Teachers' attitudes toward African American
Vernacular English: influence of contact with
linguistics on ambivalent attitudes

Von der Philosophischen Fakultät der Rheinisch-Westfälischen
Technischen Hochschule Aachen zur Erlangung des
akademischen Grades einer Doktorin der Philosophie
genehmigte Dissertation

vorgelegt von
Judith Bündgens-Kosten, geb. Bündgens

Berichter:
Universitätsprofessor Dr. phil Paul Georg Meyer
Universitätsprofessor Dr. phil. Rudolf Beier

Tag der mündlichen Prüfung: 9.7.2009

Diese Dissertation ist auf den Internetseiten der
Hochschulbibliothek online verfügbar.

Judith Bündgens-Kosten

2009

Teachers' attitudes toward African American Vernacular English: influence of contact with linguistics on ambivalent attitudes

This work is licensed under the Creative Commons

Attributions-Non-Commercial-No Derivative Works 3.0 Germany License.

To view a copy of this license, please visit

http://creativecommons.org/licenses/by-nc-nd/3.0/de/deed.en_GB;

or send a letter to Creative Commons, 171 2nd Street, Suite 300, San Francisco, California, 94105, USA.

Contents

Acknowledgments	x
A word concerning political correctness and choice of terminology	xii
1 Introduction	1
2 Theoretical background	3
2.1 The attitude object	3
2.1.1 AAVE as a linguistic concept	4
2.1.2 AAVE as a popular concept	7
2.1.3 Teachers' perspective	10
2.2 Defining "attitude"	10
2.2.1 A short history of research	10
2.2.1.1 Pre-Thomas	11
2.2.2 Attitudes in social psychology	11
2.2.2.1 Thomas and Znaniecki	11
2.2.2.2 "Attitudes can be measured"	12
2.2.3 History of research on language attitudes	13
2.2.3.1 History of AAVE-attitudes research	16
2.2.4 A working definition of "attitude"	18
2.2.4.1 Purpose of attitudes	18
2.2.4.2 Origin of attitudes	19
2.2.4.3 Delineation from related terms	19
2.2.4.4 Attitudes as hypothetical constructs	20
2.2.5 Different measures of attitudes	21
2.2.6 The internal structure of attitudes	22
2.2.7 Attitudes: the sociolinguistic perspective	23
2.3 Measuring attitudes	25
2.3.1 Theoretical problems in measuring attitudes	25
2.3.1.1 A hypothetical construct	26

2.3.1.2	Artifacts	26
2.3.1.3	The nature of the stimulus	27
2.3.1.3.1	Matched guise	28
2.3.1.3.2	By any other name: The politics of choosing a label	30
2.3.1.4	Scale level	32
2.3.2	Short overview over attitude measuring techniques . . .	33
2.3.2.1	Stimulus-then-person-scaling	34
2.3.2.1.1	Thurstone	34
2.3.2.1.1.1	Method of equal-appearing in- tervals	34
2.3.2.1.1.2	Method of paired comparisons	34
2.3.2.1.1.3	Method of successive intervals	35
2.3.2.1.1.4	Problems	35
2.3.2.1.2	Magnitude estimation (Stevens) . . .	36
2.3.2.2	Simultaneous-stimulus-and-person-scaling . .	36
2.3.2.2.1	Guttman	36
2.3.2.2.2	Edwards-Kilpatrick scale	37
2.3.2.3	Person scaling	38
2.3.2.3.1	Likert scaling(s): Method of sum- mated ratings	38
2.3.2.3.1.1	“Likert-type scale”	38
2.3.2.3.1.2	Sigma method	43
2.3.2.3.1.3	Simpler method	43
2.3.2.3.1.4	Summary	44
2.3.2.3.2	Osgood: Semantic differential	44
2.3.2.4	Attitude measures linked to specific indicator classes	47
2.3.2.4.1	Using cognitive indicators	47
2.3.2.4.2	Using affective indicators	49
2.3.2.4.2.1	Self-Report of affect	50
2.3.2.4.3	Using behavioral indicators	50
2.3.2.4.3.1	Content analysis	52
2.3.3	The shortlist: Likert, Osgood and Thurstone scaling . .	53
3	Review of literature	56
3.1	Osgood and pseudo-Osgood studies	57
3.1.1	Williams	57
3.1.1.1	The Chicago study	57
3.1.1.1.1	The subjects	57
3.1.1.1.2	The audio samples	58

3.1.1.1.3	The questionnaire	58
3.1.1.1.4	Results	60
3.1.1.1.4.1	The factors	60
3.1.1.1.5	Evaluation	64
3.1.1.2	The Texas study	64
3.1.1.2.1	The subjects	64
3.1.1.2.2	The video samples	64
3.1.1.2.3	The questionnaire	65
3.1.1.2.4	Results	66
3.1.1.2.5	Evaluation	67
3.1.2	Cecil	67
3.1.2.1	The subjects	68
3.1.2.2	The audio samples	68
3.1.2.3	The questionnaire	68
3.1.2.4	Results	69
3.1.2.5	Evaluation	70
3.1.3	Summary	70
3.2	Likert and pseudo-Likert studies	70
3.2.1	Di Giulio	71
3.2.1.1	Subjects	71
3.2.1.2	The questionnaire	71
3.2.1.3	Findings	72
3.2.1.4	Evaluation	74
3.2.2	Studies employing the Language Attitude Scale	74
3.2.2.1	Taylor	74
3.2.2.1.1	Subjects	74
3.2.2.1.2	The questionnaire	74
3.2.2.1.3	Findings	78
3.2.2.1.4	Evaluation	80
3.2.2.2	Covington	80
3.2.2.2.1	Subjects	81
3.2.2.2.2	The questionnaire	81
3.2.2.2.3	Findings	82
3.2.2.2.4	Evaluation	82
3.2.2.3	Ford	83
3.2.2.3.1	Subjects	83
3.2.2.3.2	The questionnaire	83
3.2.2.3.3	Findings	84
3.2.2.3.4	Evaluation	85
3.2.2.4	Bowie	86
3.2.2.4.1	Subjects	86

3.2.2.4.2	The questionnaire	86
3.2.2.4.3	Findings	86
3.2.2.4.4	Evaluation	88
3.2.3	Hoover	88
3.2.3.0.5	Evaluation	93
3.2.3.1	Abdul-Hakim	93
3.2.3.1.1	Subjects	93
3.2.3.1.2	The questionnaire	93
3.2.3.1.3	Findings	94
3.2.3.1.4	Evaluation	95
3.2.4	Blake and Cutler	95
3.2.4.1	Subjects	95
3.2.4.2	The questionnaire	96
3.2.4.3	Findings	97
3.2.4.4	Evaluation	99
3.2.5	Summary	99
3.3	Studies employing other methods	99
3.3.1	Hoover	100
3.3.1.1	The audio samples	100
3.3.1.2	The questionnaire	101
3.3.1.3	Evaluation	103
3.3.2	Woodwort & Salzer	103
3.3.2.1	Subjects	104
3.3.2.2	The stimulus	104
3.3.2.3	Results	105
3.3.2.4	Evaluation	105
3.4	Overview	105
4	Changing attitudes	108
4.1	The theory of attitude change	108
4.2	Popular claims	109
4.3	Language attitudes and linguistics: a side study	110
4.3.1	The subjects	110
4.3.2	The questionnaire	112
4.3.3	Results	114
4.4	Language attitudes and linguistics: outlook on the main study	123
5	Study	125
5.1	Methods	125
5.1.1	The questionnaire	125
5.1.1.1	Choice of terminology	126

5.1.1.2	Constructing the scale	128
5.1.1.2.1	Selection of items	128
5.1.1.2.2	Paired comparisons	131
5.1.1.2.3	Scale construction	131
5.1.1.2.4	Internal consistency check	135
5.1.1.2.5	The finished scale	135
5.1.1.3	Additional elements of the questionnaire . . .	138
5.1.1.3.1	Self-evaluation of contact with lin- guistics	138
5.1.1.4	Technical characteristics of the questionnaire	139
5.1.2	Conducting the study	139
5.1.2.1	Subjects	139
5.1.2.1.1	Make-up of the sample	140
5.2	Findings	146
5.2.1	Overall attitudes	146
5.2.1.1	Latitude of acceptance	147
5.2.1.2	Contact with linguistics	151
5.2.2	Contact with linguistics and effect on language attitudes	151
5.2.2.1	Perception of attitude change	154
5.2.2.2	Linguistic epiphanies	156
5.2.2.3	Latitude of acceptance and contact with lin- guistics	157
5.2.3	Summary	158
6	Conclusion	159
6.1	Ambivalence and prior research	159
6.2	Ambivalence and attitude change	160
6.3	A final thought	168
	Bibliography	170
	Index	179
	Calculations belonging to the scale construction	181
	Calculations belonging to the internal consistency check	183
	Text of the main questionnaire	185

List of Figures

2.1	‘Levels’ of AAVE	8
2.2	Timeline	17
2.3	Likert vs. Thurstone items	55
4.1	Cognitive Dimension - Aachen	115
4.2	Affective Dimension - Aachen	116
4.3	Behavioral Dimension - Aachen	117
4.4	Cognitive Dimension - Siegen	118
4.5	Affective Dimension - Siegen	119
4.6	Behavioral Dimension - Siegen	120
4.7	Question 4	123
5.1	Flowchart of the paired comparisons questionnaire	133
5.2	Sample: Census Regions and Divisions	142
5.3	Extract from data	148
5.4	Latitude of acceptance	149
5.5	Types of attitude change	153

List of Tables

2.1	Hoffman: three incarnations of the Ebonics debate	15
2.2	Phases of AAVE attitude research	16
2.3	Labels used	30
2.4	Scalogram	37
3.1	Summary: samples	56
3.2	Summary Osgood scales	70
3.3	Subjects according to region	75
3.4	Percentage of negative, mildly negative, neutral, mildly positive and positive options chosen	79
3.5	LAS scores	84
3.6	Trichotomies Hoover / Abdul-Hakim	94
3.7	Blake and Cutler 2003 results	98
3.8	Overview over studies employing (pseudo-)Likert scales	100
3.9	Summary all methods	107
4.1	Age of subjects (Aachen)	111
4.2	Age of subjects (Siegen)	112
4.3	Overview questionnaire items	113
4.4	Coding	114
4.5	Binomial test results Aachen	121
4.6	Binomial test results Siegen	122
5.1	Labels recognized by at least 50% of teachers	126
5.3	List of items selected for scaling	132
5.5	Scale values of items	136
5.7	Sample: Ethnicity	141
5.8	Sample: Census Regions and Divisions	141
5.9	Teacher population: Sex	143
5.10	Teacher population: Ethnicity (simplified)	144
5.11	Teacher population: States	145

5.12	Teacher population: School type	146
5.13	Agreement and disagreement with items	146
5.14	Agreement and disagreement with items in order of scale value	147
5.15	Latitude of acceptance	148
5.16	Contact with linguistics	151
5.17	Chi-square results never contact vs. lots contact group (rounded)	152
5.18	Fisher's exact test results never contact vs. lots contact group	152
5.19	Chi-square results contact vs. no contact group (rounded) . .	152
5.20	Fisher's exact test results contact vs. no contact group . . .	153
5.21	Significant difference in answer behavior lots contact vs. never contact group	155
5.22	Significant difference in answer behavior contact vs. no con- tact group	155
5.23	Latitude of acceptance across contact groups	157
1	F-matrix	181
2	P-matrix	181
3	Z-matrix	182
4	Column differences	182
5	Z' matrix (rounded)	183
6	P' matrix (rounded)	184
7	Difference between P and P' matrix (rounded)	184

Acknowledgments

Prof. Dr. Paul-Georg Meyer acted as thesis supervisor for this work. I would like to thank him for his support during every phase of this thesis project.

Prof. Dr. Rudolf Beier gave valuable feedback, especially concerning Section 6.2.

I would like to thank Dr. Sabine Arndt-Lappe and Dr. Maria Braun (University Siegen), who collected data for me for the study discussed in Chapter 4. I cannot thank the German Postal Service that lost Dr. Braun's data.

Robert Kosten did all PHP and SQL programming needed for the electronic questionnaires. He also designed Figure 2.2 and gave feedback on the readability of my first draft.

Dr. Paula Niemietz corrected my draft of the attitude questionnaire and proofread this text.

Two anonymous reviewer gave valuable feedback to an article version of parts of Chapter 4.

I would like to thank the administration, teachers and students of the following Californian schools for giving me the opportunity to observe 'teaching in action': Castlemont Business and Information Technology School, Oakland (Principal Rick Gaston), Monroe Elementary School, San Francisco (Principal Mark Bolton), Dorsey High School, Los Angeles (Principal George Bartleson).

My thanks also extend to all individuals who volunteered their time to answer my questions and to fill out my questionnaires, all across the U.S., and at the Universities of Siegen and Aachen.

Medienartig/SourceLounge, Aachen, provided free webhosting for the online questionnaires.

