TEST TAKERS' LISTENING COMPREHENSION SUB-SKILLS AND STRATEGIES

doi.org/10.61425/wplp.2010.04.59.85

Éva Barta
Budapest Business School, College of Commerce, Catering and Tourism
evabarta@kalasznet.hu

Abstract: The present research uses the verbal report methodology to examine what listening comprehension sub-skills and strategies can be identified in test takers' thought processes during the task-solving procedure. The relevance of the research lies in providing a new aspect to and thus complementing the existing listening comprehension taxonomies, which are based on either theoretical speculation or quantitative research methods. The input for the retrospectees consisted of the listening stimuli (two texts) and two tasks: multiple choice questions and table completion. Fourteen Hungarian retrospectees of level B1-B2 performed both tasks and provided verbal protocols on their thought processes. One part of the protocol data was used to modify and finalize a literature-based preliminary coding scheme by measuring the data against the scheme, which yielded a taxonomy of listening sub-skills and strategies. It was then tested on the rest of the double-coded protocol segments by establishing inter-coder reliability through percent agreement and Cohen's kappa. The analysis of this taxonomy shows that listening for test-taking purposes, which is a non-interactive, stressful, high-stake situation, is characterized by direct reliance on schematic associations to potential referents, a continuous interaction of the aural and written input in the cognitive processes and a strong tendency to elicit forced responses.

Keywords: listening comprehension, verbal report, skills, strategies, testing

1 Introduction

This study investigates listening comprehension test takers' thought processes during the task solving procedure from the pre-listening task preview till the post-listening final decision-making. The aim is to compile and validate a verbal protocol-based taxonomy of listening comprehension sub-skills and strategies, which can be relied on in future research. Furthermore, this research will hopefully be able to contribute to user-friendly, yet valid and reliable testing of listening comprehension and go even further on a practical level by supporting the item writer activity of examination boards.

As the first step of the research, the literature was studied to investigate the general background of one of the main variables of performance in testing listening comprehension: listening comprehension ability. The key concepts of the listening construct, listening sub-skills and strategies were outlined including the various taxonomies of listening comprehension sub-skills and strategies. The review of literature was followed by collecting data via intro- and retrospective interviews with test takers and developing a preliminary coding scheme with listening comprehension sub-skills/strategies as categories. The abundance of verbal protocol data made it possible that only a part of it was used to improve the preliminary coding scheme by scrutinizing all the utterances and code assignments proposed by the two coders. As a result of this, a final coding scheme came into existence which was then piloted on the rest of the verbal protocol data, double coding was conducted

and inter-coder reliability was calculated. This is the procedure through which the answer was sought to the research question guiding this investigation, which is formulated as follows:

What listening comprehension sub-skills/strategies can be identified in the test takers' thought processes during the task-solving procedure?

2 Review of literature

2.1 Psycholinguistic background

2.1.1 First-language comprehension

One of the main fields of psycholinguistics is the process of speech perception and comprehension, in other words, how the listener is able to convert the continuous acoustic signal into discrete linguistic units and, as a further step, into a meaningful interpretation. Several models of listening comprehension have been put forward as simplified representations of the process. Gósy (2005) and Harrington (2001) survey and analyse a wide range of such models and find that they vary in their attempts to describe the entirety or a minor part of the process, in their theoretical concept, functional approach, final goal and in their interaction with other sources of knowledge in real-time interpretation. However, Gósy (2005) claims that experimental results support the language specificity of speech comprehension and raises the issue that the dominance of the English language in psycholinguistic research might question the universal validity of these models.

Speech comprehension is an active process which leads to interpreting the perception of speech phenomena at higher levels. This operation consists of two phases: the perception of acoustic signals denoting linguistic codes on the one hand, and the interpretation of this code system on the other (Lurija, 1979), both phases encompassing a complex series of continuously interacting phases or levels. Harrington (2001) adapted Caron's (1992) model and broke down the second phase into six major steps. At first the sound stream is segmented into a string of linguistic units, in which the individual lexical items are accessed. It is followed by assigning a syntactic structure to the word string and deriving a meaning for the words and syntactic structure as a unit. The last two steps of establishing the real-life referent of the string and recovering the speaker's intention comprise higher order interpretative processes.

Rost (2002) defines listening in a much broader sense in terms of subsequent, overlapping or parallel orientations. According to his model, the process of listening comprehension comprises receptive orientation (receiving what the speaker actually says), constructive orientation (constructing and representing meaning), collaborative orientation (negotiating meaning with the speaker and responding), and transformative orientation (creating meaning through involvement, imagination and empathy). This concept is much broader in the sense that it attempts to model both of what Dunkel (1991a) calls participatory and nonparticipatory listening; the first term is characterized by the listener having "the opportunity to seek clarification or modification of the discourse from the speaker" (Dunkel, 1991b), whereas the latter does not allow for aural-oral interaction during the input.

The hierarchical and interactive psycholinguistic models of speech comprehension exert the greatest influence on the most widely accepted construct in testing listening

comprehension both in their concepts and terminology. The hierarchical models conceptualize the process as various levels built on each other either from bottom to top or from top to bottom between the acoustic input and understanding (Clark & Clark, 1977). The interactive model is one of several other cognitive models, which supposes that the various levels work and interact simultaneously and emphasizes the role of general knowledge, predictions, and contextual expectations (Osgood, 1986; Pléh, 1998).

2.1.2 Second-language comprehension

Both Harrington (2001) and Buck (2001) lament the limited attention speech comprehension issues have received in L2 so far. Harrington finds, however, that research on L2 comprehension has started to appear.

Existing research suggests that unlike the acquisition processes of listening comprehension in native and foreign languages, the cognitive operations of listening comprehension in the case of native and foreign languages are not fundamentally different in any way. Both Dunkel (1991a) and Buck (2001) come to the same conclusion in their reviews of the relevant literature.

Faerch and Kasper (1986) state that second-language listeners differ from first-language listeners only in that they have a restricted knowledge of the language and they might be influenced by transfer from their first language. Furthermore, second-language listeners might lack crucial content or textual schemata (Long, 1989) or important background knowledge that would make it possible for them to compensate for their insufficient linguistic skills (Bremer, Roberts, Vasseur, Simonot, & Broeder, 1996).

Buck (2001) adds, however, that there is a difference in emphasis. According to him problems in native-language listening are often due to distraction, disinterest or thinking about something else, whereas in second-language listening, in addition to these factors, more problems arise due to insufficient knowledge of the language system and lack of knowledge of the socio-cultural content of the message. He also describes compensatory skills as a significant aspect of second-language listening, which are put into force to bridge considerable gaps in the listeners' knowledge of the language.

2.2 The listening construct in testing

2.2.1 Listening comprehension models

Whatever research paper, textbook or handbook on testing we consult there is unanimous agreement that validity is the most important issue in language testing, since if a test is not valid for the purpose for which it is used, then the result does not mean what it is supposed to mean. As we are aiming at measuring listening comprehension, the starting point is answering the question: what is listening comprehension? In the testing literature there has been a move away from the concept of listening as auditory discrimination and decoding of contextualized utterances towards a "much more complex and interactive model which reflects the ability to understand authentic discourse in context" (Brindley, 1998, p. 172). In

spite of the wide variety of terms used in the literature to describe this construct, there seems to be a broad consensus that listening is an active rather than a passive skill and, what is more,

Vandergrift (1999) declares that "listening comprehension is anything but a passive activity" (p. 168). According to Rost (1990), listening involves 'interpretation' rather than 'comprehension' because listeners do much more than just decode the aural message; among others listeners are involved in hypothesis-testing and inferring. Brown (1995) argues in a similar way stating that listening is a process by which listeners construct 'shared mutual beliefs' rather than 'shared mutual knowledge'. Anderson and Lynch (1988) suggest the same notions in terms of metaphors, regarding listeners as 'active model builders' rather than 'tape recorders' (p. 15).

