

Assessing foreign language listening comprehension:

Exploring construct validity and task difficulty

The poster reports dissertation research which focuses on the development and evaluation of the English listening comprehension test used in DESI, a large-scale assessment project.

The study, which will have entered the reporting phase shortly before June 2005, seeks to provide very detailed and specific evidence for aspects of construct validity and explores student perceptions of task difficulty. Both objectives draw on qualitative data from the points of view of the test-takers, taking on a perspective of test evaluation Shohamy (2001) advocates for the investigation of the meanings and uses of tests. The study follows in the veins of current research into the listening construct (Buck 2001), test-taking processes (Cohen 2000) and strategy use (Purpura 1999).

The subjects are 18 test-takers (9 low and 9 high general foreign language ability as assessed by 4 C-test tasks) sub-sampled from a sample of 120 students that took the DESI tests. The sampling of extreme cases, it is hoped, will shed more light onto the possibly variable processes than other sampling procedures would have allowed (cf. Patton 1990). Subjects were asked to think aloud while trying to solve the multiple-choice listening items, as suggested by Grotjahn (1986). This procedure was used with the hope of tracing in the verbal protocols the kinds of knowledge and information understood and retained in working memory while the test-takers were trying to solve the tasks. Additionally, stimulated recall interviews were conducted to elicit data on perceptions of item difficulty.

Henning Rossa
University of Dortmund, Germany
Department of English and American Studies

henning.rossa@udo.edu

**ASSESSING FOREIGN LANGUAGE LISTENING:
EXPLORING CONSTRUCT VALIDITY AND TASK
DIFFICULTY**

Type of Study:

Introspective Multiple Case Study Research

Brown (2002), Denzin & Lincoln (2000), Yin (2003)

Subjects / Informants / Participants: Purposeful Random Sampling of
Extreme Cases

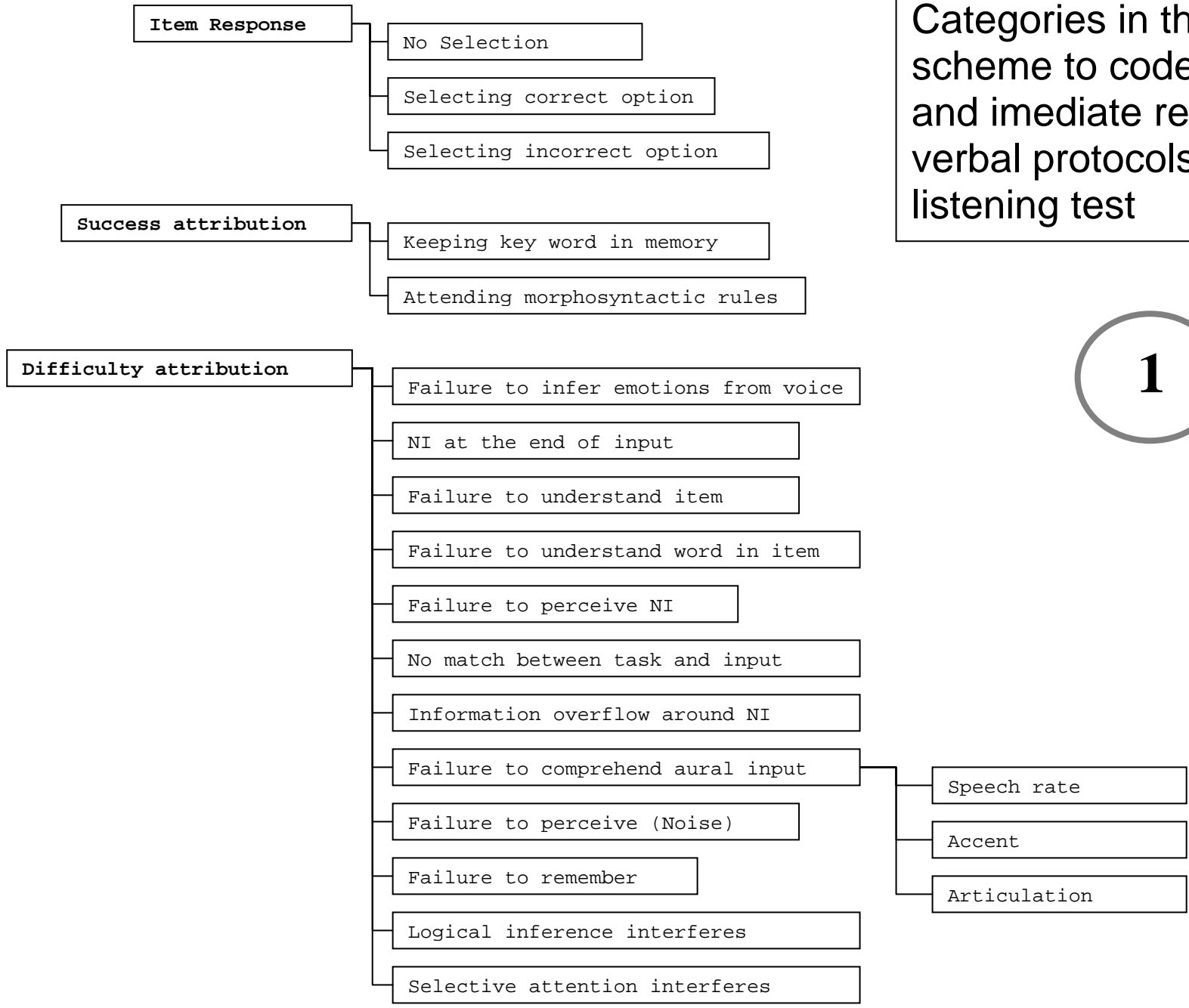
De Jong (2004), Maxwell (1996), Merkens (2003), Patton (1990), Johnson
(1990)