A word concerning political correctness and choice of terminology

For a non-native speaker it is not always easy to evade all the pitfalls of political correctness a language has to offer. This is more a problem of cultural than of linguistic knowledge. I would like to stress that, in case any individual feels slighted by my choice of terminology, that this has never been my intention.

Throughout this thesis I use the terms “African American” and “European American” to refer to the groups traditionally and/or historically referred to as “Blacks” / “Negroes” on the one hand and “Whites” / “Caucasians” / “Anglos” on the other. Some of these terms do not share exactly identical extensions, e.g. an American citizen of Egyptian origin might be considered an African American (since, indisputably, Egypt is an African country), but not to be a “Black” (as traditionally identified by skin color). This might, theoretically, have a minor distorting effect.

In some contexts, especially when discussing older research on this topic, I will present quotations that contain words strongly frowned upon today. These quotations should be viewed in their historical contexts.

For the form of speech under discussion here I have chosen the term “African American Vernacular English” / “AAVE”. I am conscious of the fact that this term is viewed critically by those who stress the influences of African languages on AAVE more than the influence of varieties of English. I also know that my choice of calling AAVE a “variety” instead of a “language”, and of referring to AAVE as “basilect” rather than “low-prestige language” might be debatable. These choices have no bearing on my research, though, since the measurement of attitudes functions in identical ways, whether I consider the attitude object to be a language or a variety.

Chapter 1

Introduction

If the ‘Pygmalion effect’ (or ‘Rosenthal effect’) assumption holds, the attitudes teachers have toward their students - including their students’ language - can influence teacher expectations and, in the long run, students’ educational chances. Such a relationship would be especially critical in the context of basilectal language varieties/low prestige languages, where it could lead to lowered expectations in teachers with negative attitudes toward these languages and language varieties. If there existed a method of influencing language attitudes that were proven effective, cost-efficient and ethical, such effects could be avoided¹. Claims have been made in the linguistic literature that teaching linguistics is such a means of influencing language attitudes (cf. Chapter 4.2).

Even if the Rosenthal effect assumption does not hold, the relationship between knowledge about language and attitudes toward languages and language varieties is still of theoretical interest. Previous studies focusing on this question have come to different conclusions, and different models of attitude change treat the question of how knowledge/new information influences attitudes in different ways (cf. Chapter 4.1).

In this thesis, I will discuss two studies I have undertaken to clarify this question. The first one, a pre-test post-test format study using a Likert-style questionnaire checked the claim that linguistic knowledge improves language attitudes for a German context. The second one, a ex-post-facto format study using a Thurstone scale derived via the method of paired comparisons tested it for U.S. teachers’ attitudes toward AAVE.

Before presenting my own research, I will first discuss the notion of “attitude” in some detail (Section 2.1), assess different forms of attitude measurement

¹There is another approach to reducing or avoiding the Pygmalion effect, which I will not discuss further in this thesis. Some researchers believe that making teachers aware of this effect may already help in reducing it (cf. Cecil 1988, 36).

(Section 2.3), and give a detailed review of previous research on teachers' attitudes toward AAVE (Chapter 3).

Chapter 2

Theoretical background

2.1 The attitude object

Attitudes, i.e. psychological tendencies that are expressed by evaluating a particular entity with some degree of favor or disfavor (cf. Eagly and Chaiken 1993, 1f; see discussion at 2.2.4), are always attitudes *toward* something, toward an *attitude object*. Attitude objects can be abstract or concrete, they can be particular objects or classes of objects, and they can also be behaviors and classes of behaviors (cf. Eagly and Chaiken 1993, 5). You can have an attitude toward truth, and toward the chair you are sitting on at this moment, toward this specific book and books in general, toward the act of writing a PhD thesis, and acts of academic scholarship in general. “Virtually anything that is discriminable can be evaluated and therefore can function as an attitude object” (Eagly and Chaiken 1993, 4f).

The attitude object of interest in this study is African American Vernacular English (AAVE) - as a *langue* it is abstract, as *parole* it appears as concrete behavior. In measuring attitudes toward AAVE, attitudes toward AAVE as an abstract unit and those toward AAVE being spoken by actual individuals in actual situations cannot easily be distinguished.

Before I discuss the nature of attitudes, I will first try to throw some light on the attitude object of relevance to this study. Using the linguistic understanding of this term as a basis, I will contrast it with the folk linguistic interpretation and attempt to position the subgroup of teachers between these two understandings of AAVE. When you measure teachers’ attitudes toward AAVE, you will measure attitudes toward what **teachers** comprise under this label, not what a linguist’s definition of *AAVE* encompasses.

2.1.1 AAVE as a linguistic concept

The linguistic study of AAVE began in the 1880s - at this time still labeled *Negro English* (cf. Smitherman-Donaldson 1988, 149). Times have changed since the 19th century, and so have labels. *AAVE* is among the most common terms, strongly contested by *Black English* and *Ebonics*¹. The number of alternatives is legion; some are still used today, others have only historic significance, and some never developed much significance to begin with: *African American English*, *African American Language*, *Afro-American English*, *Afro-American Speech*, *Black English Vernacular*, *Black Idiom*, *Black language*, *Black slave language*, *Black speech*, *Negro Dialect*, *Negro Non-Standard*, *Nonstandard Negro English*, *Substandard Negro English*, *Vernacular Black English*, and *Spoken Soul* (cf. Rickford and Rickford 2000, 169; Wolfram and Thomas 2002, xiiv; Smitherman-Donaldson 1988, 151; Taylor 1975, 29; Smitherman 1973, 828) is probably only an incomplete list of labels used by linguists over the decades.

This multitude of names already implies that the study of AAVE is not characterized by universal agreement. There are, though, certain points with which most linguists would agree:

- AAVE is spoken by many African Americans and some non-African Americans. E.g. Dillard 1972 states:

The best evidence we have at the present time - and it is admittedly incomplete - indicates that approximately eighty percent of the Black population of the United States speaks Black English. (*Black population*, in this case, would mean all those who consider themselves to be members of the “Black” or “Negro” community). There are others who speak the same dialect; for example, many members of the Puerto Rican community in New York have learned Black English in addition to Puerto Rican Spanish. In the past, there were - as has been indicated above - many white speakers of the Black variety of English, especially among the Plantation-owning class. (Dillard 1972, 229)

¹If you prefer to settle the question of popularity of a term via the ‘Google fight’ method, *Ebonics* is clearly the winner. On 05/26/2006, Google.com found circa 974.000 hits for *Ebonics*, 301.000 hits for *AAVE* (not all of these *AAVE*s must necessarily refer to the language variety) plus 86.000 hits for *African American Vernacular English* and 298.000 hits for *Black English*. In the linguistic literature, on the other hand, *Ebonics* is far from being as dominant as on the internet. A detailed discussion can be found in Buendgens-Kosten 2006.

- AAVE is characterized by certain morphological, syntactic, lexical, and phonological features. A comprehensive list of such features can be found in Rickford 1999 (4ff). Not all speakers of AAVE will employ all features, or employ them with equal frequency (cf. Rickford 1999, 9), therefore such lists can offer only a rough orientation.

Most features are also shared by (other) varieties of English², though differences in degree might exist, e.g. AAVE consonant cluster simplification is more frequent in AAVE and occurs in more positions than in White working class speech (cf. Rickford 1999, 11). Even some stereotypical AAVE features also occur with other varieties of English, e.g. the often mentioned negative concord/double negatio

The fact that not all speakers of AAVE use all features to the same degree poses a problem for the identification of speakers. The extremes are simple: An African American person who never uses any features of AAVE (not even the distinct intonation pattern), nor has done so in his/her childhood, and who has occasional difficulties interpreting AAVE sentences correctly (e.g. the frequent misinterpretation of the remote present perfect), is definitely not a speaker of AAVE. An African American person who uses most or all AAVE features constantly (including very formal occasions), both in speech and in writing, and who ‘fails’ sentence repetition tasks (automatically ‘translating’ Standard American English (SAE) sentences into AAVE sentences while repeating them) is without doubt a speaker of AAVE, and probably a monodialectal one (though potentially passively bidialectal). A person using AAVE at home, and SAE in formal contexts is obviously both, bidialectal. But what about a person using SAE grammar and AAVE intonation? Or who uses SAE grammar in most situations, but proverbs and certain lexical items from AAVE? And a person who uses SAE grammar, but

²Wolfram 1991 reviews a list of eight candidates for exceptions originally compiled by Fasold 1981, (1) “devoicing of voiced stops in stressed syllables”, (2) “present tense, third person *-s/es* absence”, (3) “plural *-s/es* absence on general plurals (as opposed to such absence found on plurals of weights and measures)”, (4) “remote time *been*” (also known as *remote present perfect*, (5) “possessive *-s/es* absence”, (6) “reduction of final consonant clusters when followed by a word beginning with a vowel or when followed by a suffix beginning with a vowel”, (7) “copula and auxiliary absence involving *is* forms (as opposed to more generally deleted *are* forms)”, (8) “use of habitual or distributive *be*” (Wolfram 1991, 108). This list is not unproblematic. Third person singular *-s* absence occurs among non-AAVE speakers, but never in the frequencies observed among AAVE speakers (cf. Wolfram 1991, 109), habitual *be* has also been observed in Southern dialects (cf. Wolfram 1991, 109). He adds as further potential candidates *be done*-constructions, and certain constructions employing *steady* and *come* (cf. Wolfram 1991, 110). How many features of AAVE exist that are not shared by speakers of other varieties has not yet been established.

AAVE pronunciation?

It has been found that certain stereotypical (“marked”) features, even in isolation, can influence the perception of a person’s speech to a high degree (e.g. Labov 2008). This influences not only the overall evaluation of a speaker, but also the variety a speaker is perceived to be using. A person may be perceived as a speaker of AAVE by the general population if he or she employs the linguistic factors stereotypically associated with AAVE. For example, a person using AAVE pronunciation but SAE grammar will be perceived most likely as an AAVE speaker. A person using only a few grammatical features of AAVE in his/her speech, but features of the type that are stereotypically associated with this variety might be perceived as ‘more of an AAVE speaker’ than a person using the same number of ‘deviations’ from SAE, but only those that are not stereotypically associated with AAVE. If I ask a non-linguist about his opinion toward AAVE, he/she will most likely be unable to differentiate between “real” (“core” and “lame”) AAVE speakers and individuals who are practically SAE speakers with a few isolated, but stereotypical (often concerning pronunciation rather than syntax or morphology) AAVE features (“periphery speakers”). I call this the “*ax*”-*phenomenon*, based on the anecdotes quoted in Lippi-Green 1997, 179f)³. This group of speakers might normally not be included in studies of AAVE, but in this context, an isolation from “core speakers” is practically impossible, since their speech is perceived similar to that of “core speakers” by the general population⁴.

³The variant pronunciation of *ask* is one of those stereotypical features which are strongly associated with AAVE. Lippi-Green quotes a statement by Edward I. Koch, then mayor of New York City:

On the last day that I met with my adopt-a-class last year, I told the students that they will have to learn to read, write, do math, and speak English properly if they are going to get a first-rate job and be a success. I told them there was one word that will mark them as uneducated ... A young girl raised her hand and said, “The word is *ax*.” ... I asked her if she could pronounce the word properly. She said, “Yes, it is *ask*.” ... I felt terrific. By simply raising that one word on an earlier occasion, I had focused their attention on something that I think is important, and I am sure you do as well. (Lippi-Green 1997, 179f)

⁴By stating that certain features stereotypically associated with AAVE move individuals overproportionally closer to AAVE in the perception of the hearer, I do not claim that listeners are unable to react differently to different ‘degrees’ of non-standardness. William Labov has demonstrated that speakers can distinguish between language samples with differing frequencies of word-final SAE /ŋ/ pronounced as /n/ (Popularly known as “*g*-dropping”). The evaluative reaction to a speaker with e.g. 50% and 70% occurrence were noticeably different (cf. Labov 2008). In a study on language attitudes not utilizing

For the purpose of this paper, therefore, I will consider those individuals to be speakers of AAVE who use AAVE grammar and/or pronunciation (excluding the use of AAVE intonation without any other phonetic or phonological characteristics) at least to some degree, at least occasionally. These are the “core speakers”, who use most or all features of AAVE to a high frequency, the “lames”⁵, who use most of these features, but with somewhat lower frequencies, and the “periphery speakers”, which use significantly fewer features, but are still **perceived** as AAVE speakers. Of course, all three types of speakers can, in addition, be functional speakers of any other language and/or variety, including SAE. Those in the periphery are very likely to be so, some of those belonging to the core might only be passively bidialectal (and possibly even that only imperfectly). I do exclude, though, those belonging to the corona:

- Individuals speaking SAE, but employing AAVE verbal strategies.
- Individuals speaking SAE, but employing AAVE intonation.
- Individuals speaking SAE, but (occasionally) using isolated AAVE vocabulary items and/or proverbs.