The next step in defining the listening construct is to look into how 'active model builders' interpret, infer, test hypotheses and construct shared beliefs. It is obvious that a number of different types of knowledge are involved, both linguistic knowledge (phonology, lexis, syntax, semantics, discourse structure, etc.) and non-linguistic knowledge (knowledge about the topic, about the context, general knowledge about the world, etc.). The latter categories are frequently referred to as schemata, mental structures that organize the listeners' knowledge of the world that listeners rely on when interpreting texts. Much research has been conducted on the apparent dichotomy between two views as to how these two types of knowledge are applied by listeners or readers in text comprehension (Alderson, 2000). These views refer to the order in which the different types of knowledge are applied during listening comprehension. The bottom-up model represents the traditional view of comprehension and was typically proposed by behaviourists in the 1940s and 1950s. It assumes that the listening process takes place in a definite order, starting with the lowest level of detail (acoustic input, phonemes, etc.) and moving up to the highest (communicative situation, non-linguistic knowledge). According to the top-down model (Goodman, 1969; Smith, 1971), the reader and listener uses the schemata (non-linguistic knowledge) to comprehend a text by interpretation, prediction and hypothesis testing, that is comprehension is seen primarily as the result of applying the schemata the listener brings to the text. Both Alderson (2000) and Buck (2001) rely on a third model of comprehension in their most comprehensive books on assessing reading and listening, respectively. They outline comprehension as the interaction of bottom-up and top-down processing and emphasize that these complex mental actions can be performed in any order, simultaneously or cyclically rather than in any fixed order. This is the interactive (Grabe, 1991) or interactive compensatory (Stanovich, 1980) model. Rost's, Brown's and Anderson and Lynch's listening constructs described above seem to rely more on the top-down model, however, as Alderson (2000) reports, recent research tends to emphasize the important contribution of bottom-up or data-driven text processing.

Buck's (2001) summary of the listening construct constitutes the most widely accepted model in testing listening comprehension:

To summarise the process, the listener takes the incoming data, the acoustic signal, and interprets that, using a wide variety of information and knowledge, for a particular communicative purpose; it is an inferential process, an ongoing process of constructing and modifying an interpretation of what the text is about, based on whatever information seems relevant at the time (p. 29).

2.2.2 Listening sub-skills

There have been numerous attempts to describe listening comprehension in terms of taxonomies of sub-skills. This approach is based on the notion that these skills underlie the process, and the act of listening consists of the application of various separate skills.

One of the first taxonomies is the division of listening into a two-stage process: the extraction of the basic linguistic information as the first step and the utilisation of that information for the communicative purpose. This two-stage division occurs again and again in the literature with various modifications and under various terms: 'the recognition level' and 'the level of selection', 'construction process' and 'utilisation process', 'micro-comprehension' and 'macro-comprehension'. Buck (2001, p. 52) gives high credit to the two-stage view of listening since scholars have worked out similar concepts and the different terminology suggests that they have arrived at these concepts more or less independently.

The cognitive skills taxonomy (Valette, 1977) was meant to be the basis for developing listening tests, which is why it describes a series of increasingly complex cognitive skills so that it can be used to measure various levels of ability. Other attempts were aimed at describing listening skills in communicative terms. These taxonomies go beyond linguistic processing and consider a wide variety of skills necessary for relating the basic linguistic processing to the wider communicative situation. Weir (1993) calls his taxonomy a 'checklist of operations that listening tests should require' and makes it clear that he does not attempt to provide an exhaustive list of listening sub-skills. The most detailed communicative taxonomies were built on various listening purposes and suggested that different lists of 'micro-skills' are required for listening in a social action, listening for information, pleasure, academic purposes or for some other reasons (Richards, 1983). Mention should be made of Munby's (1978) very influential taxonomy of what he calls 'enabling-skills', which includes one of the most detailed lists of sub-skills of both productive and receptive skills.

In addition to the above skills, which are built mainly on theoretical speculation, there are some taxonomy-based studies. Using the newly developed rule-space methodology, Buck, Tatsuoka, Kostin, and Phelps (1997) looked at multiple choice items, and Buck and Tatsuoka (1998) examined short-answer comprehension questions and found 14 and 15 abilities respectively, which they claimed were most important. Buck (2001) warned, however, that the results should be treated with caution as the analysis is based on item characteristics, not abilities and there was no intercoder reliability study.

Besides the assumption that there are identifiable listening skills, there seems to be agreement in the language testing literature that these skills can be arranged in a hierarchy from lower order (e.g., understanding utterances at the literal level) to higher order (e.g. inferencing and critical evaluation) (Buck, 1991; Rost, 1990; Weir, 1993). Some of the numerous graded taxonomies have been applied as a basis for identifying the comprehension operations to be sampled in the listening tests, such as understanding main ideas, listening for specific information, inferring meaning, etc. However, scholars treat the skills-based approach of comprehension with caution; Alderson (2000) states that "the existence of such skills is in some doubt, at least as far as it is possible separately to identify and test them" (p. 93) and Buck (2001) goes even further by stressing that "the empirical support for these taxonomies is usually lacking" (p. 51). Alderson (2000) disapproves that too many lists of skills have been theorised or speculated upon and argues that "the key thing is not how many skills we can

dream up, but how many can be shown to exist on tests" (p. 94). This suggestion of appropriately handling rather than discarding the sub-skills approach is underpinned by referring to valuable studies that investigated how many empirically separable skills there are, whether it can be distinguished which skills the items are testing, which skills contribute most to the performance, which skills are the easiest to test, which skills are the most important to test, etc. Buck (2001) declares that although there is no evidence that these lists of sub-skills constitute a complete unified description of the listening process, there is no doubt that many of the components are of crucial importance in listening. He adds that "collectively they are useful because they tell us what scholars in the field have come to think is important in listening comprehension" (p. 51).

2.2.3 Strategies vs. skills

Strategies have been labelled and classified in various ways and several strategy taxonomies have been proposed (Cohen, 1998a; Dörnyei & Scott, 1997; Oxford, 1990), however, the concept of learning strategy has proved difficult to determine. The definitions of strategies are settled along a scale with one end denoting the perception that strategies are behavioural and as a result of this, observable, whereas other experts state that strategic processes are mental and cannot be observed. Cohen (1998b) represents the former, more accepted view, since he strongly links the consciousness to language use strategies defining them as "mental operations and processes that learners consciously select" (p. 92), yet, when defining test-taking strategies, he allows that respondents are "at least to some degree" conscious of selecting these processes. While Bialystock (1990) denies that strategic processes can be conscious, Ellis (1994) makes an attempt to merge the two views in his concept. Although experts' views on the role of consciousness vary to a great degree, they mostly share the concept that strategies are related to cognitive information processing and are used to define strategic competence as a component of communicative competence (Bachman, 1990; Bachman & Palmer, 1996; Bialystock, 1990). According to Dörnyei (2005), the concept of learning strategies was marginalized by the 1990s since it was found to focus on 'surface manifestations', like tactics and techniques employed by learners. Consequently, the new construct of 'self-regulation' was introduced and the research perspective shifted to examining more dynamic and process-oriented variables.

Test-taking strategies are strategies that respondents apply to tasks in language tests. Cohen (1998a) lists four language use strategies (retrieval, rehearsal, cover and communication strategies) and points out that the same steps or actions constitute test-taking strategies when "they are used to help produce responses to testing tasks" (p. 219).

Since the present study aims at exploring the thought processes of respondents working on listening comprehension tests, the following question seems relevant: what is the difference between a skill and a strategy? Oxford (1990) integrates all strategies described in the literature and draws up the best known qualification of strategies, yet she devotes only a short paragraph to defining skills as the four modalities to be developed in language learning (listening, reading, speaking, writing), which simply mean ability, expertness or proficiency and are gained incrementally during language development.

Ignoring one of the terms (skill or strategy) and applying the other throughout a study or book, which quite frequently happens, might be consistent but the two terms are too close not to be defined. It is clearly pointed out by Grabe (2000), who demonstrates the need for

terminological clarification and recategorization by challenging the experts to define the differences between e.g. 'inferencing skills' and 'strategy to extract and use information' or 'ability to synthesize information'. The examples could certainly be extended even by similarly worded expressions like 'compensatory strategy', which is 'compensatory skill' in another handbook or study. Similarly, Alderson (2000) finds it unclear why 'infer meaning of unknown words from text' – traditionally an important skill in the reading literature – is classified as a 'strategy' in other studies.

The tendency to reclassify well-known variables as strategies, which Alderson (2000) points out, is underpinned by the observation that language teaching is prone to pendulum movements of language teaching theory, which in the case of listening means a shift from a focus on linguistic processing to schematic processing in general (Lynch, 2002). Similarly, Field (1998) argues that the strategic approach has been taken too far and a better balance should be found between skills and strategies. He defines sub-skills as competencies which native listeners have acquired and L2 listeners still need to acquire, whereas strategies as strictly compensatory, which are required less and less as the listener's ability develops. This can be regarded as another aspect of the above mentioned views which consider the element of consciousness as a distinguishing feature of strategies from other processes that are not strategic (Alderson, 2000; Cohen, 1998b).

There have been a number of descriptive models of communicative competence that attempt to describe all the knowledge, skills and strategies that non-native learners need in order to use the language effectively. The most well-known and influential general description of language ability among testers is the communicative language ability (CLA) model developed by Bachman and Palmer (1996). Buck (2001) introduced a similar descriptive model to describe listening ability, which is partly based on Bachman and Palmer's CLA model, partly relies on the works of Purpura (1999) and McNamara (1996).