RESEARCH PURPOSES	RESEARCH QUESTIONS	METHODS OF DATA COLLECTION	METHODS OF DATA ANALYSIS
Construct Validation of an EFL Listening Test	What are the cognitive processes German grade 9 learners of EFL engage in when they try to solve the listening comprehension tasks in DESI?	<p>Qualitative: Think-Aloud Protocols (TAP) Bråten & Strømsø (2003) Cohen (1997, 2000) Ericsson & Simon (1980, 1987, 1993) Ericsson (2002) Grotjahn (1986) Someren et al. (1994)</p> <p>Quantitative: Test scores, IRT-scaling of person-ability parameters, CTT and IRT test and item parameters</p>	<p>Qualitative: Verbal Protocol Analysis (VPA) Chi (1997) Green (1998)</p> <p>External categories in the coding scheme from: Buck (1990, 2001) Rost (2002) Vandergrift (2003)</p> <p>Quantitative: Chi-square analyses of co-occurrence of processes and text-types / task-formats</p>
	How do these processes differ between high- and low-ability (general language proficiency) test-takers?		
	How do processes differ between different text types (dialogues and expository texts) and task formats (multiple-choice and oral summary)?		
	To what extent do these processes reflect elements relevant to the test construct and to what extent do they represent construct-irrelevant language processing?		

RESEARCH PURPOSES	RESEARCH QUESTIONS	METHODS OF DATA COLLECTION	METHODS OF DATA ANALYSIS
Validation of a System of Task Characteristics to Support Development and Evaluation of Listening Test Tasks	What is the nature of the test-takers' experiences of task difficulty in the EFL listening tests?	<p>Qualitative: Stimulated Recall Interviews (SRIV) Bloom (1954) Di Pardo (1994) Gass & Mackey (2000)</p> <p>Quantitative: Test scores, IRT-scaling of item-difficulty parameters</p>	<p>Qualitative: Interpretative Phenomenological Analysis (IPA) Smith (1996) Smith et al. (1999) Giles (2002)</p> <p>Quantitative: Regression analysis of task-characteristics and item-difficulty parameters</p>
	How does the phenomenon of difficulty differ in the perceptions of high- and low-ability (general language proficiency) test-takers?		
	How can these perceptions be linked to expert ratings of task characteristics and how can they be used as a resource for the development and evaluation of listening test tasks?		