These individuals partake in the culture surrounding AAVE, but are not fully functional speakers of this variety. I assume that they are not identified as AAVE speakers by the general population. Whether or not this is the case might be an interesting question for perceptual dialectology.

2.1.2 AAVE as a popular concept

Awareness⁶ of the existence of AAVE can be expected from the absolute majority of the U.S. population. Sandra L. Barnes asked 420 students whether they knew “Ebonics” or the “Oakland resolution” (Barnes 1998, 22) and found that more than 90 % knew one or both:

355 respondents are aware of both Ebonics and the Oakland resolution and 30 respondents are aware of Ebonics but are not knowledgeable about the Oakland resolution. Twenty-six students are

spoken language samples, this should have very little effect, though.

⁵This term has been borrowed from William Labov (cf Labov 1972). I use it in a wider sense here than the one it was originally designed to convey.

⁶You can be (a) aware of the existence of an object, i.e. “I know that there is a form of speech called *Ebonics*”, or you can (b) be aware of specific features of an object, i.e. “Ebonics uses double negation, and people drop the ends of words.”. In this section, I use both senses. The research of Barnes 1998 uses *awareness* in the first sense, the part of the work of Niedzielski and Preston 2003 discussed here uses it in its second sense.

Figure 2.1: ‘Levels’ of AAVE

aware of the Oakland resolution, but have not previously heard of the Ebonics debate before exposure to the Oakland controversy. (Barnes 1998, 23)

This would mean that a total of 97.9% of students at this college had at least some awareness of AAVE. Even if the awareness-levels in the general population might be expected to be lower than in a sample of college-students, relatively high awareness-levels even in the general population are made plausible by these results.

Niedzielski and Preston, in their volume on “Folk linguistics” (Niedzielski and Preston 2003), analyze interviews with non-linguists regarding their attitudes to and (pseudo-)knowledge⁷ of language. Their findings include the following: Regarding language in general:

- Non-linguists are prescriptionists who see language as a “Platonic ab-

⁷By using the term *pseudo-knowledge*, I stress that many assumptions held by laypeople are not in agreement with current linguistic knowledge. I do not wish to imply that lay theories of language are inherently worthless.

straction” (Niedzielski and Preston 2003, 18).

- Language varieties other than the standard variety are seen merely as deviations from “The language”. They are considered “performance deviations from competence, not alternative competencies” (Niedzielski and Preston 2003, 22).
- Awareness of specific language features is often limited (cf. Niedzielski and Preston 2003, 10ff, 24).

Regarding AAVE:

- Grammatical features that were mentioned by interviewees: negative concord, invariant *be* (also misinterpreted), hypercorrected *be*, third person indicative leveling, leveled participial (“I have went.”).
- AAVE is labeled as “accented”, but not every interviewee was able to give phonological details, either by explicitly stating phonological rules or by giving examples. Those details that are usually offered are: *r*-lessness, diphthongization, substitution of stops for labiodental fricatives (cf. Niedzielski and Preston 2003, 138f).
- Examples of lexicon differences were often mentioned (cf. Niedzielski and Preston 2003, 138f., 236).
- The non-linguist population does not always systematically distinguish AAVE from slang (cf. Niedzielski and Preston 2003, 139).
- It is generally considered to be rather easy to acquire a second variety. An inability to speak SAE is interpreted as an unwillingness to do so (cf. Niedzielski and Preston 2003, 140f).
- The majority position seems to be a “modified ‘bidialectal’ position” (Niedzielski and Preston 2003, 234). SAE has to be acquired in addition to the home variety, but the home variety as such is not respected beyond the acknowledgment of certain utilitaristic reasons (cf. Niedzielski and Preston 2003, 234).

They summarize their observations regarding folk linguistic opinions toward AAVE the following way:

Linguists may be heartened by the apparent fact that information about AAVE seems to be influencing folk knowledge, but all may be disappointed that it has not yet influenced attitudes as strongly as one might have hoped. (Niedzielski and Preston 2003, 141)

While I assume that there is sufficient awareness of AAVE to make possible that AAVE functions as an attitude object, non-linguists may differ from linguists in what exactly they subsume under this and under similar labels.

2.1.3 Teachers' perspective

In a side study (cf. Buendgens-Kosten 2006), I looked at how teachers are positioned between linguists and laypeople. The linguistic knowledge of the U.S. teachers in my sample (as assessed by recognition of basic and intermediate linguistic terms and names of linguists) was higher than is to be expected for the general population. In their language use concerning AAVE (recognition of and statements on usage of labels for AAVE), this group was much closer to the lay population than to linguists. This has ramifications for the design of programs aiming at teachers as well as of questionnaires used for studying this group, since when targeting teacher populations, researchers must be ready to deal with previous (and possibly outdated) knowledge, without being able to presuppose it.

I will discuss the influence of this side study on my choice of terminology in the questionnaire belonging to the main study in Section 5.1.1.1.

2.2 Defining “attitude”

In this subchapter, I will discuss the concept of ‘attitude’, including its (past and present) role in social psychological and sociolinguistic debate. The definition of *attitude* this study is based on will be discussed toward the end of this chapter.

2.2.1 A short history of research

Attitude is a term used by several disciplines, and to write a detailed history of attitude research would include writing its history not only in the framework of (social) psychology, sociology and sociolinguistics, but also ethnology, discourse analysis, and to a certain degree even the arts and biology⁸. This,

⁸In art theory, “(T)[t]he ‘attitude’ of the figure [of an actor or a statue, for example, addition mine] was the visible arrangement of its parts into a meaningful pattern, as in ‘an attitude of waiting’, or of devotion, or of nonchalance, for example.” (Danziger 1997, 134) In biology, *attitude* was used for example to refer to the expression of emotion in animals, or, even more generally, as ‘motor attitudes’, for reflex postural adjustments of animals (cf. Danziger 1997, 135). This usage of *attitude* is very distant from today’s usage in social psychology or sociolinguistics, and even though the modern usage of *attitude* developed in part out of these early scientific usages of the term, they do not help in clarifying the

of course, cannot be achieved here. I will limit myself to a short delineation of its history in the then newly emerging discipline of social psychology, and the subsequent adoption of the term *attitude* by sociolinguists.

2.2.1.1 Pre-Thomas

Even though the term *attitude* can be found in earlier (pre-1920s) psychological literature, it is used there in a sense very different from current usage. *Attitudes* then usually meant either *conscious attitudes*, a rough translation of the German term *Bewußtseinslage*, referring to what subjects described as going on mentally while they were solving tasks. Or it could refer, especially in behavioristic psychology, to the concept of ‘motor attitudes’, i.e. reflex postural adjustments of animals (cf. Danziger 1997, 135). These senses of *attitude* have to be kept separate from the sense of ‘social attitude’ with which it is usually used today.

2.2.2 Attitudes in social psychology

2.2.2.1 Thomas and Znaniecki

Social attitudes, used in a way that already approaches our contemporary understanding, were imported into psychology from sociology, especially from W.I. Thomas’s and Florian Znaniecki’s⁹ influential study “The Polish peasant in Europe and America” (cf. Danziger 1997, 140). Under *attitude* they sum all “subjective characteristics of the members of a social group” (Thomas and Znaniecki 1918, 20).

By attitude we understand a process of individual consciousness which determines real or possible activity of the individual in the social world. Thus, hunger that compels the consumption of the foodstuff; the workman’s decision to use the tool; the tendency of the spendthrift to spend the coin; the poet’s feelings and ideas expressed in the poem and the reader’s sympathy and admiration; the needs which the institution tries to satisfy and the response it provokes; the fear and devotion manifested in the cult of the divinity; the interest in creating, understanding, or applying a scientific theory and the ways of thinking implied in it - all these are attitudes. The attitude is thus the individual counterpart of

modern sense of the term.

⁹Although the study itself is already half footed in the emerging new subdiscipline of social psychology, Thomas and Znaniecki themselves are generally considered to be sociologists rather than social psychologists.

the social value; activity, in whatever form, is the bond between them. (Thomas and Znaniecki 1918, 22)¹⁰

Thomas and Znaniecki give some examples of attitudes that would not be considered attitudes today, e.g. hunger. Other examples, such as the tendency of a spendthrift to treat money carelessly, get much closer to today's usage.

This term was consequently adopted by psychologists - but not without a twist: the psychological aspect was strengthened, the sociological aspect weakened (cf. Danziger 1997, 141).

2.2.2.2 “Attitudes can be measured”

In the following decades, the term *attitude* came to dominate the still young subdiscipline of social psychology. Several renowned scientist went so far as to define social psychology as the study of attitudes (cf. McGuire 1986, 89)¹¹. Attitudes were ‘hot’ in the 1920s and this did not change when Emory Bogardus and L.L. Thurstone developed first means to quantify them - Bogardus¹² in 1925, Thurstone in 1928 (see also the separate chapter on Thurstone

¹⁰On the previous page, “a foodstuff, an instrument, a coin, a piece of poetry, a university, a myth, a scientific theory” (Thomas and Znaniecki 1918, 21) have all been defined as “values”.

¹¹Among those who defined social psychology this way were also Thomas and Znaniecki, contrasting social psychology as “science of attitudes” (Thomas and Znaniecki 1918, 27) and “general science of the subjective side of culture” (Thomas and Znaniecki 1918, 33) with sociology as the study of “one type of (...) values – social rules” (Thomas and Znaniecki 1918, 33), as “theory of social organization” (Thomas and Znaniecki 1918, 33), comprising both of them under “social theory” (cf. Thomas and Znaniecki 1918, 33).

¹²Emory S. Bogardus measured social distance, which allows conclusions to be drawn about attitudes (the higher the social distance, the more negative the attitude). He asked individuals into which groupings they would allow members of certain ‘races’. These groupings were graded according to social distance:

1. “To close kinship by marriage”
2. “To my club as personal chums”
3. “To my street as neighbors”
4. “To employment in my occupation in my country”
5. “Citizenship in my country”
6. “As visitors only to my country”
7. “Would exclude from my country” (Bogardus 1925, 301)

He used the results to calculate three indices: The Social Contact Range Index (S.C.R.), the Social Contact Distance Index (S.C.D.) and the Social Contact Quality Index (S.C.Q.).

2.3.2.1.1). Society's growing interest in attitudes toward social questions such as immigration and sexual morale, the success of modern means of opinion polling and the availability of government funding (especially during World War II (cf. Danziger 1997, 150f)) did nothing to reduce interest in attitude research, but rather helped to establish a growing 'attitude business' in the United States.

2.2.3 History of research on language attitudes

In order to position the research on attitudes toward AAVE historically, I will first discuss two attempts to structure the history of attitude research in social psychology (McGuire and Caspar I), then one attempt at providing a framework for the history of language attitude research (Bradac, Cargile and Hallett), and one that looks at language attitude research not from a social psychological, but from a sociolinguistic perspective (Caspar II). Finally, I will explain Hoffman's "Incarnations of the Ebonics Debate" and suggest my own interpretation of the historic development of AAVE attitude research, contrasted with the before-mentioned frameworks.

McGuire suggests that the history of attitude research within social psychology can be divided into three eras:

1. The **attitude measurement era**, 1920s and 1930s
2. The group dynamics interlude, 1935-1955
3. The **attitude change era**, 1950s and 1960s
 - (a) The convergent approach in the 1950s
 - (b) The divergent approach in the 1960s
4. The social cognition interlude, 1965-1985
5. The **attitude structure era**, 1980s and 1990s (prognosis only!)
 - (a) Structure of individual attitudes
 - (b) Structure within systems of attitudes

These three indices can be considered precursors of attitude measurement, especially of Guttman scaling (see Section 2.3.2.2.1). It was one of the first attempts to scale attitude-related concepts (Bogardus-'scaling' in itself does not measure attitudes but a closely related construction ("degrees and grades of understanding and feeling that persons experience regarding each other" (Bogardus 1925, 299))).

(c) Structures linking attitudes to other personal systems

(cf. McGuire 1986, 90) The first era, the attitude measurement era, corresponds to what was discussed above (“Attitudes can be measured”). The third era is only a prognosis, but it seems to be corroborated in part by Caspar I, discussed below.

Casper structures the history of attitude research within social psychology into seven phases: pre-Thomas (until 1918), Thomas (1918-1920), attitude measurement (circa 1925-1957), pessimism (60s), methods orientation, theoretical orientation (80s), phase of doubts (end of 80s) (cf. Casper 2002, 14-16).

What is quite obvious is that McGuire and Casper roughly agree in vital points, but differ greatly in details. Casper dates the phase of attitude measurement as lasting circa from 1925 till 1957¹³, while McGuire’s attitude measurement phase starts five years later and ends seventeen years earlier. Casper’s “pessimism” phase fits somehow with McGuire’s second era, since the pessimism dominating attitude research sprang from the recognition that attitudes and behavior were not as simply and systematically related as assumed before, a recognition that was, to a certain degree, an outcome of the research on attitude change. The more theoretical orientation of research she observed concerning the 80s supports McGuire’s prognosis of a “attitude structure era”. As a synthesis of both attempts of structurization, one might summarize: At the beginning of the twentieth century, stress was on developing means of measuring attitudes. After World War II, the practical relevance of attitude research was stressed (e.g. the relationship between attitudes and actions, and the manipulation of these). More recently, theoretical questions have become more salient.