Buck's model, the 'Framework of listening ability', has two main components: a) language competence and b) strategic competence. Language competence is defined as "the knowledge about the language that the listener brings to the listening situation", which includes "fully automated procedural knowledge and controlled or conscious declarative knowledge" (p. 104). Its sub-categories are based on the main sub-fields within linguistics (grammatical, discourse, pragmatic and sociolinguistic knowledge), which provide either the basis or significant parts of the sub-skill taxonomies mentioned above. Strategic competence in Buck's model is defined as the ability to use language competence. It "includes the cognitive and metacognitive strategies, or executive processes, that fulfil the cognitive management function in listening" (p. 104), and all the compensatory strategies used by second-language listeners. Buck relates the cognitive strategies to comprehending, storing and retrieving input, and the metacognitive strategies to performing "an executive function in the management of cognitive strategies" (p. 104). The 'Framework of listening ability' is of great relevance because it can be considered as a taxonomy where listening comprehension sub-skills and strategies are listed within the same system.

2.3 Verbal report in listening comprehension research

Despite (and after attentively considering) both the criticism of sub-skill theories and the doubts about strategies summarized above, the present study aims at adding yet one more to the existing repertoire of listening sub-skill and strategy taxonomies. This is justified by the

fact that although such lists are not proven and do not claim to provide a complete unified description of the process, it certainly can be accepted that breaking a complex process down to almost grain size and analysing the constituents contribute to understanding the entirety of the process itself, consequently a new method to approach and describe the test takers' behaviour should be beneficial.

The novelty of the present research lies in making an attempt to identify the test takers' listening comprehension sub-skills and strategies based on what the test takers report about their own thought processes. To my knowledge, the taxonomies available in the literature are built on either speculation and hypothesising, or on applying the method of analysing the test item, text, result or the combination of these in empirical research. The studies which apply verbal report methods in researching listening comprehension target a wide range of topics, some of which are related to sub-skills but none of them set the aim of compiling a verbal report-based taxonomy of sub-skills and strategies.

When discussing the research of listening comprehension, Ross (1997) states that this method "has considerable potential as a tool for investigating the psycholinguistic validity of item response patterns and can offer detailed qualitative data to supplement traditional and probabilistic approaches to test analysis" (p. 219). Buck (1991) gives a more focused direction for future research in this field: "Although this method may not be very suitable for testing clearly formulated research hypotheses, it does seem likely to provide a broad view of second-language listening processes and indicate how listening tests work" (p. 68).

3 Description of method

3.1 Participants and material

The participants of the research were 14 Hungarian students – 13 females and 1 male – from intermediate courses that the researcher teaches at the Budapest Business School,. Since these courses lasted for a year and the researcher had a good insight into the students' language performance, the criterion for selecting the 14 students from the volunteers was to make sure that they represent a wide range of B1-B2 level and to avoid that usually the best performers volunteer. None of the students had any prior specific training or spent more than 2 weeks in a native environment, which are factors that could have biased the results.

As the first step in the preparation of the interviews, two tasks representing two different task types (multiple choice and completing a table) were selected from *Are you listening?*(Barta, 2004), a book containing validated intermediate tasks for general listening comprehension tests. Both tasks are built on authentic texts; one of them (Depicted as an ape) is an extract from a BBC radio programme on Darwin's life and age. The second one (Getting heard) is part of an interview with an English mayor conducted by the author of the book about the various petitions citizens submit to him. (The two tasks and the transcripts can be found in Appendices A and B.)

In addition to this, the Interview prompts were compiled according to the research purposes. The pre-listening prompt was meant to be eliciting introspective comments on the task with the aim of exploring the interviewee's thought processes while reviewing the task before the listening, whereas the Retrospective interview prompts were applied after listening

to the sections of the texts. It was complemented by one further prompt which aimed at the cognitive processes during finalizing the answers after listening. (See Interview prompts in Appendix C.)

3.2 Data collection

The data collection took place in a small room with good acoustics at the foreign language department of the college under undisturbed, quiet circumstances in the year 2005. The interviews, each of which took 1–1.5 hours with feedback, were conducted by the researcher who met individually with each informant according to a mutually agreed appointment. Since the participants were native Hungarian speakers, the interviews were conducted in Hungarian in order to guarantee unhindered expression of their ideas.

Before the interview began, the researcher and the interviewee engaged in informal small talk in order to put the interviewee at ease and establish rapport. Then the researcher explained the purpose of the interview very briefly, the procedure of the interview in more detail and the Interview prompts were read and interpreted. The researcher illustrated retrospection on one item of a multiple choice task (other than those involved in the research) and the participant could rehearse on another item or a simple arithmetic task. The procedure of the interview was as follows:

Before listening, the interviewee provided introspective accounts of her thoughts while reviewing and reading the task sheet. During the first listening of the text the interviewee was working on the task sheet while listening to the input. It was followed by the second listening of the text section by section: the researcher played a cohesive section of the text, with a pause after the section, which covered 2-4 task items each. During listening, the interviewee was working on the task sheet. When the researcher paused the text of the task, the interviewee verbalized their thoughts retrospectively.

In order to replicate every possible detail of the test-taking situation and to gain more insight into the decision making process, the interviewees were encouraged to verbalize their thoughts introspectively while making the final decisions on the task sheet after listening to the section. When the interview was completed, the researcher thanked the interviewee and gave feedback both on the result of the test and on the strengths and weaknesses of the interviewee as a listening test taker.

It is to be noted that the data collection was planned and implemented in a free interview format and interview prompts were used rather as guidelines for the interviewee. Since experience showed that the quality of the verbal reports depends on the probing procedure employed by the interviewer, purposeful but unobtrusive probing was applied throughout the interviews for the purpose of accessing and eliciting essential listening comprehension test-taking processes.

3.3 Data analysis

The abundance of data in this study made it possible to improve and finalize the coding scheme on one part of the protocol data and then testing it on the rest. This is why after transcribing all the retrospective interviews of the 14 students (verbalizing on two tasks

each) only the data provided by 10 students were used on formulating the coding scheme, the other part of the data collected from 4 students was applied in the pilot of the coding scheme.

3.3.1 Developing a preliminary coding scheme

Developing a coding scheme is critical in abstracting useful and valid information from verbal protocols as it shapes and constraints the inferences that may be drawn from the data. Since it is also the main focus of the present study, special emphasis was put on it in the research.

It was taken into consideration on the one hand, that – as the survey of relevant literature shows – there is confusion in the fields of sub-skills/strategies due mainly to the difficulty of making distinctions. On the other hand it is also relevant that the study aims to analyse any mental acts that can be identified in the thought processes reported by the test takers. As a consequence, in order to develop a coding scheme with listening comprehension mental acts as categories several L2 comprehension inventories of both sub-skills and strategies were examined, among others: Munby (1978), Richards (1983), Oxford (1990), Weir (1993), Bachman and Palmer (1996), McNamara (1996), Buck et al. (1997), Buck and Tatsuoka (1998), Cohen (1998b), Purpura (1999), Nikolov (2006). These inventories were relied on as a starting point to develop a draft preliminary taxonomy, which consisted of 56 items, including those that were added by the researcher based on her previous research, experience and intuition. After attentive consideration these were reduced to a taxonomy of 34 items by combining the categories covering the same or very similar processes and by eliminating categories irrelevant due to the aimed level or listening purpose. This 34-item coding scheme was discussed and slightly modified with an expert of testing and reading comprehension and used as the preliminary taxonomy of listening comprehension sub-skills in the first phase of the research. In this taxonomy the number of categories is rather high, yet many of the categories themselves are vague and verbose so that they can trigger discussion as to both the content and wording of the final categories.

3.3.2 Finalizing the preliminary coding scheme

The purpose of this phase, as mentioned above, was to finalize the preliminary coding scheme. It is done by the method of measuring the scheme, which was compiled mainly from data gained from hypothesising and quantitative research on the content analysis of the input, against data from a qualitative source, namely verbal protocols.

The first part of the verbal protocols was segmented by the researcher and the segmentation was then reviewed by the expert who suggested extending, merging, deleting or splitting segments or identifying new ones. After negotiating the final segmentation in this way, which can be regarded as a type of double-segmenting, the 459 segments of the verbal protocols were thus double-coded. It was also agreed that in case of uncertainty about the content of the segment in question, this can be noted by using a coding category labelled ‘miscellaneous’. Since – as mentioned above – the purpose of this phase was to finalize the coding scheme by disambiguating both the protocol and the coding scheme, the co-coders agreed that utterances can be assigned to two categories. It was expected that this method combined with expert discussion enhances refining, interpreting and setting boundaries to the categories, elicits new categories and makes it easier to discover recurrent patterns.