The DESI listening construct

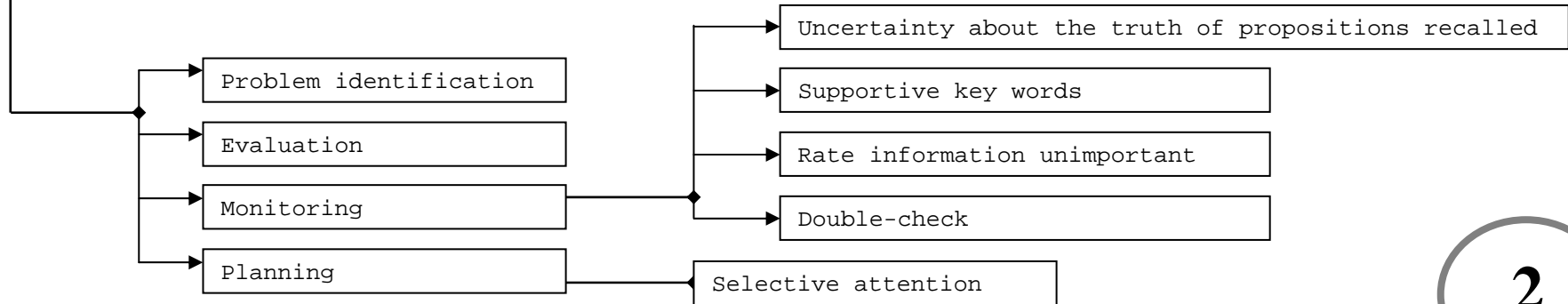
- Processing short and extended samples of spoken language (English [Near-RP and General Canadian], authentic speech rates, generally clear articulation, scripted texts) in real time.
- Understanding the linguistic information that is presented on the local level of the input text (understanding details)
- Connecting pieces of information in order to develop a mental model which allows comprehension on the global levels of the input text (understanding gist)
- Matching explicitly and implicitly presented information (actions, emotions, intentions) with language knowledge and background knowledge to recognize and retrieve, to infer, and to interpret this information.
- Constructing a representation of information presented in the aural mode that allows the listener to understand paraphrases of that information in other (written) contexts.



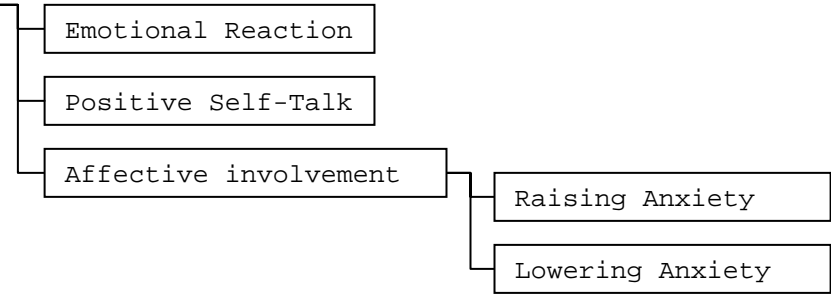
Categories in the coding scheme to code concurrent and immediate retrospective verbal protocols of taking the listening test

1

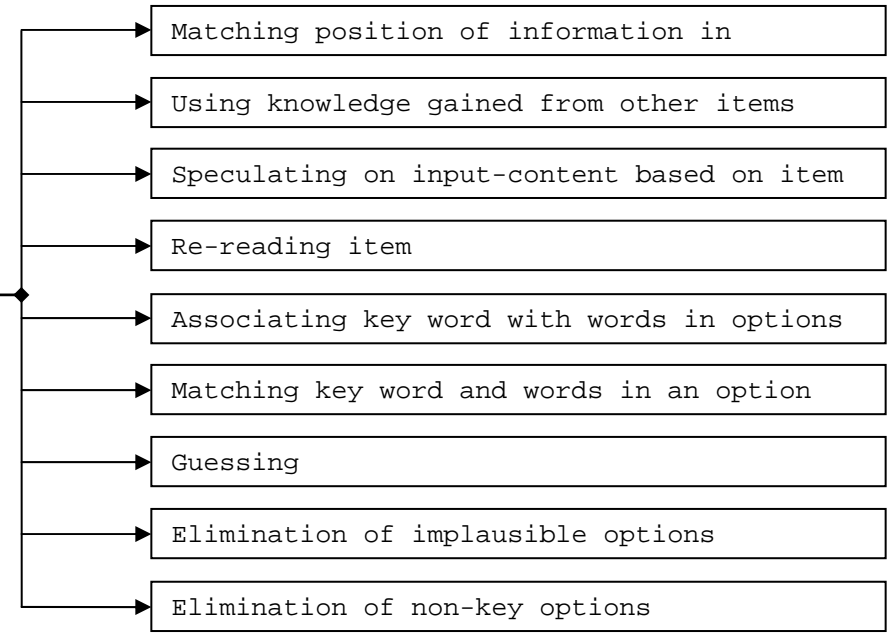
Metacognitive Strategies



Affective Strategies



Test-taking Strategies



Cognitive Strategies