Bradac, Cargile and Hallett see two main ‘paradigms’ of language attitude research. The first paradigm, “roughly 1960s to the mid-1970s” (Bradac et al. 2001, 138) was dominated by the work of Lambert. It was *atheoretical* and *acontextual*, depending on matched-guise (cf. 2.3.1.3.1) paper-and-pencil experiments (cf. Bradac et al. 2001, 139-141). The second phase developed out of this first one. It is characterized by a higher degree of theoretical orientation (cf. Bradac et al. 2001, 145) and attempts to remediate the *acontextuality* of research (cf. Bradac et al. 2001, 141). The most important detail here is that the “golden days” of language attitude research began relatively late. Attitude research had “had its go” for half a century before languages moved seriously into the focus of this discipline.

¹³Strictly speaking, she does not date this phase at all, but these years can be deduced from her list of relevant publications.

Caspar, in taking a sociolinguistic perspective, arrives at similar conclusions. She dates the beginning of sociolinguistic study of language attitudes around 1960. She suggests three phases: a phase of peripheric interest (1960s and 1970s) followed by one of growing interest (1980s) and one characterized by a concentration on empirical research (1990s).

Hoffman does not discuss attitude research at all, but focal points of public and scientific discussion of ‘Ebonics’/AAVE. He structures the hot phases of AAVE debate into three “incarnations” (see Table 2.1) With the growth of

Date	Trigger
1964	Educational attainment levels - Civil Rights movement
1979	Ann Arbor case
1996/7	Oakland Ebonics resolution

Table 2.1: Hoffman: three incarnations of the Ebonics debate

non-segregated education and the rise of the Civil Rights movement, differences between (non-mainstream) African American and (mainstream) European American children’s reading and writing attainment levels became more visible (cf. Hoffman 1998, 77). Against that backdrop, the year 1964 saw an impressive number of influential conferences, publications and research¹⁴(cf. Hoffman 1998, 78). While the second and third incarnations were to be rather shortlived, the first incarnation had sufficient drive to survive into the late seventies, even though the energy levels declined somewhat over the years (cf. Hoffman 1998, 80). The second incarnation was triggered by the Ann Arbor case, in which parents of AAVE speaking students sued their school board to improve the educational chances of their children. This court case is especially relevant for this thesis, since in it the judge decided that the AAVE-speaking students were hindered academically by their teachers’ negative attitudes toward their variety and ordered additional teacher training (cf. Joiner 1980).

The third incarnation was brought about by a resolution by the Oakland Board of Education, demanding a greater consideration of students’ native variety “Ebonics”/AAVE. The resolution’s wording was not unambiguous or uncontroversial, and reactions to it were immediate, passionate, and nationwide. Linguists and non-linguists, with few exceptions, represented contrary opinions, linguists laudating the basic idea while recommending slight changes in the formulation of the resolution, non-linguists misinterpreting it

¹⁴Even though Labov’s seminal work “The stratification of English in New York City” was published in 1966, it was finished in 1964

as a call for the demolition of Standard English teaching in schools. The debate was characterized by a high degree of emotional involvement (cf. e.g. Pandey 2000, Kretzschmar, Jr. 2008).

2.2.3.1 History of AAVE-attitudes research

Interest in research on attitudes toward AAVE peaked twice within the 20th century. The first peak could be observed during the 1970s, the second peak during the 1990s (see Table 2.2 for explorative list of studies). If we try to place these two phases of increased interest in AAVE attitude research within Hoffman’s structure of the AAVE debate, we see that the first phase of increased interest lingers somewhere between the first and second incarnation, preceding a phase of hot public debate rather than following one. Something similar can be observed for the 1990s rejuvenation in interest. It is likely to have influenced the 1996/7 Ebonics debate to a similar degree as it was, itself, influenced by that debate. If we compare the peaks in interest with the three eras of attitude research, we see that the first peak is positioned relatively late - only after the end of the second era. Sociolinguistics did not probe deep into the second era, at least not as far AAVE attitudes were concerned, but stayed for the most part on the level of the first era, i.e. was at least 40 years ‘out of date’ at its very beginning (cf. 2.2; an asterisk indicates that the article discusses an instrument for measuring attitudes toward AAVE without presenting results). The problem that caused the demise of the first era - the low correlations between measured attitudes and actions - was neglected entirely. Even in relationship to the ‘hot’ phases of language attitude research (both from a social psychological and a sociolinguistic perspective), AAVE attitude research was nearly a decade ‘late’.

Decade	Studies
1970s	Woodworth and Salzer 1971; Williams 1973; Taylor 1973; Di Giulio 1973*; Rosenthal 1974; Covington 1975; Williams et al. 1976; Shores 1977; Ford 1978
1980s	Cecil 1988
1990s	Speicher and McMahon 1992; Bowie and Bond 1994; Hoover et al. 1996*; White et al. 1998
2000s	Abdul-Hakim 2002; Blake and Cutler 2003

Table 2.2: Phases of AAVE attitude research

Naturally, ‘being late’ does not imply ‘being irrelevant’, especially since the dominant topics within social psychology do not necessarily possess the same

Figure 2.2: Timeline

importance for sociolinguists. This observation merely illustrates that, even though the term *attitude* was borrowed into sociolinguistics from social psychology, the research on attitudes by sociolinguists has developed rather independently from the research by social psychologists. This might have advantages, such as creating the possibility of a genuine sociolinguistic view on attitudes, but it also means that in some instances, research is conducted using techniques or concepts for which better alternatives might have been available.

2.2.4 A working definition of “attitude”

This study uses the definition of attitude brought forward by Eagly and Chaiken 1993 as a working definition:

Attitude is a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor. As we will explain in more detail, psychological tendency refers to a state that is internal to the person, and evaluating refers to all classes of evaluative responding, whether overt or covert, cognitive, affective, or behavioral/conative. This psychological tendency can be regarded as a type of bias that predisposes the individual toward evaluative responses that are positive or negative. (Eagly and Chaiken 1993, 1f)

In the main, Eagly’s and Chaiken’s definition reflects today’s scientific mainstream. It differs from other common ‘textbook’ definitions (e.g. Zimbardo 1992, 578) in avoiding the disputed term of ‘disposition’ and using the more general term ‘tendency’. Attitudes are usually considered to last a certain amount of time, but the minimum duration that a evaluative trend must exist before being called an ‘attitude’ is not generally agreed upon. Using the term ‘disposition’ implies a relatively long duration, while the term ‘tendency’ can encompass a wide range of different durations, including relatively short ones. Thus the second choice of term will find approbation among more social psychologists than the first one and is therefore preferable.

2.2.4.1 Purpose of attitudes

In this thesis, I will treat attitudes as a problem. Attitudes can lead to all kinds of actions, including discrimination. But attitudes per se are useful elements of our cognitive make-up, since they help structure information and can speed up cognitive processes or serve a number of other functions for the

individual who holds them (cf. Eagly and Chaiken 1993, 19f). It is neither possible nor expedient to aim at eliminating attitudes in general.

2.2.4.2 Origin of attitudes

Attitudes can be genetically encoded (e.g. disliking blue food), be transmitted through society (e.g. attitudes toward objects one has not encountered) or be derived from experience (e.g. attitudes toward dogs after having been bitten by a dog). In many instances, attitudes will stem from a combination of origins, e.g. learned but influenced by experience (cf. Eagly and Chaiken 1993, 3). This will also be the case with attitudes toward AAVE.

Language attitudes are acquired at a very young age. Marilyn S. Rosenthal (cf. Rosenthal 1974, also see section 2.3.2.4.3) has demonstrated the existence of attitudes toward AAVE in children aged 3-5.

2.2.4.3 Delineation from related terms

A number of constructs similar to that of ‘attitude’ exist. By discussing the differences between attitudes on the one hand and prejudices, stereotypes, values, character traits and moods, ideologies and schemas on the other hand, it should be possible to clarify the term *attitude* even further.

Attitudes versus prejudices Theoretically, a prejudice is an attitude toward any social group. Generally, though, the term is employed only to refer to negative attitudes toward minority groups (cf. Eagly and Chaiken 1993, 5; Taïffel 1972, 37). It might be argued that attitudes toward AAVE (or, more exactly, toward speakers of AAVE) constitute prejudices in so far as the language variety is associated with a minority group and is often negatively evaluated.

Attitudes versus stereotypes Stereotypes are “beliefs held about (...) [a] group” / “the attributes that an individual ascribes to a social group” (Eagly and Chaiken 1993, 104)¹⁵. Since beliefs (as forms of cognitive responding toward an attitude object) can function as indicators for attitudes (and prejudices), stereotypes on the one hand, and attitudes and prejudices on the other hand, are related: “people who are prejudiced in relation to a group

¹⁵There is also the notion of “a priori group stereotypes”, which stresses that stereotypes are *group* judgments, i.e. judgments shared *by* a group (cf. Schenk 1972, 273).

are generally thought to have a negative stereotype about group members” (Eagly and Chaiken 1993, 104).

Attitudes versus values A value is a specific type of attitude, one “toward relatively abstract goals or end states of human existence (e.g., equality, freedom, salvation)” (Eagly and Chaiken 1993, 5). Attitudes toward AAVE may be influenced by such values (cf. Katz et al. 1986, 42ff).

Attitudes versus character traits and moods An attitude can only be observed when it is elicited by a stimulus (e.g. seeing the attitude object). Character traits and moods, on the other hand, can be observed independently from such stimuli (cf. Eagly and Chaiken 1993, 5).

Attitudes versus ideologies Ideologies are systems of attitudes. An ideology such as “Conservatism” might include matching attitudes toward e.g. certain ethnic groups, political questions, different forms of dress, etc. Not every attitude is part of an ideology, and not every person needs to have an ideology (cf. McGuire 1986, 116f). Language attitudes may be part of an ideology, but do not need to be.

Attitudes versus schemas Eagly and Chaiken 1993 argue that attitudes may be regarded as subtypes of schemas (Eagly and Chaiken 1993, 18). Schemas are cognitive structures based on past experience as are attitudes. It is assumed that schemas influence behavior - the same assumption is held concerning attitudes. But schemas do not necessarily possess an evaluative dimension (cf. Eagly and Chaiken 1993, 18f), and they cover more phenomena than the term ‘attitude’ does: “The concepts differ because the term attitude refers to evaluation, whereas the term schema has been used more broadly” (Eagly and Chaiken 1993, 18).

2.2.4.4 Attitudes as hypothetical constructs

Attitudes are not directly observable. What is observable is that certain events (presentations of attitude objects; stimuli) co-vary with certain other events (evaluative responding). An assumption that cannot be verified through direct observation but that nonetheless has heuristic value is called a *hypothetical construct* (cf. Eagly and Chaiken 1993, 2). By stating that attitudes are hypothetical constructs, I do not commit myself to an analysis of attitudes either as latent processes, i.e. truly existing, or as mere conceptual conveniences (cf. Eagly and Chaiken 1993, 5). This question lies beyond the scope of this thesis.

2.2.5 Different measures of attitudes

an attitude is a complex affair which cannot be wholly described by any single numerical index. For the problem of measurement this statement is analogous to the observation that an ordinary table is a complex affair which cannot be wholly described by any single numerical index. (Thurstone 1967, 77)

When people speak about measuring attitudes, they usually refer to measuring attitude value, i.e. the degree of positiveness or negativeness of an attitude¹⁶. This is also the measurement I will focus on for the purpose of this study. There are other measurements though, which I will discuss here very briefly.

Attitude strength Strength of the link between the attitude and its object. High attitude strength also means high attitude accessibility (see below) (cf. Eagly and Chaiken 1993, 196). A person presented with a formerly unknown attitude object and asked about his/her attitude toward this object may develop an attitude ‘on the spot’. This attitude, though, will have very low attitude strength.

Attitude accessibility The ease with which attitudes can be activated. Attitude accessibility has been taken into account to explain why some attitudes seem to influence behavior stronger than other attitudes (cf. Eagly and Chaiken 1993, 195).

Accessibility can be viewed as the strength of the associative link between object and evaluation, such that for highly accessible attitudes, the evaluation of an object is automatically activated from memory when that object is encountered (...). Alternatively, accessibility could be conceptualized as represented in the connection weights within a connectionist model. In this model, accessibility would correspond to the ability of partial stimulus input to *quickly* and accurately produce the entire pattern of activation for the attitude (...). (Fabrigar et al. 2005, 81)

Attitude accessibility is measured through the speed with which a person can react evaluatively to an attitude (cf. Fabrigar et al. 2005, 81).