The double coding was followed by re-scrutinizing those segments which produced disagreement between the coders or turned out to be ambiguous for either of them. As at this step it was not intended to finalize the coding scheme, problem statements were re-coded only if the code was found to have been assigned due to lapse of attention or when the expert miscoded a segment due to lack of non-verbal information (prosodic or paralinguistic elements, observations, etc.) during the interview process.

The majority of the 23 segments coded as miscellaneous were found to be fake segments and were not entered into the further analysis, whereas 8 segments were found to capture very similar information and were included as a new category in the final coding scheme.

Then the numbers of co-occurrences of each and every category with every other category on the 436 segments which were entered into the further analysis were arranged in a table format which clearly presented the recurrent patterns. In this way by simple eyeballing it could be decided (relying also on the experience gathered during the expert discussions) which categories are sufficiently similar to merit their fusion to form one category since they proved to be capturing the same content. Furthermore, three categories were completely deleted as they were not sufficiently distinguishable to elicit coding agreement. Finally, the wording of the categories of the final coding scheme was disambiguated and finalized.

All these changes and adjustments resulted in the 27-item final coding scheme called 'Taxonomy of listening comprehension sub-skills/strategies', which is presented and commented on in the Results and discussion section below.

3.3.3 Piloting the final coding scheme

The purpose of this phase is to pilot the finalized coding scheme in order to probe and hopefully prove the appropriacy of its reliability level.

The second part of the verbal protocols (four students verbalizing on the two tasks) was segmented through the same process as described above. Unlike in the first phase though, since the purpose of this phase was to establish inter-coder reliability rather than improving the coding scheme, the double coding was conducted in the traditional method: both coders assigned one code unambiguously to each of the 188 identified segments of the protocols.

Inter-coder reliability was calculated in two ways. Percent agreement, which reflects the number of times that the raters agreed on the exact categorization, seems to be used most widely and is both intuitively appealing and simple to calculate. The percent of total agreement across the two raters on the 188 segments representing identified mental acts was 78.72 %.

However, the methodological literature is consistent in identifying percent agreement as an inappropriately liberal and consequently misleading measure that overestimates true inter-coder agreement, and if it is reported the researcher is expected to add an index that tests whether agreement exceeds chance levels and if so to what extent. Cohen's kappa coefficient does so and is generally thought to be a more robust measure than simple percent agreement calculation. That is why this calculation was done on the same data and resulted in 0.771. According to expectations, the liberal measure yields a higher index than the much more

trustworthy conservative one. The difference is slight though, and both figures show quite a high reliability level.

4 Results and discussion

4.1 Grouping skills/strategies

The answer to the research question of the present study (What listening comprehension sub-skills/strategies can be identified in the test takers' thought processes during the task-solving procedure?) is the 27-item taxonomy presented below, which was compiled and validated in the process described above. However, the validity of the taxonomy is planned to be enhanced by measuring it against the data gained from verbal protocols on 3 various text types, each different from the 2 involved and processed in the study so far. This further step in the never ending validation process will show to what extent different texts yield the categories of sub-skills and strategies that are set up in the taxonomy.

The 27 categories of mental acts (skills and strategies) identified in the protocols were further analysed in order to reveal possible patterns which would make it possible to group them. Firstly, based on the researcher's experience gained through segmenting, coding, discussing and analysing the segments of the protocols they were investigated as for what kind of grouping system they lend themselves to. Then possible grouping systems were put forward. Finally, the grouping systems were compared with the taxonomies which had been consulted when compiling the preliminary coding scheme. Through this process the grouping system was set up and the 27 categories of mental acts were grouped by assigning them to one of the sub-groups within the agreed grouping system by the researcher and the expert independently. Finally, the grouping was concluded through expert discussion.

In summary, the 27 categories of mental acts identified in the verbal protocols are divided in two main groups: language competence, which consists of 7 listening sub-skills and strategic competence comprising 20 listening strategies. The sub-skills are further broken down into sub-groups of knowledge (grammatical, discourse, sociolinguistic) depending on the field of linguistics the sub-skills are aiming at. Within strategic competence, 3 cognitive, 8 compensatory and 9 metacognitive strategies were separated. Some of them could be further categorized, e.g. the application of checking/skipping example was reported both before and while listening, whereas self-correction, self-approval and self-evaluating both while and after listening. Similarly, excluding options proved to be an umbrella category denoting the method of working out the answer but during the process of considering/probing and rejecting/accepting the three options in turn, several other sub-skills and strategies were applied.

The skills and strategies identified in the protocols were divided into groups as listed below in an inventory of listening sub-skills and strategies. The numbers in brackets after the taxonomy items give their frequencies with which they occurred within the 188 segments of the second part of the verbal protocols. Furthermore, each item has been provided with an example to illustrate both the content of the given category and the character of the mental acts that the participants gave account of. The translated excerpts from the protocols are in italics, words that the interviewees say in English are in capital letters, the short clarifications added by the researcher are in brackets. For ease of reading and interpretation the excerpts

have been slightly edited, merely leaving out redundancies, without altering the essential meaning.

INVENTORY OF LISTENING SUB-SKILLS AND STRATEGIES

LISTENING SUB-SKILLS (Language competence)

GRAMMATICAL KNOWLEDGE

- **Understanding phonological features** (4)

  Example 1. *I misheard it at first. I wrote CROWDED so the libraries are overcrowded and at second listening I heard quite the opposite that they might be closed. Perhaps because both words begin with letter c.* (Participant 4 on Task 2 Item 5)

- **Understanding lexis** (8)

  Example 2. *I really don't understand too much here, well, the words and it is obviously a problem, too.* (Participant 13 on Task 1)

- **Understanding syntactic structures** (2)

  Example 3. *He said it in negation several times, that no, he didn't do it, I could exclude that he refers to it at all*. (Participant 4 on Task 1 Item 6)

- **Recognising cues of oral punctuation** (4)

  Example 4. *I wrote PARK AREA because of the pause before it again. It wasn't such a big pause but it was emphasized pretty well here, too.* (Participant 4 on Task 2 Item 7)

DISCOURSE KNOWLEDGE

- **Understanding gist/main ideas** (11)

  Example 5. *At the library I could understand by and large that it is closed. I don't know when and how. All the time or what is the problem.* (Participant 13 on Task 2 Item 5)

- **Understanding specifics/important details** (3)

  Example 6. *I couldn't fill in this part of the table because I couldn't grab exactly those words in spite of globally understanding this part.* (Participant 2 on Task 2)

SOCIOLINGUISTIC KNOWLEDGE

- **Making inferences by social/situational contexts, background/real-world knowledge** (13)

  Example 7. *She spoke about both A and C. I chose C but if I think again that he started to publish them at the age of 50 I might find it too much… although Darwin… in that period he might as well have published at the age of 50.* (Participant 3 on Task1 Item 1)

LISTENING STRATEGIES (Strategic competence)

COGNITIVE STRATEGIES

- **Keeping pace with the text** (12)

  Example 8. *I could not pay attention because I was dealing with the only word NEGLECTED, because I don't understand it so I got stuck there and couldn't pay attention to the rest.* (Participant 12 on Task 2 Item 8)

- **Retaining chunks of language for short periods** (5)

  Example 9. *There was something with the opening hours. But I can't remember if it was OPEN or CLOSE. I tend to confuse such things, like left and right.* (Participant 1 on Task 2 Item 5)

- **Constructing a response quickly and efficiently** (3)

  Example 10. *When you have to write in numbers, when he says long numbers in English by the time I put it together in Hungarian, and then write it down, I have the biggest problem with it.* (Participant 7 on Task 2 Item 11)

COMPENSATORY STRATEGIES

- **Recognizing synonyms/paraphrases** (6)

  Example 11. *It is said that he doesn't speak, that HE DOESN'T SPEAK ABOUT IT SIMPLY than it means that he doesn't mention anything in his study, that he doesn't mention anything in connection with religion.* (Participant 7 on Task 2 Item 6)

- **Deducing meaning of unfamiliar lexical items from context** (3)

  Example 12. *What kind of document he gives to the man who handed in a petition…I think…such a …RECEIPT… but I'm not sure. Well, RECEIPT means income, too. And I hope it means some kind of acknowledgement, too, because the other thing he said was that he gives a promise to answer and a promise is certainly not a document… so I think it must be the word*. (Participant 4 on Task 2 Item 14)

- **Making text-based inferences** (9)

  Example 13. *I circled C, she didn't say exactly that he* (Darwin) *decorated his house with the cartoons but she said that he hung them on the walls*. (Participant 4 on Task 1 Item 4)

