learning vocabulary in another language. Cambridge: Cambridge University Press.

North, B. 2000

Defining a flexible common measurement scale: descriptors for self and teacher assessment. In G. Ekbatani and H. Pierson (Eds.) *Learner-directed assessment in ESL.* New Jersey, 13-47.

Oscarson, M. 1989

Self-assessment of language proficiency: Rationale and applications. *Language Testing*, 6, 1-13.

Oskarsson, M. 1978

Approaches to self-assessment in foreign language learning. Strasbourg: Council of Europe.

Oskarsson, M. 1980

Approaches to self-assessment in foreign language learning. London: Pergamon Press.

Oskarsson, M. 1984

Self-assessment of foreign language skills: a survey of research and development work. Strasbourg: Council of Europe.

***Oxford Russian dictionary*. 2000**

Third edition. Oxford: OUP

Palmer, A., and L. Bachman. 1981

Basic concerns in test validation. In J. Alderson and A. Hughes (Eds.) *ELT documents 3: Issues in language testing.* London.

Paribakht, T., and M. Wesche. 1993

Reading comprehension and second language development in a comprehension-based ESL Program. *TESL Canada Journal* 11,1, 9-29.

Paribakht, T., and M. Wesche. 1995

Second language vocabulary acquisition through reading and text-based exercises. In Courchene, R. , Burger, S. , Cornaire, C., LeBlanc, R., Paribakht, S. and Seguin, H. (Eds.). *25 years of second language teaching at the University of Ottawa.* Ottawa: Second Language Institute, University of Ottawa, 44-67.

Paribakht, T., and M. Wesche. 1996

Enhancing vocabulary acquisition through reading: a hierarchy of text-related exercise types. *The Canadian Modern Language Review*, 52, 1, 155-178.

Paribakht, T., and M. Wesche. 1997

Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady and T. Huckin (eds.) *Second language vocabulary acquisition: a rationale for pedagogy.* Cambridge: Cambridge University Press, 174-200.

Peirce, B., Swain M., and D. Hart. 1993

Self-assessment in two French immersion programmes. *Applied Linguistics*, 14, 25-42.

Postman, L. 1970

The California norms: Association as a function of word frequency. In L. Postman & G. Keppel (Eds.), *Norms of word association.* New York: Academic Press, 241-320.

Raasch, A. 1979

To evaluate oneself: Is that a neologism? *Le Français dans le Monde,* 149, 63-67.

Rea, P. 1981

Formative assessment of student performance: The role of self-appraisal. *Indian Journal of Applied Linguistics*, 7, 66-88.

Read, J. 1988

Measuring the vocabulary knowledge of second language learners. *RELC Journal*, 19 (2), 12-25.

Read, J. 1989

Towards a deeper assessment of vocabulary knowledge. Washington, DC: ERIC Clearinghouse on Languages and Linguistics. (No. ED 301 048)

Read, J. 1993

The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10, 355-371.

Read, J. 1994

Refining the word associates format as a measure of depth of vocabulary knowledge. Paper presented at the 19th Annual Congress of the ALA of Australia, Melbourne.

Read, J. 2000

Assessing vocabulary. Cambridge: CUP.

Richards, J. 1976

The role of vocabulary teaching. *TESOL Quarterly* 10, 77-89.

Rivers, W. 2001

Autonomy at all costs: An ethnography of metacognitive self-assessment and self-management among experienced language learners. *The Modern Language Journal*, 85, 279-290.

Roehr, K. 2007

Metalinguistic knowledge and language ability in University-level L2 learners. *Applied Linguistics 29/2, 173-199.*

Ross, S. 1998

Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15 (1), 1-21.

Rott, S. 2005

Processing Glosses: A Qualitative Exploration of how Form-Meaning Connections are Established and Strengthened. *Reading in a Foreign Language* 17, 2, 95-124.

***Russian-English dictionary*. 1989**

Smirnitsky A. (Ed.), Moscow: Russky Yazyk Publishers.