¹⁶Fabrigar, MacDonald and Wegener call this “valence and extremity” (Fabrigar et al. 2005, 79).

Attitude stability Some attitudes may be stable, others may only be passing. Often, attitudes with low attitude strength/accessibility will be less stable than those with high attitude strength/accessibility. All other factors equal, it is likely that the higher the attitude strength/accessibility, the less prone to attitude change this attitude will be (cf. Fabrigar et al. 2005, 108).

This selection of measurements for different aspects of attitudes is only a small sample of measurements used in different theories. Which measures are used, or even considered to be meaningful, depends very much on the theoretical orientation of a researcher. Attitude stability, for example, is a non-measure for scientists assuming that attitudes are episodic by nature, i.e. are not stored in memory, but assembled anew upon each contact with an attitude object (“attitudes-as-construction” model, cf. Kruglanski and Stroebe 2005, 324ff). What you measure and how you measure is always a reflection of your theoretical background.

Attitudes may also differ concerning their internal structure, which I will discuss in the following section.

2.2.6 The internal structure of attitudes

The assumed internal structure of attitudes has an effect on the method of attitude measurement chosen, as well as on the interpretation of any data derived. Originally, during the design of the studies discussed here, I followed the traditional model of the internal structure of attitudes, assuming that each individual’s attitudes are positioned at one specific “point” or along a limited “range of acceptance” on an attitude continuum conceptualized similar to a number ray. This model lies at the basis of all measuring techniques that aim at reaching a single “attitude score”. Although sociolinguistic attitude research tends to be atheoretical (cf. section 2.2.7), the choice of the instrument of measurement always implies certain theoretical assumptions, especially about the internal structure of attitudes. While the side study on German dialects using a Likert-like measuring technique (see section 4.3) did not, and could not, due to its low scale level, raise questions concerning the validity of this assumption, the main study (see Chapter 5), using interval scale level Thurstone scale data, did exactly that. Therefore, while assuming the traditional notion of attitude structure usually purported by sociolinguistic research on language attitudes during the design phase of this research, the results forced me to reconsider my assumption and to argue that the attitudes of U.S. teachers toward AAVE are characterized by ambivalence, i.e. by the endorsement of opposing evaluations at the same time (for a more

detailed discussion, see section 6.2).

2.2.7 Attitudes: the sociolinguistic perspective

Above, I have outlined the social psychological attitude model, and suggested in how far this model differs from the historical sociologist one. I have also explained the fact that sociolinguistics took up the concept of ‘attitude’ relatively late. In this section I will discuss the question to what extent the sociolinguistic usage of this term depends on that of social psychology, and whether there is a genuine sociolinguistic perspective on attitudes. For that purpose, I will first discuss the use of the term in empirically oriented research papers¹⁷, and then will continue with a delineation of main positions found in theoretically oriented papers by sociolinguists. As a last step, I will present and criticize the analysis by Ryan et al. 1982.

The most salient fact about the sociolinguistic usage of the attitude term in empirically oriented papers is that the majority of authors on this topic do not attempt a formal definition. A consensus on the exact meaning of this term seems to be assumed - an assumption that would prove fallacious within social psychology. Often, the method of attitude scaling employed might service as an indicator for the definition used (cf. 2.3). And one has to admit that a wordy discussion of theoretical standpoints might indeed be inappropriate for some ‘hands-on’ practice-oriented studies. In those papers the names of the developers of these scaling methods are systematically misspelled (“Lickert” vs. “Likert” scaling), or labels are used for the techniques employed that are outrageously farfetched (cf. what I call *pseudo-Likert* studies, see section 3.2), this credit might be undue. In some cases it might be assumed that the authors were not aware of the fact that attitudes are not an unproblematic term.

Those authors who do discuss the attitude term usually do it in one of two ways: (a) they refer to traditional social psychological usages of the term, or (b) they discuss the practical implications of attitudes within speech communities. The first type is rarely found in the literature on attitudes to AAVE. One noteworthy exception to this trend is Williams et al. 1976, who use this as an introduction to type (b) that they employ as well. Other authors that use type (a) are e.g. Robinson 1996. Type (b), even though it is not, strictly speaking, a definition, might be the ‘best bet’ for a genuine sociolinguistic

¹⁷For convenience I review those articles that will also be discussed in Chapter 3. All discuss teachers’ attitudes to AAVE, and might therefore not reflect the character of language attitude studies in general. Since a review of all language attitude studies by sociolinguists is not possible within the scope of this thesis, I must accept this potential distortion.

approach to defining language attitudes. Beside Williams et al. 1976, this type can be found with e.g. Cecil 1988 and Bowie and Bond 1994. For no apparent reason, at least among the selection of studies I used as a basis for this short assessment, Osgood-based studies seem to deal with this topic somewhat more thoroughly. Bowie and Bond 1994 is the only article in this list that does not employ Osgood-scaling.

A number of theoretical oriented papers on language attitudes written by linguists do exist. Among them are e.g. St. Clair 1982, Ryan et al. 1982, Garrett et al. 2003, and Cooper and Fishman 1974. Giles and Ryan 1982 might be added here, from a communication science perspective. A lot of this literature discusses traditional techniques of attitude measurement and similar ‘1930s’ questions (cf. 2.2.3). Few papers leave this span of topics, among them, for example, Cooper and Fishman 1974. I suggest that the questions discussed in the theoretically oriented literature on language attitudes roughly falls into one of two compartments:

(a) Social psychology revisited:

- Definition of language attitudes
- Attitude measurement (especially matched guise)

(b) A genuine sociolinguistic approach

- Differences between social psychological, communication science, sociological and sociolinguistic views
- The genesis of language attitudes
- Functions of language attitudes in speech communities
- Comparability between different language attitude studies
- Systematization of different types of language attitudes
- Critique of the ‘blind’ copying of social psychological methodology

This, of course, is a coarse simplification of research reality and can serve as little more than a basic orientation on the state of much sociolinguistic literature on that topic.

Ryan et al. 1982 offer an analysis of the differences between the social psychological/communications, the sociological and the sociolinguistic approach to language attitudes. According to this article, sociology views “the symbolic values of language [...] within societal and situational contexts” (Ryan et al. 1982, 2): language serves as an indicator of social statuses and personal

relationships and as marker of societal goals. The preferred measurement techniques are the questionnaire or interview method, but content analysis is also used (cf. Ryan et al. 1982, 2).

Social psychology/communication science puts “emphasis [...] upon the individual and his/her display of attitudes toward ingroup and outgroup members as elicited by language and as reflected in its use” (Ryan et al. 1982, 2). Matched guise (cf. 2.3.1.3.1) is considered its preferred technique (cf. Ryan et al. 1982, 2).

They consider two questions to be specific to the sociolinguistic perspective: “understanding the association between specific linguistic features [...] and characteristics of the societal, social group and situational contexts in which they occur” (Ryan et al. 1982, 2) and “understanding the inferences listeners make about these associations” (Ryan et al. 1982, 2). Their characterization of sociolinguistic research on language attitude is in harmony with the work of e.g. Cooper and Fishman 1974, but, from my perspective, much of sociolinguistic research is rather *atheoretical* than closely bound with these questions. None of the studies discussed here uses matched-guise, though verbal-guise, a closely related method (see footnote in section 2.3.1.3.1), has been used in both studies by Williams et al. (see section 3.1.1).

2.3 Measuring attitudes

In this chapter, I will delineate several methods for the measurement of attitudes in general and language attitudes in particular. I will first outline some basic theoretical problems of measuring psychological phenomena, then present collection of methods in form of a short overview, before discussing the methods most relevant for this study in more detail and giving arguments for my choice of method.

2.3.1 Theoretical problems in measuring attitudes

In this section, I would like to discuss some theoretical problems that attitude measurement - utilizing whichever method - has to face: the fact that attitudes are hypothetical constructs and therefore not directly observable, and the problem of social desirability. I will also discuss the different options concerning choice of stimulus here, offering a short introduction into the technique of matched guise and its specific needs concerning stimulus material. At the end of this section, I will explain the relevance of scale level for the development of attitude scales.

2.3.1.1 A hypothetical construct

Attitudes are hypothetical constructs (see also 2.2.4.4):

Like other hypothetical constructs, attitudes are not directly observable but can be inferred from observable responses. The relevant observations are responses that are elicited by (or occur in close conjunction with) certain stimuli. As a general strategy in psychology, when certain types of responses are elicited by certain classes of stimuli, psychologists infer that some mental state (e.g., mood, emotion, attitude) or disposition (e.g., personality trait) has been engaged. It is this state or disposition that is said to explain the covariation of stimuli and responses. (Eagly and Chaiken 1993, 2)

We cannot measure attitudes directly, we can only measure responses to stimuli and try to infer the intensity and orientation of attitudes from these. The form responses can take are as multifold as the techniques for measuring them: cognitive responses such as beliefs and opinions (expressed beliefs), affective responses like fear, dislike, aesthetic enjoyment, conative responses such as leaving the room, sharing a gift, or discriminating against. Responses can sometimes be observed directly, sometimes technical apparatus is needed (e.g. affective responses are reflected in physical reactions such as rising blood pressure), and often a questionnaire is the tool of choice. However, all of these indicators, whichever way they are measured, are merely indicators for the existence of attitudes. We have no way of measuring attitudes *directly*.

2.3.1.2 Artifacts

It goes without saying that the measurement of attitudes is always subject to a multitude of artifacts created by the act of measuring itself. Social desirability is only one of those, albeit a very important one. It skews the results away from the true attitudes of subjects toward what subjects consider the social or professional ‘standard’. This will especially be the case when you ask teachers about their behavior at school, a context in which teachers will be very conscious about what answers/behaviors will be expected from them by their school administration, their school board or the parents whose children their school serves. Guaranteeing anonymity can alleviate, but not solve, the problem.

If questionnaires are used, order effects contribute another major source of distortion. Careful arrangement of items may help to reduce the effect. If questionnaires are distributed or interviews are conducted face-to-face, the

presence of the experimenter will have a further distorting effect. An African American experimenter is likely to elicit different responses than a European American experimenter, for example. By conducting studies online, this effect cannot be totally avoided, because many characteristics of the experimenter beyond his/her physical appearance (e.g. the name) will be interpreted as cues for ethnicity.

2.3.1.3 The nature of the stimulus

When we measure attitudes, we measure attitudes toward attitude objects. If you want to study attitudes toward snakes, your attitude object is a snake. In your research, you can use either the term *snake*, present a picture of a snake, or use a living snake as stimulus. Even though the word *snake* ought to represent the object ‘snake’, measured attitudes differ based on the stimulus chosen - at least in this specific case, where the attitude object can evoke strong emotions, especially fear, that the word itself may not evoke (cf. Eagly and Chaiken 1993). Yet, presenting a particular snake creates the problem that this is just *one* snake out of a multitude of different types of snakes, and the results of attitude measurement might differ based on the species and particular features of the snake used as a stimulus. Replacing a class of entities by a particular entity might skew the results as well. In this case, our attitude object is a language variety. Since this is an abstract object, we do not have all the options we have with snakes as the attitude object: We cannot hold “an AAVE” into the air or put it onto a table. Our options are basically the following:

1. We can use the word *AAVE* (or an equivalent term) as a label.
2. We can use examples of AAVE: example sentences, voice recordings, etc.
3. We can use an abstract description of AAVE.
4. We can use some form of reference to a particular speaker or a class of speakers.

The first two types have been used repeatedly in the studies reviewed here. Type three has been used only in combination with stimulus of type one. The fourth type seems to be used less often; it does not appear in the selection of studies reviewed here, though it can be found in studies outside the scope of the literature review. Shuy 1970 used the term *disadvantaged* to refer

to students whose language teachers were asked to describe.¹⁸ Speicher and McMahon used “examples in popular culture” (Speicher and McMahon 1992, 389) if subjects did not understand the label *BEV* (i.e. the initialism for *Black English Vernacular*). Furthermore, it is likely that an abstract description of AAVE would include some form of reference to, for example, the ethnicity of many speakers of AAVE.

2.3.1.3.1 Matched guise In matched guise, a technique developed during the 1950s by Wallace E. Lambert to study language attitudes in the Montreal area (cf. Lambert 1967, 94), **one** speaker produces **two** sets of stimuli, e.g. a bidialectal speaker is recorded speaking (a) SAE and (b) AAVE¹⁹. The subjects are presented with both sets of stimuli, without being told that both recordings stem from the same speaker. When these two sets of stimuli yield different evaluative responses (typically, matched guise is used in combination with a semantic differential (cf. 2.3.2.3.1), although it can be used with other other techniques as well), it can be assumed that these differences were triggered not by differences between individual speakers (e.g. differences in voice quality, etc.), but by differences between the varieties used. Of course, even though the difference in attitudes measured was triggered by a difference in language, the measurement does not reflect attitudes toward that language variety, but merely attitudes toward *speakers* of that variety. That is, the fact that a subject is of the opinion that a speaker of AAVE is less “warm” than a speaker of “SAE” does not automatically guarantee that he/she also considers AAVE as a variety, independent of its speakers, to be less “warm” than SAE. Even though this conclusion seems plausible at first sight, it does not automatically follow from the data gathered (cf. Edwards 1982, 22). Nor does a negative evaluation of one speaker using a specific variety warrant that all speakers will be judged the same way (e.g. some form of speech may be considered to be more acceptable with male than with female, or with young rather than old (or vice versa) speakers). The advantages of such a method are obvious. An indirect approach can reduce the amount of distortion due to social desirability. The subject assumes that he/she is evaluating different speakers, and is not aware that the purpose of the study is to measure language attitudes.