- **Identifying what type of information to search for** (10)

  Example 14. *I have a feeling that it is the parking that is problematic, may be not. He emphasizes that it is problematic… so I look at it* (the table) *what the problem is, and it is in connection with the traffic so I think it may be the parking problem.* (Participant 1 on Task 2 Item 3)

- **Using previous or subsequent items to help information location** (3)

  Example 15. *At item 12 the EXCEPT FOR PUBLIC HOLIDAYS* (text on the worksheet) *helped. Then you realize that bingo something has been said that should have been written down.* (Participant 4 on Task 2 Item 12)

- **Constructing a response from words heard before or after key words of task item** (8)

  Example 16. *I concluded that it is the answer because it was close to VANDALISM, that is SCHOOL was close to VANDALISM* (a word in the written input). (Participant 13 on Task 2 Item 6)

- **Excluding options** (14)

  Example 17. *I chose A. I excluded B because she says he* (Darwin) *didn't like to be in the centre at all and now I managed to exclude C at second listening because if I could understand it well she said that his friends did the public side of his work for him. Although she doesn't say that he liked RESERVED LIFE but she said he liked to be with his family and gardening. During the first listening I thought of C as well, because I could catch PUBLIC SIDE OF HIS WORK but during the second listening I heard that his friends do it for him.* (Participant 4 on Task 1 Item 5)

- **Strategies of last resort (Blind guessing or Blind construction)** (11)

  Example 18. *Many times when I have no idea which one to circle I circle the longest one* (option). *And it works in 90%. It is really so.* (Participant 1 on Task 1)

METACOGNITIVE STRATEGIES

- **Checking rubrics** (4)

  Example 19. *The wording of the task frightened me a little bit because I found words in it which I don't know and I felt they are important.* (Participant 2 on Task 2)

- **Skipping rubrics** (8)

  Example 20. *I skipped the instruction. I sometimes do this that I skip the instruction of the tasks, which is not very advantageous.* (Participant 1 on Task 2)

- **Reviewing task before listening** (10)

  Example 21. *The table is very well constructed and since it contains lots of data you can understand the task even without the rubric.* (Participant 4 on Task 2)

- **Predicting answer before listening** (4)

  Example 22. *I guess it will be A. I'm not sure but researchers like reserved life, they are unsociable, like the people of arts, special creatures, at least for me.* (Participant 1 on Task 1)

- **Checking example before/while listening** (3)

  Example 23. *I always try to pay attention also to question zero, to when it is over. During the first listening I wouldn't have known it, because I expected that one of them is mentioned, but both PARADOX and MAKING were mentioned so I didn't understand why. But during the second listening it was clear why it is C.* (Participant 7 on Task 1)

- **Skipping example before/while listening** (7)
  Example 24. *I don't like examples, mostly I don't even pay attention to them. It disturbs me that an example is given and I can never clearly make it out and then I get nervous that goodness me I can't hear item zero which is surely in the text.* (Participant 1 on Task 2)
- **Self-correction while/after listening** (9)
  Example 25. *First I wrote this solution into the wrong box, because I didn't read it properly. This is my fault. So I am crossing it now.* (Participant 3 on Task 2)
- **Self-approval while/after listening** (8)
  Example 26. *It seemed A during the first listening because he was not a sociable man. At first I wasn't sure but it cleared up at the second listening.* (Participant 6 on Task 1 Item 5)
- **Monitoring and self-evaluating while/after listening** (6)
  Example 27. *I am thinking a lot now. I feel now again, as during the whole listening that I am not sure about myself. That is not about the test but myself.* (Participant 2 on Task 2)

It must be noted that the separation of language competence and strategic competence is problematic since it is difficult to distinguish knowledge of the language from the ability to apply that knowledge (McNamara, 1996). Similarly, it is hard to sub-categorize both sub-skills and strategies because they intertwine in complex mental processes and rarely manifest themselves in unambiguously definable pure forms. However, as described above and in the method section, validity in the study has been enhanced by meticulous processes of double-segmenting, double-coding, probing and piloting the system of categorizing and grouping the reported mental acts identified in the verbal protocols.

4.2 Framework of listening ability

Measuring the outcome of the present research against Buck's Framework for describing listening ability (2001) was encouraged by his statement, in which he put forward his model as work in progress. He added that the model was not intended to be exhaustive and was meant to provide a basis for discussion.

As far as the descriptive model of listening ability set up by him is concerned, the comparison of the theory-based Framework and the categories/groups of categories identified in the verbal protocol shows that there is considerable consistency between them. Firstly, the items of the list set up based on the verbal protocols fell into two categories, according to denoting either strategic competence (listening strategies) or language competence, similarly to the framework. Language competence comprises various kinds of 'knowledge' in Buck's framework and these sub-headings were retained in the present taxonomy, according to which the sub-skills were subdivided. Buck does not call these fields of knowledge 'sub-skills' in the Framework itself, however, he introduces his model as an attempt to describe "all the knowledge and skills" necessary for effective language use (p. 102) and states that performing on listening tests needs a combination of knowledge, skills and strategies (p. 256). The application of the term 'sub-skill' in the present taxonomy is also justified by the fact that it seems to be the most widely accepted term in the literature to name what the listener's knowledge about the language can be broken down into.

Secondly, test takers in the present study gave account of thought processes fitting all but one sub-category of the Framework's two competences. It suggests that the findings of the present study concur with the theory-based framework. Nevertheless, let us examine what might be behind the only main difference between them. The only sub-category of the framework which was not found to be referred to in any of the segments of the database is one

of the language competencies: pragmatic knowledge. Some reasons come into view as possible explanations for the lack of emergence of this important sub-skill within the context of the present study. Since the research aims at investigating listening comprehension at levels B1- B2, both the types and items of tasks on the one hand and the difficulty level and topic of the texts on the other hand were selected and developed according to this purpose. At this level, tests mostly intend to measure the understanding of stated information rather than unstated implications, such as pragmatics. Nevertheless, it is only partly due to the level. It is also due to the practical reason that it is difficult to measure the comprehension or identification of the function or any illocutionary force of an utterance. An item aiming at this sub-skill by asking 'what is meant by the text?' mostly disrupts the logical sequence of the other task items, which ask 'what does the text mean?', and as a consequence, might confuse the test taker.

Although systematic item analysis of the two tasks has not been carried out as for which aspect of the listening ability each item is aiming at, the survey of the items suggests that none of them aims or has the potential to trigger pragmatic knowledge. That might be the reason why no reference could be traced to pragmatics in the data of the protocols.

When determining the components of Strategic competence, a group of strategies, namely compensatory strategies emerged as exceedingly dominant in the verbal protocols. That is why, unlike in the Framework where they are mentioned but not separated within strategic competence, in this study compensation strategies were judged to deserve a separate subgroup. There seems to be a simple explanation behind this disparity in emphasis. The Framework is proposed to testers as a list of components that testers can choose from to include in their tests and several of the compensatory strategies are undesirable in valid and reliable testing, whereas these strategies do exist and the test taker resorts to them quite frequently as their dominance in the verbal reports demonstrate.

4.3 Listening purpose

The analysis of the items of this inventory supports the view that different lists of sub-skills and strategies are required for various listening purposes (Richards, 1983). Clearly, the situation in which the listening takes place can have considerable effect on various aspects of the listening processes, including the sub-skills to be activated and the strategies to be applied. The participants were listening for test-taking purposes, which is a non-interactive, stressful, high-stakes situation. Based on the data of the protocols, it led to a high number and frequency of applying compensatory strategies, as it is discussed in the previous section. Other effects of the listening purpose that can be observed in the reported sub-skills and strategies are over-exploitation of schemata, emphasized reliance on the written input and various types of forced responses. Let us look into these effects in more detail below.

According to Rost (1990) "non-interactive comprehension tasks require a more direct reliance on schematic associations to potential referents" (p. 102). Participants frequently made inferences of various types by relating utterances to social or situational contexts or by using background or real-world knowledge, as in example 28.

> *Example 28 Relying on schemata – Making inferences (Participant 11 on Task 1 Item 3)*
> *I excluded C because I know Darwin's activity to a certain extent and I know that his theory was accepted in his time. So it is not probable that he was closed out because of it.*

Interestingly, participant 12's schemata about the acceptance of Darwin's ideas greatly differ from the above and leads to the opposite inference.

Example 29 Relying on schemata – Making inferences (Participant 12 on Task 1 Item 7, referring to Darwin's wife)
C seems realistic because she had good reason to worry about her husband after all because he came up with very big innovations in that time.

Similarly, participants relied on schemata to a certain extent when identified the response by determining what type of information to search for, as in example 30.