Schärer, R. 1983

Identification of learner needs at Eurocentres. In Richterich, R. (Ed.), *Case studies in identifying language needs.* Oxford: Pergamon Press.

Schmitt, N., and P.M. Meara. 1997

Researching vocabulary through a word knowledge framework: word associations and verbal suffixes. *Studies in Second Language Acquisition* 19, 1, 17-36.

Schmitt, N. 1998

Tracking the incremental acquisition of second languag vocabulary: a longitudinal study. *Language Learning* 48, 2, 281-317.

Schraw, G., Bruning, R., and C. Svoboda. 1995

Sources of situational interest. *Journal of Reading Behaviour*, 1-17.

Stake, R. 1995

The art of case study research. Thousand Oaks, London, New Delhi: Sage.

Stolz, W. S., and J.Tiffany. 1972

The production of "child like" associations by adults to unfamiliar adjectives. *Journal of Verbal Learning and Verbal Behaviour, 11, 38-46.*

Strong-Klause, D. 2000

Exploring the effectiveness of self-assessment strategies in ESL placement. In G. Ekbatani and H. Pierson (Eds.) *Learner-directed assessment in ESL.* New Jersey, 49-73.

Tobias, S., and H.T.Everson. 1998

Research on the assessment of metacognitive knowledge monitor presented at a Symposium on "Metacognition: Assessment and Training", American Educational Research Association, San Diego, CA.

Tzeng, W-T., Dornyei, Z., and N. Schmitt. 2006

A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics* 27/1, 78-102.

Vakar, N. 1966

A word count of spoken Russian. US: Ohio State University Press.

Van Passel, F. 1974.

L'enseignement des langues aux adultes. Paris: Nathan.

Vermeer, A. 2001

Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics 22,* 217-234.

Von Elek, T. 1981

Self-assessment of Swedish as a second language. (*Sjalvbedomning av fardigheter i svenska som andra språk*). Goteborg, Sweden: University of Goteborg, LTRC.

Von Elek, T. 1982

Test of Swedish as a second language: an experiment in self-assessment. Goteborg: Goteborg Universitet.

Von Elek, T., Lee, Y., and A. Fok. 1985

New directions in language testing. Oxford: OUP. 47-57.

Wan-a-rom, U. 2010

Self-assessment of word knowledge with graded readers: a preliminary study. *Reading in a Foreign Language* 22, 2, 323-338.

Wesche M., and T. Paribakht. 1996

Assessing second language vocabulary knowledge: depth versus breadth. *The Canadian Modern Language Review*, 53, 1, 13-40.

Wilks, C., Meara, P. and Wolter B. 2005

A further note on simulating word association behaviour in a second language. *Second Language Research, 21(4),* 359-372.

Wilks, C. and Meara, P. 2007

Graph theory and words association networks. In H. Daller, J. Milton and J. Treffers-Daller (eds) *Modelling and Assessing Vocabulary Knowledge.* Cambridge: CUP, 167-181.

Wolter, B. 2001

Comparing the L1 and L2 Mental Lexicon. *Studies in Second Language Acquisition*, 23, 41-69.

Wolter, B. 2005

V_Links: A new approach to assessing depth of word knowledge. Unpublished PhD Dissertation, University of Wales Swansea.

Yin, R.K. 2003

Case Study Research: Design and Methods (3rd edition). Thousand Oaks, London, New Delhi: Sage.

Zareva, A. 2007

Structure of the L2 mental lexicon: how does it compare to native speakers' lexical organisation? *Second Language Research* 23, 2, 123-152.

Аграев. А, Бородин. В. 1977

Частотный словарь русского языка. Засорина. Л. (Ред.), Москва: Русский Язык.

Чехов, А. 1976

Дама с собачкой. In *Рассказы.* Москва: Художественная литература, *332-347.*

Чехов, А. 1976

Крыжовник. In *Рассказы.* Москва: Художественная литература, 271-280.