No technique is without its drawbacks, though. Garrett et al. 2003 mention the following problems:

¹⁸Roger W. Shuy was interested in the abilities of teachers to describe the language of their students. He did not conduct, strictly speaking, attitude research.

¹⁹A variant of this method, called *verbal-guise*, exists. Here, different speakers are used for the different varieties (cf. Garrett et al. 2003, 53). This is sometimes unavoidable, when a bidialectal or bilingual speaker cannot be found.

1. the salience problem: “The routine of providing judges with the repeated message content of a reading passage presented by a long series of speakers may exaggerate the language contrast compared to what would otherwise be the case in ordinary discourse (...). [T](t)he MGT [matched guise technique, addition mine] may systematically make speech/language variation much more salient than it otherwise is” (Garrett et al. 2003, 58)
2. the perception problem: “One cannot be sure (...) how reliably judges have perceived the manipulated variables” (Garrett et al. 2003, 58)
3. the accent-authenticity problem: Some variables, e.g. in intonation or discourse patterning, might inadvertently be eliminated from the stimulus material. The material is not authentic (cf. Garrett et al. 2003, 59).
4. the mimicking-authenticity problem: Occasionally, one speaker is asked to produce a wide range of different guises. In such cases, it is unconvincing that the speaker is actually multi-dialectal to such a high degree, and more plausible that some of these guises are merely mimickings of the varieties. Even though a good imitation might elicit similar responses from the judges, this is a problem that should be taken into account (cf. Garrett et al. 2003, 59). Verbal guise might be a more appropriate method when bi- or multidialectal/lingual speakers cannot be found for a specific combination of languages or language varieties.
5. the community-authenticity problem: Within a variety, there may be further variation. Garrett et al e.g. criticize using labels as broad as *Welsh English* or *south-Welsh English* (the label *AAVE* is even broader than the examples Garrett et al mention), and recommend the use of “more specific or localized” (Garrett et al. 2003, 60) labels.
6. the style-authenticity problem: When speakers read out passages, the resulting style will differ from that of naturally occurring speech (cf. Garrett et al. 2003, 60). Either authentic spoken passages should be used, or the subjects must be told that what they are going to hear are read-aloud text passages.
The use of read passages for the study of attitudes toward AAVE is especially problematic since AAVE is normally less accepted in formal (e.g. at school) than in informal situations (cf. Buendgens 2002).
7. the neutrality problem: It is difficult to construct a neutral text. A sentence such as “I didn’t know what to think” might, with one type of

speakers (e.g. older speakers), be interpreted as sign of confusion, with another type of speakers (e.g. younger speakers) as a sign of careful evaluation of a complex problem - thus influencing speaker evaluations (cf. Garrett et al. 2003, 60f).

Most of these problems can be summarized as follows: It is difficult to find or produce ‘natural’ or seemingly natural stimulus that is, at the same time, neutral. Matched guise can be a valuable tool, but it has to be designed carefully. While verbal-guise avoids the mimicking-authenticity problem, all other problem apply to this method as well.

2.3.1.3.2 By any other name: The politics of choosing a label

When a variety is identified through a label for it, this label should be chosen very carefully, and the influence of label choice on attitude measurement should be considered.

In the different studies on language attitudes toward AAVE, a multitude of different labels have been utilized. Table 2.3 offers an overview over the different decisions of researchers concerning choice of label.

Generally, the choice of labels used in a questionnaire is not explicitly dis-

Label	Studies
<i>African American English</i>	Hoover et al. 1996 & Abdul-Hakim 2002
<i>African American English & AAE & Ebonics</i>	Blake and Cutler 2003
<i>Black English</i>	Di Giulio 1973
<i>Black English & nonstandard English & nonstandard dialects</i>	Taylor 1973 & Covington 1975 & Ford 1978 & Bowie and Bond 1994
none	Cecil 1988 & Williams et al. 1976 & Woodworth and Salzer 1971

Table 2.3: Labels used

cussed. I assume that in most cases researchers picked those terms they considered to be the best known ones, or at least the best known ones that they considered to be politically correct or otherwise acceptable for the purpose of an attitude study among teachers.

Speicher and McMahon 1992, who used a relatively free interview style for assessing attitudes,²⁰ discuss in detail the difficulties their subjects had with

²⁰Their study is not included in the Review of Literature section (cf. Chapter 3), since it does not deal with teachers/pre-service teachers and therefore falls outside the scope of

the offered label “BEV”:

Most of our informants said they knew what BEV referred to, a few thought they did but were unsure. Four did not know. From examples in popular culture, two of those eventually understood the referent, but the other two informants objected to the term for political reasons (see later discussion). However, as we began to explore what the informants thought BEV entailed, we realized that although we were using the same label, each informant’s understanding of the code to which the term referred differed dramatically.

When asked what label they would use to refer to BEV, nearly half of the informants said “slang.” Several more used terms such as “street talk”, “jive,” and “non-standard English.” The older male staff person called BEV “ghetto language” and characterized it as “idiotic.” On a more neutral note, the terms “Black English,” “Afro-American English,” and “Ebonics” were provided, along with our personal favorite: “Blanglish,” credited to a former college teacher of one of the male student informants. Three said they would not call it anything. A female student reported, “If I was sitting around with my friends that I grew up with, we wouldn’t call it anything. It was just the way we talked.... There was no label for it.” (Speicher and McMahon 1992, 389)

During a side study (cf. Buendgens-Kosten 2006, also briefly discussed in section 2.1.3) I collected data on the usage and recognition of terms by teachers. “Ebonics” was both used and recognized most frequently, but “Black English”, “African American English” and “African American Vernacular English” were also recognized often. They all might be considered potential labels for an attitude study aimed at teachers, possibly in combination with each other, e.g. as “Ebonics, also known as African American Vernacular English or Black English”²¹. Accompanying a label with a short explanation to clarify the term might also be an option (e.g. “The form of speech called ‘Ebonics’”), but this carries the risk of influencing the answer behavior of subjects, especially when the explanation has been worded carelessly (e.g. “Ebonics is a language used by...” followed by items asking whether the subject considers Ebonics/AAVE to be a language or not.). If it were plausible

the literature review.

²¹Blake and Cutler 2003 used this procedure: They used the term “AAE” in their questionnaire items, but preceded them with the following explanation: “African American English and AAE are used interchangeably with Ebonics” (Blake and Cutler 2003, 190), to ensure the term was correctly interpreted and/or more individuals understood their label.

to assume that there are teachers who are *absolutely* unacquainted with the idea that something like an African American sociolect (whether considered to be a language, a lect or merely ‘broken English’) exists, such definitions should not be used. In such a case it should be avoided that those individuals who cannot have attitudes toward AAVE since they are totally unacquainted with this attitude object even in its widest sense fill out the questionnaire. In studies working with matched guise or, more generally, using audio stimulus, the variety may not be named at all.

I will discuss my choice of terminology in detail in Chapter 5.

2.3.1.4 Scale level

The type of scaling used determines the scale level of the data gained by applying the scale. Four different levels of measurement (or scale levels) exist: nominal scales, ordinal scales, interval scales and ratio scales. Simply speaking, the higher the scale level, the more informative the data.

Nominal scales reflect equivalence and difference of data. If one person has a score of “1” and the other person has a score of “2”, you know that they differ in their scores. You do not know whose score is higher, or how far apart the two scores are (cf. Eagly and Chaiken 1993, 23f).

Ordinal scales have a fixed order. A score of “1” is less than a score of “2”, which, again, is less than a score of “3”. When using ordinal scale data, you can sensibly speak about a more or less, but you cannot speak about the degree of more or less. “1” might be only very little less than “2”, and “2” and “3” might be worlds apart. Only the order is known, not the distance (cf. Eagly and Chaiken 1993, 24).

With interval scales, on the other hand, you know not just the order of scale value, but also their unit of measurement, i.e. you know the exact distance between a score of “1” and one of “2” (cf. Eagly and Chaiken 1993, 25f).

The highest scale value is ratio scale value. A ratio scale has not only a fixed and known unit of measurement, but also a true zero point. The exact origin, the “0”, of the scale is known (cf. Eagly and Chaiken 1993, 25f).

The differences between these scales become highly relevant during statistical analysis. Nominal scales allow = and $\neq$; ordinal scales, additionally, < and >; interval scales, again additionally, + and -; and ratio scales, additionally, * and : as mathematical operators. If you have a scale that codes “male” as “1” and “female” as “2”, the arithmetic mean is just not as informative as, for example, when used on data sets detailing size (which is of a ratio scale level) (“The average group member has a sex of 1.8” vs. “The average group member has a size of 1.8 m”).

The question in how far it is appropriate to use methods of statistical anal-

ysis that depend on the use of e.g. division on ordinal scale level data, is controversial. As long as “an observed measured variable is a continuous ordinal variable that is a monotonically increasing function of an underlying variable (and the standard assumptions of homogeneity of variance and normality hold)” (Eagly and Chaiken 1993, 27) important statistical tests, such as the t -test, might be applicable.

During literature review, I will sometimes note that a statistical operation has been used with data of a scale level not strictly speaking compatible with this operation. This does not imply that the results are necessarily invalid, of course. During my own research, I avoided such problematic operations, either by using forms of analysis that are unproblematic for the scale level at hand, or by choosing a scale with an appropriately high scale level.

2.3.2 Short overview over attitude measuring techniques

In this section, I am going to discuss a number of different techniques for measuring attitudes. The list, although extensive, is not exhaustive.

I have arranged the different methods in a variant of Eagly and Chaiken 1993’s systematization. The first three categories differ in their relationship between scale development (or, in other words, scaling of the stimulus) and data collection. In category 1, the stimulus is scaled in a first step, and only in a second step is data gathered concerning the subjects’ attitudes. This means that two sets of subjects are needed: One group of judges for scaling the stimulus, and one group of subjects whose attitudes are scaled (the ‘subjects proper’) (cf. Eagly and Chaiken 1993, 30). In category 2, only one group of subjects and one set of data is needed: The same data that will be used to assess subjects’ attitudes also functions as material for scale construction (cf. Eagly and Chaiken 1993, 44). In category 3, no empirical data is used to determine scale distances, etc. The subjects’ responses only serve to assess their attitudes, not for creating the scale (cf. Eagly and Chaiken 1993, 50).

In most cases, these three categories are realized as ‘paper-and-pencil’ methods, even though they could, of course, be adapted to other media, such as speaking and signing, without altering their basic concepts. Some of them carry restrictions as to which classes of evaluative responses (i.e. cognitive, affective and conative responses) are used as indicators, but most can (and often are) adapted to serve other classes as well. To these I add a fourth category, which comprises those techniques that are limited to only one class of evaluative responding, i.e. cognitive, conative or affective responding.

2.3.2.1 Stimulus-then-person-scaling

Since stimulus-then-person-scaling demands two sets of subjects, one group of judges for scale creation and one group of individuals whose attitudes are assessed, it might seem to be rather laborious. They do often yield high scale levels, though, and are therefore of special interest.

2.3.2.1.1 Thurstone Strictly speaking, there is no ‘Thurstone-scaling’, but ‘Thurstone-scalings’. Louis L. Thurstone has developed three methods for scale construction, the method of equal-appearing intervals, the method of successive intervals and the method of paired comparisons, which I will discuss in this section.

Thurstone entered the field of attitude measurement with a strong background in psychophysical measurement. In psychophysical measurement, individuals are asked to judge characteristics which can also be measured using traditional physical measurements, e.g. weight, brightness, shades of grey, etc. From experiments of this type, two scales can be derived: the so-called “*R*-scale”, the scale of the actual physical stimulus with the unit gram, centimeter, etc., and the “*S*-scale”, reflecting the psychological continuum, with the unit ‘equally often noticed stimulus difference’ (cf. Thurstone and Chave 1929, 2). Attitude measurement differs from psychophysical measurement only in so far as there is no *R*-scale.

2.3.2.1.1.1 Method of equal-appearing intervals In the method of equal-appearing intervals, a large number of judges are asked to sort items into a certain number of piles (e.g. 11), so that these piles “represent an evenly graduated series of attitudes from those extremely against (...) to those which are very much in favor (...)” (Thurstone and Chave 1929, 30). From this data, scale values for each item can be calculated.