Example 30 Relying on schemata – Identifying what type of information to search for (Participant 13 on Task 2 Item 8)
I am not sure which one is right, maybe something in connection with cleanliness. In connection with keeping something clean, the park. I thought that perhaps it fits in here, it is acceptable here.

Listening comprehension test takers are expected to follow a task while listening and the very strong interaction of the aural and written input in the cognitive processes was also quite frequently verbalized, often inseparably. In reading comprehension tests, by comparison, testees can work around the reading text itself, whereas the real-time nature of spoken language makes it impossible and the testees' attention is or should be shared between the aural and written inputs continuously while listening, with more focus shifted onto the written input when finalizing the answers. Let us see some examples within compensatory strategies how the two inputs interact in the thought processes.

In example 31, the participant considers each option of the written input, compares them with extracts from the aural input alternately and goes back to an option to confirm why he excludes this before he makes the final decision about the correct answer.

Example 31 Interaction of inputs – Excluding options (Participant 11 on Task 1 Item 7)
His wife was a Christian woman and she wrote her thoughts to him in a letter that if he thinks in that way he will not get into heaven, but it is said that "she doesn't want to change him", that is "doesn't want to push his views" or something like this. After all "totally shared" can be excluded because then she wouldn't have written the letter that he will not get into heaven. So C is left, which is a kind of worrying.

It follows from the nature of the retrospective interview that it is impossible to define if this complex segment of cognitive process is the delayed but exact account of while-listening-thoughts or it is completed/modified after listening when the participant could take his time to make or justify his choice. Either way, it gives good insight into the reading and working load the testee has to cope with when working on one listening comprehension multiple choice item.

Participants quite frequently reported constructing a response from words heard before or after key words of the task item on the written input, constituting a compensation strategy based on informed guessing that item writers must be aware of as a potential threat to validity. This kind of reliance on the written input might assist test takers to pick the right answer without real comprehension of the extract as the participant herself states it in example 32.

Example 32 Interaction of inputs – Constructing a response from words heard before or after key words of the task item (Participant 13 on Task 2 Item 2)
At item 2 there is something in connection with houses. I wrote HOUSING BENEFIT but to tell the truth I can't understand the second part of it. I wrote it because I heard it at DISTRESS. I don't know if it's good or not. It was close to it, to the word itself. (Participant points at word 'distress' on the task sheet).

Extreme or exclusive reliance on the written input, however, proved to be exerting a detrimental effect on performance in each reported case. Example 33 illustrates how it leads to a breakdown in listening comprehension on the one hand and further confusion in processing the rest of the text on the other.

Example 33 Interaction of inputs (Participant 8 on Task 2)
I got stuck here because I couldn't hear any of the words that can be seen here and I didn't notice that we have already passed beyond this.

In this example the participant recounts that concentration on the written input diverts her attention from attempting to comprehend the listening text since she focuses on recognizing the spoken form of the written input. She lags behind in following the aural input as a consequence of her false expectation, which probably results from inexperience, inappropriate practice or lack of test-taking technique.

The high-stakes situation of test taking tends to elicit forced responses from the test takers who reported a wide repertoire of what can be called 'strategies of last resort'. Although the idiosyncrasy of processing the listening text is beyond the focus of the present study, the data show that the participants' portfolios of sub-skills and strategies that they bring to the situation greatly vary. However, nearly each participant reported the application of various strategies they resort to when they are forced to produce an answer under the pressure of the test-taking situation. Some of them gave account of real blind guessing as in example 33.

Example 33 Strategies of last resort (Participant 6 on Task 1 Item 4)
I would decide in favour of C because I had to choose something and I just chose this. I looked at the ceiling and chose this.

In many cases the declaration of applying blind guessing is complemented with some kind of justification in which participants refer to their test-wiseness stating that the longest option is the correct answer (see Example 18) or that they select the least frequently occurring option out of A, B and C. In other cases participants refer to mysterious information sources like the belief that the first impression always works for them or to intuition, which is demonstrated by example 34.

Example 34 Strategies of last resort (Participant 5 on Task 1 Item 6)
I just guess here. I don't know, I didn't understand it. I just chose C. Intuition suggested it but I can't support it with anything because I couldn't make out which is appropriate, which fits here.

The phenomenon of uninformed or informed guessing was not, however, confined exclusively to selecting a multiple choice item. Participants reported making efforts to coerce themselves to produce an answer through various ways of blind construction as a last resort when they failed to understand the spoken text.

Example 35 Strategies of last resort (Participant 14 Task 2 Item 3)
It is sure that I wouldn't leave it blank at the exam. I would write something. Well, perhaps I would write to item 3 that OUTSIDE…Although it is already in the question, so… But I would try OUTSIDE.

In example 35 the participant makes a desperate attempt to find something she could fill in the gap. She retrieves the word 'outside' from her memory, but hesitates when she realizes that this word is already given in the written input. In spite of this she writes the word in the gap, which should in no way be left blank, according to her.

5 Conclusion

The main aim of the present research was to investigate what sub-skills and strategies of listening comprehension can be identified in the test takers' thought processes which are not normally accessible through quantitative research methods. Results support the expectations that the methodology of intro- and retrospective interviews can provide valuable insights into many aspects of comprehension and language processing. Since "listening is hard work and deserves more analysis and support" (Vandergrift, 1999, p. 168) and "the assessment of listening abilities is one of the least understood, least developed and yet one of the most important areas of language testing" (Alderson & Bachman, 2001, as cited in Alderson & Banerjee, 2002, p. 89), the contribution from an underexploited methodology in this field is beneficial.

Through a meticulous process of double-segmenting, double-coding and piloting the interviewee's utterances, a 27-item taxonomy of listening sub-skills and strategies were compiled, which was subdivided into subgroups. The findings were found to concur with the theoretical framework of listening ability and also with the concept that the listening purpose exerts an impact on which sub-skills and strategies are mobilized by the listener. The non-interactive, stressful, high-stakes situation of listening during test-taking tends to activate the schemata considerably, highlighting for item-writers the importance of selecting input texts that lend themselves to tasks with high passage dependency. Also, participants gave account of applying a wide repertoire of compensation strategies, some of which should and can be avoided by writing valid and reliable items.

The participants reported a strong interaction of the aural and written input in the cognitive processes, which, in the case of comprehension breakdown led to overreliance on the written input. What is more, exclusive concentration on the written input, resulting either from expecting to hear the same word strings or from excessive reading load, diverted the listeners' attention from the listening text and could even cause loss of control over the test-taking process. During listening comprehension tests, unlike during measuring any other skill, test takers are not in control of their own speed of working since time allotment is determined by the sequence of items and the character of the aural input. This fact again points to the responsibility of the listening comprehension test designer and item writer who hopefully can make use of the lessons learnt in this study in creating valid and reliable tasks and items which are also fair to the test taker.

In sum, according to its aim, this study led to compiling a verbal report-based taxonomy of mental acts during the listening comprehension test-taking procedure, which was then compared and found to comply with the theoretical concepts of the listening ability. However, the analysis of the results highlighted many ways and provided numerous examples of how the construct validity of listening comprehension tests might be threatened by what

Messick (1996) calls 'construct-underrepresentation' (inadequate or incomplete sampling of the construct) and 'construct-irrelevant variance' (measuring things that are not relevant to our construct definition). It calls for further exploitation of this study's database with a separate analysis focusing on investigation into the above threats.

Acknowledgement

I would like to express my appreciation to Ágnes Loch for all her contribution as a co-coder and co-thinker in this study.

Proofread for the use of English by: Nick Chandler, Budapest Business School, College of Finance and Accountancy.

References

Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Alderson, J. C., & Banerjee, J. (2002). Language testing and assessment (Part 2). *Language Teaching*, *35*(2), 79-113. https://doi.org/10.1017/S0261444802001751

Anderson, A., & Lynch, T. (1988). *Listening*. Oxford: Oxford University Press.

Bachman L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Barta, É. (2004). *Are you listening?* Budapest: Akadémiai Kiadó.

Bialystock, E. (1990). *Communication strategies in second language acquisition.* Oxford: Blackwell.

Bremer, K., Roberts, C., Vasseur, M., Simonot, M., & Broeder, P. (1996). *Achieving understanding: Discourse in international encounters*. London: Longman.

Brindley, G. (1998). Assessing listening abilities. *Annual Review of Applied Linguistics, 18*, 171-191. https://doi.org/10.1017/s0267190500003536

Brown, G. (1995). *Speakers, listeners and communication*. Cambridge: Cambridge University Press.

Buck, G. (1991). The testing of listening comprehension: An introspective study. *Language Testing*, *8*(1), 67-91. https://doi.org/10.1177/026553229100800105

Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing*, *15*(2), 119-157. https://doi.org/10.1177/026553229801500201

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

Buck, G., Tatsuoka, K., Kostin, I., & Phelps, M. (1997). The sub-skills of listening: Rule-space analysis of a multiple-choice test of second language listening comprehension. In A. Huhta, V. Kohonen, L. Kurki-Suonio, & S. Luoma (Eds.), *Current developments and alternatives in language* assessment (pp. 589-624). Tampere: University of Jyvaskyla.

Caron, J. (1992). *Introduction to psycholinguistics*. Toronto: University of Toronto Press.

Chalhoub-Delville, M. (2001). Task-based assessments: Characteristics and validity evidence. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks, second language learning, teaching and testing* (pp. 210-228). Harlow: Longman.

Clark, H., & Clark, E. (1977). *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt.
Cohen, A. (1998a). *Strategies in learning and using a second language*. London: Longman.
Cohen, A. (1998b). Strategies and processes in test/taking and SLA. In L. F. Bachman, & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 90-111). Cambridge: Cambridge University Press.
Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah: Lawrence Erlbaum.
Dörnyei, Z., & Scott, M. L. (1997). Communication strategies in a second language: Definitions and taxonomies. *Language Learning, 47*(1), 173-210. https://doi.org/10.1111/0023-8333.51997005
Dunkel, P. (1991a). Listening in the native and second/foreign language: Toward an integration of research and practice. *TESOL Quarterly, 25*(3), 431-457. https://doi.org/10.2307/3586979
Dunkel, P. (1991b). Computerized testing of nonparticipatory L2 listening comprehension proficiency: An ESL prototype development effort. *Modern Language Journal, 75*(1), 64-73. https://doi.org/10.1111/j.1540-4781.1991.tb01084.x
Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
Faerch, C., & Kasper, G. (1986). The role of comprehension in second language learning. *Applied Linguistics, 7*(3), 257-274. https://doi.org/10.1093/applin/7.3.257
Field, J. (1998). Skills and strategies: Towards a new methodology for listening. *ELT Journal, 52*(2), 110-118. https://doi.org/10.1093/elt/52.2.110
Goodman, K. S. (1969). Analysis of oral reading miscues: Applied psycholinguistics. *Reading Research Quarterly 5*(1), 9-30. https://doi.org/10.2307/747158
Gósy, M. (2005). *Pszicholingvisztika*. Budapest: Osiris Kiadó.
Grabe, W. (1991). Current developments in second-language reading research. *TESOL Quarterly 25*(3), 375-406. https://doi.org/10.2307/3586977
Grabe, W. (2000). Developments in reading research and their implications for computer-adaptive reading assessment. In M. Chalhoub-Deville (Ed.), *Issues in computer-adaptive tests of reading* (pp. 11-47). Cambridge: Cambridge University Press.
Harrington, M. (2001). Sentence processing. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 91-124). Cambridge: Cambridge University Press.
Long, D. R. (1989). Second language listening comprehension: A schema-theoretic perspective. *Modern Language Journal, 73*(1), 32-40. https://doi.org/10.1111/j.1540-4781.1989.tb05307.x
Lurija, A. R. (1979). *Jazik i szoznanyije*. Moscow: Nauka.
Lynch, T. (2002). Listening: Questions of level. In R. B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (pp. 39-45). New York: Oxford University Press.
McNamara, T. F. (1996). *Measuring second language performance*. London: Longman.
Messick, S. (1996). Validity and washback in language testing. *Language Testing 13*(3), 241-256. https://doi.org/10.1177/026553229601300302
Munby, J. (1978). *Communicative syllabus design*. Cambridge: Cambridge University Press.
Nikolov, M. (2006). Test-taking strategies of 12-13-year-old Hungarian learners of EFL: Why whales have migraine. *Language Learning, 56*(1), 1-51. https://doi.org/10.1111/j.0023-8333.2006.00341.x
Osgood, C. E. (1986). Mondatok megértése és létrehozása. In Cs. Pléh (Ed.), *Szöveggyűjtemény a pszicholingvisztika tanulmányozásához* (pp. 136-168). Budapest: Tankönyvkiadó.
Oxford, R. (1990). *Language learning strategies: What every teacher should know*. New York: Newbury House/Harper & Row.

Pléh, Cs. (1998). *A mondatmegértés a magyar nyelvben*. Budapest: Osiris Kiadó.
Purpura, J. (1999). *Learner strategy use and performance on language tests: A SEM approach*. Cambridge: Cambridge University Press.
Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, *17*(2), 219-240. https://doi.org/10.2307/3586651
Ross, S. (1997). An introspective analysis of listener inferencing on a second language listening test. In G. Kasper, & E. Kellerman (Eds.), *Communication strategies* (pp. 216-237). Harlow: Longman.
Rost, M. (1990). *Listening in language learning*. London: Longman.
Rost, M. (2002). *Teaching and researching listening*. London: Longman.
Smith, F. (1971). *Understanding reading.* New York, NY: Holt, Rinehart and Winston.
Stanovich, K. E. (1980). Towards an interactive compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly 16*(1), 32-71. https://doi.org/10.2307/747348
Valette, R. M. (1977). *Modern language testing* (2nd ed.). New York: Harcourt Brace Jovanovich.
Vandergrift, L. (1999). Facilitating second language listening comprehension: Acquiring successful strategies. *ELT Journal, 53*(3), 168-176. https://doi.org/10.1093/elt/53.3.168
Weir C. (1993). *Understanding and developing language tests*. New York: Prentice Hall.

APPENDIX A

Task 1

Depicted as an ape

You will hear a radio programme on Charles Darwin and how 19th century Victorian Britain reacted to his revolutionary ideas. While listening, circle a), b) or c) according to the text.

0. The title of the programme on Darwin is
   a) 'Paradoxes'
   b) 'Riddles of the Past'
   c) Making History'
1. He made his theory public
   a) right after his 5-year expedition.
   b) 12-13 years after returning to London.
   c) at the age of 50.
2. The title of Janet Browne's biography of Darwin is
   a) 'Charles Darwin: The Power of Place'.
   b) 'Darwin: The Later Part of his Life'.
   c) 'In the Public Arena'.
3. The local community
   a) didn't trust him any more.
   b) criticized but accepted him.
   c) closed him out.
4. His reaction to cartoons depicting him as an ape or monkey:
   a) Attacked them in magazine Punch.
   b) Gave them as present to friends.
   c) Decorated his house with them.
5. He enjoyed
   a) a reserved life.
   b) being in the centre of public life.
   c) the public side of his work, too.
6. In his work 'On the Origin of Species' he ….. aspects of faith.
   a) criticizes some
   b) refers to some
   c) doesn't mention any
7. His wife, Emma ….. his views.
   a) totally shared
   b) wanted to change
   c) was worried because of