2.3.2.1.1.2 Method of paired comparisons When using this method, judges are not given the complete list of items and asked to sort them into piles, but are presented with only one pair of items at a time and are asked to indicate which of these is more positive. This process is repeated until every item has been judged against every other item, resulting in a total of $\frac{n(n-1)}{2}$ pairs. While the method of equal-appearing intervals is feasible with rather large numbers of items (Thurstone reports using a total of 130 items for his scale on attitudes toward the church (cf. Thurstone and Chave 1929, 30)), the method of paired comparisons is very much limited in that regard (cf. Eagly and Chaiken 1993, 39). Since the number of pairs needed rises exponentially with each additional item, item pools with more than twenty

items are not feasible. While ten items need only 45 pairs, 15 need 105, 20 need 190, and 25 items need 300 pairs.

2.3.2.1.1.3 Method of successive intervals The method of successive intervals employs the same sorting technique as the method of equal-appearing intervals, but uses a different method for calculating the scale values for each item. It was developed in order to combine the advantages of the other two methods: the more precise results for scale values of the method of paired comparisons and the bigger ease of judgment collection of the method of equal-appearing intervals.

2.3.2.1.1.4 Problems One problem of Thurstone scaling has not been addressed yet: the anchor effect. When a subject is told to estimate the physical weight of objects, and given one object as a ‘standard’ or ‘anchor’ to compare those objects with, the anchor causes distortions in the estimation of weight:

(...) an anchor located at either end of the to-be-judged stimuli (or just beyond them) tends to produce *assimilation* – the judged stimuli are perceived to be closer or more similar to the anchor than they actually are. However, an anchor located considerably beyond the range of stimuli to be judged tends to produce the opposite effect, *contrast* – the judged stimuli are perceived to be more discrepant or more dissimilar to the anchor than they actually are. (Eagly and Chaiken 1993, 366)

M. Sherif et al assume that the attitudes of a judge can function as an anchor when this judge evaluates the positiveness or negativeness of an item (cf. Sherif and Hovland 1961, 17ff.). This could cause distortions during the scaling process. During Thurstone scale construction, the exact ‘value’ of positiveness or negativeness of items is of no relevance, but merely the *compared* positiveness or negativeness of one item compared with another item. If a distortion occurs (may it be assimilation or contrast), the perceived positiveness and negativeness is altered, but the contrast between item one (slightly more positive) and item two (slightly less positive) should – even though the ‘extremeness’ of distance might be altered – still be perceivable for judges. Therefore, a method asking judges to allot values to individual items (among them, probably, Thurstone’s method of equal-appearing intervals and method of successive intervals) is highly susceptible to distortion by the anchor effect, while Thurstone’s method of paired comparisons should suffer little or not at all of this effect.

2.3.2.1.2 Magnitude estimation (Stevens) S.S. Stevens' magnitude estimation "has received only limited attention from social scientists" (Eagly and Chaiken 1993, 43), and none in the context of language attitudes, but would, due to its potentially ratio scale level, deserve some more attention than it has.

An item, in this context called *modulus*, is presented to a judge together with a value of e.g. 100 (the actual value is entirely arbitrary). The judge is then presented with further moduli and asked to indicate the ratio between this and the first modulus by a number. If he/she for example had "Vanilla ice cream is tasty" as 'standard' modulus and "Vanilla ice cream is the best thing in the entire universe" as a second modulus, and he/she perceived the second item to be exactly twice as favorable as the first, he/she would assign a "200" (cf. Eagly and Chaiken 1993, 43).

This process is repeated for all moduli and with several judges, and then mean values for all moduli are calculated. These arithmetic means constitute the item values (cf. Eagly and Chaiken 1993, 43).

Variants of this procedure exist, e.g. asking the judges not to give their estimates in numerical values but for example by adapting a light source (twice as bright equals twice as favorable), which allows for cross-validation across different modalities.

This method of scaling has two major drawbacks: Firstly, it has not been settled whether the resulting scale is indeed a ratio scale, as Stevens claimed, or 'merely' an interval scale. Secondly, the research concentrated on the creation of scales, neglecting the actual use of these scales, resulting in a lack of actual experience with the application of this method (cf. Eagly and Chaiken 1993, 44).

2.3.2.2 Simultaneous-stimulus-and-person-scaling

In simultaneous-stimulus-and-person-scaling, the need for a second group of subjects that serve as judges of item values is eliminated, simplifying the data-collection part of scale creation.

2.3.2.2.1 Guttman Louis Guttman's scaling method (also known as *interlocking technique* or *scalogram analysis*) uses only one group of judges/subjects. The data used to create the scale is at the same time the data being analyzed by the newly created scale. How is this achieved?

The first step consists of the selection of scalable items. In order to be scalable, these items need to be cumulative: A person agreeing with a high-rating item must also be in agreement with any lower-rating item (cf. Eagly and Chaiken 1993, 47; Guttman 1944, 100ff). The items might be chosen intu-

itively, but only if this condition is fulfilled (which cannot be verified before scaling is attempted), will the construction of a scalogram (see below) be possible.

In a second step, data is elicited from judges/subjects. Guttman stresses that this may be done in any convenient modality (cf. Guttman 1944, 98), including, of course, questionnaires. Often, a questionnaire design similar to Likert-scaling (see also section 2.3.2.3.1) is used, though the number of reply options are normally collapsed into only one or two options (e.g. agreement/disagreement) during the scalogram creation process.

In a third step, the construction of a scalogram will be attempted. A scalogram is simply a visual arrangement of data, in which items and subjects are arranged in the following fashion: Those individuals who agree to the highest number of items are moved to the top of the chart, those with the lowest number to the bottom²². Then, the order of items is changed so that those items with which the highest number of subjects agree are moved to the left side, those with the lowest rate of agreement to the right²³(see Table 2.4)(cf. Guttman 1944, 99ff; Eagly and Chaiken 1993, 44ff).

If the items are indeed cumulative, the resulting scalogram will show a

	Item b	Item a	Item d	Item c
Person 1	agreement	agreement	agreement	agreement
Person 3	agreement	agreement	agreement	disagreement
Person 4	agreement	agreement	disagreement	disagreement
Person 2	agreement	disagreement	disagreement	disagreement
Person 5	disagreement	disagreement	disagreement	disagreement

Table 2.4: Scalogram

triangle-form. In many cases, though, several revisions of data are necessary to reach this form, e.g. exclusion of items from the scale. In some instances, even this will not suffice: Not every set of data proves to be scalable by this method (cf. Guttman 1944, 103). If a scale is achieved, this will be of ordinal scale level.

2.3.2.2.2 Edwards-Kilpatrick scale Allen L. Edwards and Franklin P. Kilpatrick have suggested techniques by which one of the main problems of Guttman-scales could be resolved: How can scalability be improved through

²²Individuals with identical answering behavior can be collapsed into one row.

²³The orientation of the arrangement is arbitrary and can be changed at will.

item choice? Edwards and Kilpatrick looked at Thurstone- and Likert-scaling and suggested borrowing ideas from them, especially from Thurstone-scaling:

One of the most promising [approaches, addition mine] with which we have been working involves first scaling a large number of items by the method of equal-appearing intervals. On the basis of scale and Q values, we reduce the initial set of items to a smaller number. These items are then given to another sample and subjected to item analysis. We then plot the discriminatory power of the items against the Thurstone scale values and select from within each scale interval the two or three items with the greatest discriminatory power. (Edwards and Kilpatrick 1948,113)

From this point on, scaling is continued as outlined in the preceding section.

2.3.2.3 Person scaling

Person scaling is the attitude-measurement-equivalent of a blockbuster producing company: Two types of person scaling, Likert- and Osgood-scaling, are ‘stars’ in the field of sociolinguistic attitude scaling. Person scaling makes scale creation relatively easy, but easy scale creation is paid for with lower or uncertain scale levels.

2.3.2.3.1 Likert scaling(s): Method of summated ratings The term *Likert-scale* can be used to refer to two different methods: The Sigma method and the Simplifier method. Only the Simplifier method constitutes an example of Person scaling, the Sigma method is an example of simultaneous-stimulus-and-person-scaling. Sometimes, scales that do not follow Likert’s method but have a superficial similarity to his instruments are referred to as *Likert scale*, *Likert-type scale* or, in this thesis, *pseudo-Likert scale*.

Since all three, Sigma method, Simplifier method, and pseudo-Likert, are routinely referred to by the same term, and the Simplifier method is probably the most popular one among them, I will discuss them together under this heading.

2.3.2.3.1.1 “Likert-type scale” Occasionally authors use terms like *Likert-scale* or *Likert-type scale* to refer exclusively to the outer appearance of a scale, i.e. the presentation of items. This is problematic, since Rensis Likert used different forms of item presentation:

1. Yes or no questions,

Would most negroes, if not held in their place, become officious, overbearing, and disagreeable?

YES – NO

(2) (3) (4) (Likert 1932, 18)²⁴

2. five-option multiple choice questions,

How far in our educational system (aside from trade education) should the most intelligent negroes be allowed to go?

(a) Grade school. (1)

(b) Junior high school. (2)

(c) High school. (3)

(d) College. (4)

(e) Graduate and professional school. (5)

(Likert 1932, 18)²⁵

3. standardized five-option (strongly approve, approve, undecided, disapprove, strongly disapprove) questions,

If the same preparation is required, the negro teacher should receive the same salary as the white.

Strongly Approve Approve Undecided Disapprove Strongly Disapprove

(5) (4) (3) (2) (1) (Likert 1932, 19)

4. narratives and description of outcome of the narrated situation, followed by the same response options as (3)

In a community of 1,000 whites and 50 negroes, a drunken negro shoots and kills an officer who is trying to arrest him. THE WHITE POPULATION IMMEDIATELY DRIVE ALL THE NEGROES OUT OF TOWN,

Strongly Approve Approve Undecided Disapprove Strongly Disapprove

(5) (4) (3) (2) (1) (Likert 1932, 19)(cf. Likert 1932, 14f)

Terms like *Likert-type scale*, though, are used to refer exclusively to items of the (3) type, often with a five-point gradation, but also of seven- or three-point gradations. Theoretically, even numbers of gradations are also possible, but they are rarely used.

Concerning content and wording of items, Likert introduced a number of

²⁴All examples are from Likert's "Negro Scale".

²⁵This item is similar in its structure to Bogardus' scale.

rules. These rules are, more often than not, ignored in the construction of such ‘Likert-type scales’, but also, at least occasionally, in the construction of ‘real’ Likert scales. These rules for item choice are as follows:

- “It is essential that all statements be expressions of *desired behavior* and not statements of *fact*.” (Likert 1932, 44)

A statement of fact can be true or false. An expression of desired behavior is an opinion and therefore beyond correctness or falsity. Wordings containing *should* are recommended (cf. Likert 1932, 45).²⁶

- “The second criterion is the necessity of stating each proposition in *clear, concise, straight-forward statements*.” (Likert 1932, 45); “Double-barreled statements [e.g. “AAVE is great and everybody should learn AAVE at school.” - addition mine] are most confusing and should always be broken in two [e.g. “AAVE is great” and “Everybody should learn it at school.” - addition mine].” (Likert 1932, 45); “each statement *must avoid every kind of ambiguity*” (Likert 1932, 45).

²⁶It might be difficult to cover all three dimensions of evaluative responding using this form. Most items in Likert’s “Negro scale” contain evaluations of public policy, legislation or group behavior, such as “Practically all American hotels should refuse to admit negroes.” (Likert 1932, 19). These items cover the cognitive dimension. Several items ask for statements of personal behavior, e.g. “Would you shake hands with a negro?” (Likert 1932, 18), covering the conative dimension. One item is of an affective nature: “If you heard of a negro who had bought a home or a farm would you be glad?” (Likert 1932, 18). Theoretically, all three dimensions are therefore covered. Yet, the fact that the “Negro scale” contains only one isolated item representing the affective dimension, and neither the “Imperialism scale” (Likert 1932, 19f) nor the “Internationalism scale” (Likert 1932, 15ff) contain anything similar, indicate that this form of evaluative responding might not easily be integrated into Likert-scales. This, of course, is not a prerequisite of attitude scaling, since each indicator class can be used both independently and in combination. The main problem one has when wording affective dimension items for strict Likert-scales is that one tends to ask for *actually expected* and not for *desired* behavior. The one example contained in Likert’s “Negro scale” actually does so. Many conative dimension items have the same basic problem the affective dimension items have: The fact that I would or would not “shake hands with a negro” (Likert 1932, 18) does not imply that I evaluate this as “desired behavior”. It might be a *fact* that I usually do or do not do so, but that I am actually ashamed of my behavior and consider it to be highly undesirable. As soon as a person feels slightly ambivalent about an issue, i.e. there is a contrast between e.g. how this person acts/reacts and how he/she thinks he/she should act/react, this may pose a major problem.