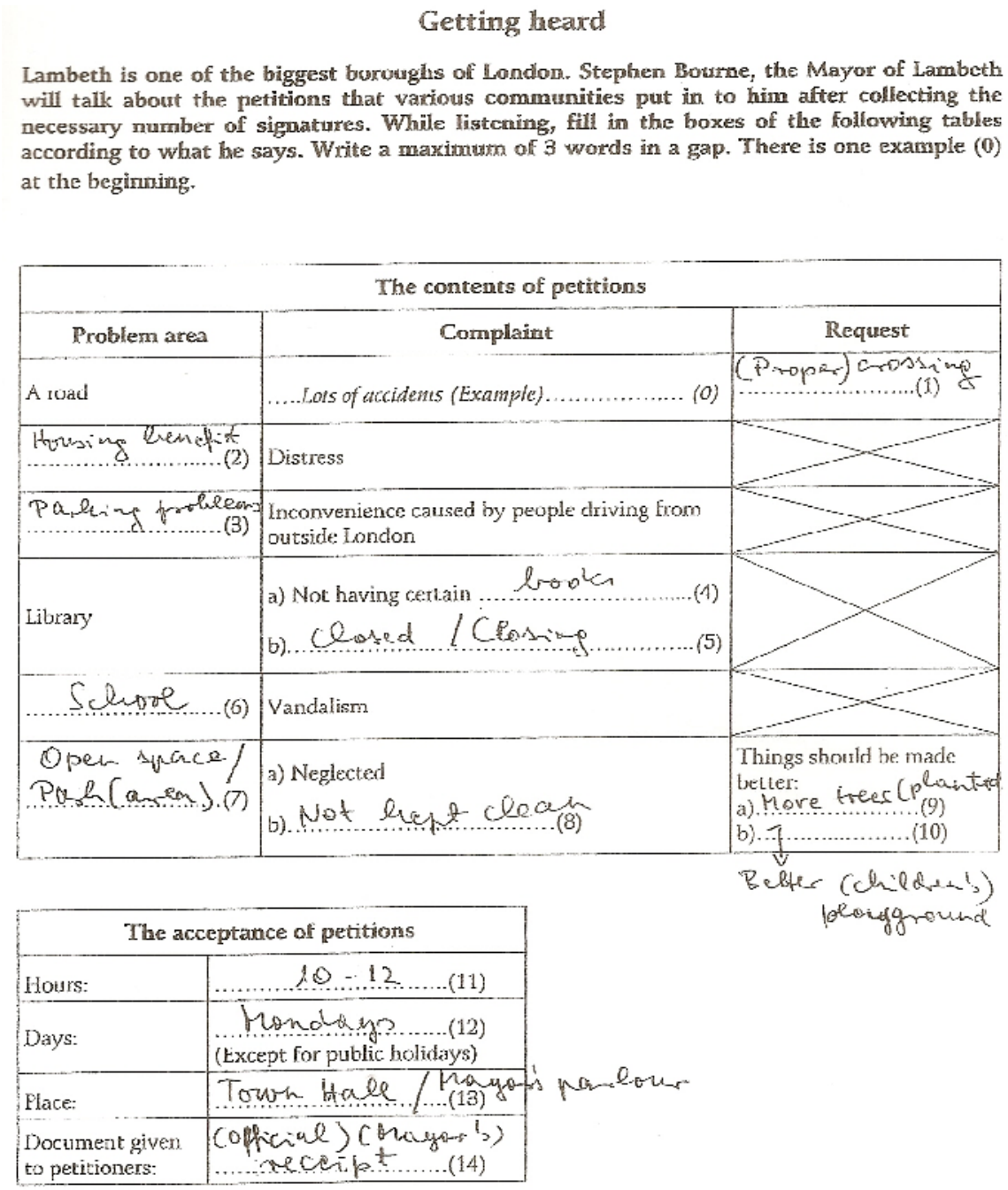
Barta, É. (2004). *Are you listening?* Budapest: Akadémiai Kiadó.

Depicted as an ape (Transcript)
- In half an hour Martin Jarvis will be reading another of the paradoxes of Mr Pond written by G K Chesterton after more riddles of the past have been untangled in Making History with Sue Cook.
- Hello again. But we'll start with the origins of life itself and Charles Darwin, the man who revolutionised our thinking about how we all began. After five years at sea as a young naturalist with a British scientific expedition, Darwin returned to London to work through all

the research notes he'd gathered and from which over the next twenty or thirty years his theories of evolution gradually grew. Theories which didn't really see the light of day until he was fifty. I'm joined now by the Darwin scholar, professor Janet Browne. The second volume of her biography of Darwin: 'Charles Darwin: The Power of Place' is just out and it deals with that later part of his life when his ideas were in the public arena. So Professor Browne, what was the general attitude to the man and his work?
- Behind that great big Victorian beard there really was a very good, decent man. He was a pillar of the local community. And he wasn't at all a fiery radical and so when he put forward these ideas the community responded to him as an expert, they responded to him as to somebody they trusted. Of course there were lots of rather vicious cartoons in the press, there were people who did criticise but in the end his social status and his evident wish to discover the truth helped that theory to become accepted.
- You mentioned cartoons. He was depicted as an ape or being accompanied by an ape or a monkey or an orang-utang on several occasions in magazines like 'Punch'. Did he take that kind of thing personally himself?
- No, he loved it. That's one of the delightful things about being able to work on Darwin. He surprises us at every turn. He collected all those cartoons and he showed them to friends, he had several framed and hung on the walls of his house and found them amusing.
- And he didn't seem to seek the limelight from what I've read about him.
- He liked very much to be in his house surrounded by his family, pottering around his garden, doing some research in the grounds of his quite substantial estate. He wasn't a public man in that way. He had lots of friends who did the public side of the work for him.
- He seemed to be quite non-confrontational about it. He ducks really any conflict with the church, didn't he?
- He never wanted to directly attack Christianity. If you read 'The Origin of Species' there's nothing in there that directly attacks faith or believers or the idea of the existence of God or the accuracy of the Bible. He just doesn't speak about that.
- Because his wife, Emma was a very devout Christian, wasn't she?
- Yes, his wife was a… simply wonderful woman and like many Victorians, a very good and devout believer. And there's some terribly tender letters written by her saying to him that the views that he was putting forward would mean that on the day of judgement he simply wouldn't go to heaven and that they wouldn't meet in the afterlife and she was anxious about the state of his soul. But I don't think ever tried to push her own views on him, as a good wife shouldn't..
- Professor Janet Browne, thanks very much indeed.
Time: 3'20"
Source of text: BBC Radio 4, Making history

APPENDIX B

Task 2



Getting heard

Lambeth is one of the biggest boroughs of London. Stephen Bourne, the Mayor of Lambeth will talk about the petitions that various communities put in to him after collecting the necessary number of signatures. While listening, fill in the boxes of the following tables according to what he says. Write a maximum of 3 words in a gap. There is one example (0) at the beginning.

The contents of petitions		
Problem area	**Complaint**	**Request**
A road	*.....Lots of accidents (Example).....* (0)	(Proper) crossing .....(1) &
Housing benefit .....(2)	Distress	
Parking problems .....(3)	Inconvenience caused by people driving from outside London	
Library	a) Not having certain books .....(4) b) Closed / Closing .....(5)	
School .....(6)	Vandalism	
Open space / Park(area) .(7)	a) Neglected b) Not kept clean .....(8)	Things should be made better: a) More trees (planted) .....(9) b) ↓ .....(10) Better (children's) playground

The acceptance of petitions	
Hours:	10 - 12 .....(11)
Days:	Mondays .....(12) (Except for public holidays)
Place:	Town Hall / Mayor's parlour .....(13)
Document given to petitioners:	(official) (Mayor's) receipt .....(14)

Barta, É. (2004). *Are you listening?* Budapest: Akadémiai Kiadó.

Getting heard (Transcript)

Hmmm. I always hold an opportunity to be in town hall for people in the community to present petitions to me. Because the way the system works in the council, they don't have necessarily very many opportunities to air a grievance by actually collecting signatures and saying we don't like this particular issue or we don't like that particular thing that's happening or there has been lots of accidents on this particular road and we need a crossing, a proper official crossing, please can you do something about it?

Other petitions would be perhaps people are having a lot of problems with their housing benefit perhaps. And it's causing so much distress in that particular location, at that particular community that they need… they feel that they have asked all the right questions but aren't getting anywhere so they want to make their voice known. So they will collect signatures and the more signatures they collect obviously it gives more weight to their problem.

It could be a parking problem. There might be a street that… the people who live in it can never ever park their cars because people who work in the city, driving from outside London, park their cars in their street and leave them there all day and they drive home in the evening. But during the day the people who live there can't find a parking space so again they will collect signatures.

It might be an issue to… with a library. That perhaps they don't have the sort of books that the library… or perhaps the library is being possibly closed. So therefore they want to make their voice known.

It could be perhaps a problem in a particular school. Either a lot of vandalism is taking place by pupils from that school… and they want something done about it. So they will collect again signatures.

It could be an open space, a park area. That's perhaps… they feel is … been very much neglected, isn't being kept clean properly or they feel there should be things that are made better in the park. More trees perhaps should be planted. Or there should be a better children's playground. So to make their voice heard, they would write as many signatures or collect as many signatures as they can from the community and present those to me, now the mayor.

So I always make myself available between ten and twelve on Mondays, every Monday except a public holiday, Monday, to be in the town hall, in the mayor's parlour. So if there are people who wish to bring such petitions to me can do so and meet me in person and I meet them in the mayor's parlour, talk to them about it, give them a receipt for it, so that they have an official mayor's receipt and promise that they will receive an answer within four weeks. So that's the sort of thing that happens on Monday.

Time: 2'40"

Source of text: Interview with Stephen Bourne (mayor of Lambeth, London) by Éva Barta

Appendix C

Interview prompts (Translation from Hungarian)

Pre-listening prompt
Please verbalize all your thoughts you have while looking at/reading the task before the listening.

Retrospective interview prompts
Please, as much as you can remember verbalize what thoughts you had while listening to the text and answering the item.

Guidelines:

- What made you give that answer? / What helped you in answering the question? (e.g.: understanding words or grammar, context, the situation suggested it, excluded improbable answers, based on logics, **anything else**)
- What other replies did you consider and why did you discard it?
- What answer did you definitely exclude as totally impossible?
- What made you uncertain about the reply? What disturbed you? (e.g.: you found something unambiguous, the speed of text, **anything else**)

Post-listening prompt
Please, verbalize all your thoughts you have while finalizing your answers after the second listening to the text.