When constructing a Likert scale, one has to decide between concentrating on the cognitive dimension of evaluative responding, or choosing some items that do not perfectly fulfill Likert’s conditions, but are still in keeping with his examples.

These restrictions are rather obvious. Only if a subject understands items perfectly can he/she mark his/her true opinion concerning these items.

- “In general it would seem desirable to have each statement so worded that the modal reaction to it is approximately in the middle of the possible responses.” (Likert 1932, 45)

I assume that “modal” in this context refers to the (statistical) mode, the most frequently chosen value. If this is correct, then version I demands that a questionnaire is built in such a way that most respondents would chose the “Undecided” option or an equivalent (‘score-of-three’) option. In a society with the extremest imaginable dislike of AAVE this would mean opting for items such as “AAVE should be branded as the single worst plague of modern society.”, while a society with the deepest affection for AAVE would require items like this: “We should invest all available money in the teaching and conservation of AAVE.” The problem here, of course, is, that it is not always possible to foresee the response behavior of one’s population.

In an otherwise identical reprint of this chapter (“The method of constructing an attitude scale”, Likert 1932, 44-53) 1967, this paragraph has been changed:

In general it would seem desirable to have the questions so worded that the modal reaction to some is more toward one end of the attitude continuum and to others more in the middle or toward the other end. (...) There is no need, however, to have questions whose modal reactions are at either extreme of the continuum. (Likert 1967, 91)

The second version is not as easy to comprehend. If “attitude continuum” is meant to refer to the traditional concept of an “attitude line” covering possible attitude positions, then it is nonsensical: It would demand that items are chosen in such a way that, e.g., for item 1, most subjects will express a negative attitude toward AAVE, for item 2 a very positive attitude, for item three a neutral one, etc. If the term “attitude continuum” refers to the response options, this would be a very unusual usage of the term, but would make more sense (e.g. as a measure to avoid order effects, that has, of course, to be taken into account when scoring the responses). It would, nonetheless, be a perfect contradiction of version I, which I cannot explain.²⁷

²⁷Other differences between the two versions are less perplexing. Two sentences were deleted; both contained references to other chapters in Likert 1932. Other changes were the relabeling of one table, and a reaction to a minor typographic change (The overview in Likert 1932, 47 was turned into a table in Likert 1967, the introductory sentence was adapted appropriately). None of these changes would point toward careless editing.

- “To avoid any space error or any tendency to a stereotyped response it seems desirable to have the different statements so worded that about one-half of them have one end of the attitude continuum corresponding to the *left* or *upper* part of the reaction alternatives and the other half have the same end of the attitude continuum corresponding to the *right* or *lower* part of the reaction alternatives. (...) These two kinds of statements ought to be distributed throughout the attitude test in a chance or haphazard manner.” (Likert 1932, 46)

This rule, aiming at avoiding order effects and similar distortion, is rather straightforward and requires no further comment.

- “If multiple choice statements are used, the different alternatives should involve only a single attitude variable and not several.” (p. 91)

This means, that multiple choice statements of this type are excluded:

The use of AAVE should be:

- (a) tolerated at schools
- (b) encouraged in literature
- (c) be punished by fine
- (d) reprimanded

In an example like this, different aspects are mixed. This item should be split into two new items, e.g.:

The use of AAVE is acceptable in the following situations:

- (a) all situations
- (b) most situations
- (c) most informal situations
- (d) a few informal situations
- (e) never

and:

The use of AAVE should be:

- (a) accepted
- (b) reprimanded by words
- (c) reprimanded by words and punished by fine

In many cases it might be easier to avoid multiple-choice formats altogether.

2.3.2.3.1.2 Sigma method Both the Sigma method and the Simpler method use the same types of items and item presentations – they differ merely in their scoring procedures.

The Sigma method is the more complex among the two scoring methods. It is based on the observation, “that a great number of the five-point-statements (...) yielded a distribution resembling a normal distribution” (Likert 1932, 21). Assuming that items form normal distributions, you can calculate sigma values for each of the answer options. The defined zero point lies at the mean; negative and positive sigma values exist. The theoretical maximum for sigma values is +3 and -3. Likert also suggests, as an alternative, placing the zero point at sigma = -3 in order to avoid using negative values. Then the extrema would lie at 0 and +6 (cf. Likert 1932, 22). The different scores across items would be combined using a median or a mean (Likert 1932, 24).

Reliability tests of one instrument derived by this method gave satisfactory results.

This method is not unproblematic, though. The basic assumption, that of a normal distribution, is one weak point (one that Likert shares with Thurstone). While many attitude objects may create attitudes in a population in the form of a normal distribution, this may not be the case for highly controversial, i.e. polarizing, topics. Furthermore, it is problematic to compare different groups (or one group over time) using this method, since another mean will automatically bring with it a different scoring.

2.3.2.3.1.3 Simpler method While the Sigma method already constitutes a simplification of Thurstone’s method, the Simpler method constitutes as simplification of the Sigma method:

The simpler technique involved the assigning of values of from 1 to 5 to each of the five different positions on the five-point statements. The ONE end was always assigned to the negative end of the sigma scale, and the FIVE end to the positive end of the sigma scale. (...)

After assigning in this manner the numerical values to the possible responses, the score for each individual was determined by finding the average of the numerical values of the positions that he checked. Actually, since the number of statements was the same for all individuals, the sum of the numerical score rather than the mean was used. (Likert 1932, 25f)

I refer to this method of assigning number values to the different reply options as “standard scoring”.

Likert compared the results of the Simpler method with those of the Sigma method for two different scales and found high correlations (cf. Likert 1932, 26). If both methods arrive at the same results, the Simpler method can be considered advantageous because of the reduced effort needed for its construction.

2.3.2.3.1.4 Summary Likert summarizes the advantages of his method the following way:

first, the method does away with the use of raters or judges and the errors arising therefrom; second, it is less laborious to construct an attitude scale by this method; and third, the method yields the same reliability with fewer items. (Likert 1932, 42)²⁸

I do not consider it to be unproblematic, though. Unlike in other methods, in Likert scaling “each statement becomes a scale in itself and a person’s reaction to each statement is given a score. These scores are then combined using a median or mean.” (Likert 1932, 24). This method *can* only work if the ‘neutral’ option for each single items lies at approximately the same point in the attitude continuum, and the “(dis-)agreement” options have comparable distances from the ‘neutral’ option. If you have two items of this type: “The government should encourage education.” and “Education should be available for those who want it.” the neutral points might lie near each other. If the items used are: “Writing a factual text about illegal drugs should be punishable by capital punishment.” and “It should not be made too easy to get access to illegal drugs.” it is obvious that the neutral points inhabit entirely different spots on the attitude continuum. This example is of course overly extreme, but the same effect takes place in more moderately chosen items. Item selection has to be conducted *very* carefully to avoid any major distortion, and the choice of potential items is thus very much limited. The advantages and disadvantages of Likert-scales will be discussed in more detail in section 2.3.3.

2.3.2.3.2 Osgood: Semantic differential Charles E. Osgood, together with George J. Suci and Percy H. Tannenbaum, developed a method of measuring connotative dimensions of concepts in a way that enables direct intercultural comparison (Eagly and Chaiken 1993, 56; see for example Osgood et al. 1975). The first step in developing a semantic differential consists of

²⁸In this passage, Likert refers to the Sigma method, but his arguments apply to the Simpler method as well.

presenting the concept²⁹ to a number of subjects in order to elicit adjectives that describe this concept. Osgood often worked with long lists of concepts, eliciting only a single adjective per person per concept (cf. Osgood 1965, 111). From the resulting list of adjectives a first selection is derived using the following criteria:

- total frequency of usage
- diversity of usage across different concepts (if a scale for different concepts and not only one single concept is to be created)
- independence of usage with respect to each other³⁰(Osgood 1965, 111)

Opposites are elicited from subjects (Osgood does not mention whether the same or a new group of subjects is to be used for this step) for those adjectives that survive this process. Those adjectives for which no appropriate opposites can be found are discarded (cf. Osgood 1965, 111). The other adjective pairs are then turned into scales, on which subjects can position themselves: rough __ : __ : __ : __ : __ : __ smooth (cf. Osgood et al. 1957, 81). A person would tick the field on the very left to indicate that he/she considered the attitude object to be very rough, the second field on the left if he/she considered to be quiet rough, etc. Each person would be presented with a number of such scales for each attitude object of interest.

In Osgood 1965, Osgood describes the development of an instrument consisting of 50 adjective pairs. The semantic differential scales used to assess attitudes are generally much shorter. According to Eagly and Chaiken 1993, Heise (1970) has demonstrated that scales with 4-5 pairs should be sufficiently reliable (cf. Eagly and Chaiken 1993, 56). The two Osgood scales reviewed in this thesis use 20 respectively 15 pairs³¹.

Strictly speaking, the instrument developed by this method does not measure attitudes per se, but rather affective meaning, which does include, among others, a measure for attitude. Osgood reports that, independent of what concepts are featured and what language or culture a study is conducted in, factor analysis yields at least three major factors, usually labeled “evaluative

²⁹Osgood does not employ the term ‘attitude object’ since he is not, in the first line, interested in measuring attitudes. He suggests, instead, the usage of ‘object of affect’ (cf. Osgood et al. 1975, 65).

³⁰This point is not described in detail, but I assume what is meant is that if several adjectives with very closely related meanings appear in this list, all but one are to be excluded.

³¹Cecil 1988 presents a three-item questionnaire, but I will argue in section 3.1.2 that this questionnaire does not constitute an Osgood scale.

factor” (or E factor), “potency factor” and “activity factor” (cf. e.g. Osgood 1965, 110). The E factor is the one representing the measure of attitude:

Despite the plethora of definitions of “attitude,” there seems to be a general consensus that (a) it is learned and implicit, (b) it may be evoked by either perceptual or linguistic signs, (c) it is a predisposition to respond evaluatively to such signs, and (d) this predisposition may rate anywhere along a scale from “extremely favorable” to “extremely unfavorable.” Without a single exception, the first indigenous factor in our analyses turns out to be Evaluation - and, being both bipolar and graded in intensity in both directions from a neutral point, our E meets the criteria for a measure of “attitude” expressed above. Although the factor itself is presumably innate and universal, *what* things in the human environment acquire + or - values on E is a function (in representational mediation theory) of learning the meanings of signs and hence is highly susceptible to cultural influences. We therefore define “attitude” as the projection of a concept onto the E factor, indigenous or pancultural as the case might be. (Osgood et al. 1975, 237)

In the two studies reviewed in this paper that employ Osgood’s technique, this distinction between semantic differential and E factor is not upheld. Instead, the semantic differential as a whole is treated as the measurement of attitudes (cf. Section 3.1.1).

One great advantage of the semantic differential is the supposed transferability of one scale to a multitude of attitude objects/object of affect, so that a direct comparison between the attitudes/affects toward e.g. ‘love’ and ‘winter’, or ‘AAVE’ and ‘success’ could be drawn. The two ‘true’ Osgood scales applied to the measurement of teachers’ attitudes to AAVE reviewed here (cf. Section 3.1.1) have been designed specifically for this topic and target group, and are adapted so well to the specific field of language attitudes/language as object of affect, that they cannot be used outside this scope and therefore do not allow comparison with most other attitude objects/objects of affect or across different populations.

Within attitude measurement, it is popular to combine Osgood scales with the technique of matched guise. It is important to note, though, that these techniques do not have to be used in combination. Osgood himself preferred presenting mere labels for rating, rather than, for example, language samples or objects.

An open problem regarding the evaluation of Osgood scales are their scale level:

[...] its representational measurement properties are essentially unknown. Consequently, it is difficult to know what level of measurement is obtained or what properties the obtained attitude scores have. (Eagly and Chaiken 1993, 57)

2.3.2.4 Attitude measures linked to specific indicator classes

The following methods specialize in only one of three classes of evaluative responding: cognitive, affective or behavioral/conative responding. While most ‘traditional’ measurement techniques can be adapted for use with any of these three classes, these techniques are strictly limited to one class (cf. Eagly and Chaiken 1993, 57). Even though language attitude research usually employs the more ‘traditional’ techniques, some of the following might prove fruitful additions to more customary measurement techniques, e.g. for the purpose of cross-validation.

2.3.2.4.1 Using cognitive indicators If attitudes do influence the way we think, the distortions in cognitive processing caused by attitudes might be used as indicators for the presence of positive or negative attitudes. Several methods that utilize this principle have been developed. The most relevant ones are:

- **Hammond: Error-choice method.** Subjects are asked to answer questions by choosing one of two answer options. Both answers are incorrect, but distorted in different directions. Individuals are assumed to pick that option which is distorted in accordance to the attitude of the subject.

Example: “The average weekly wage of the war worker in 1945 was: a. \$37 b. \$57”, the correct answer being \$47, the question being presented to businessmen and labor union workers, used as an indicator for attitudes toward labor and management (cf. Eagly and Chaiken 1993, 58).

- **Thistlethwaite: Distortions in logical reasoning.** Subjects were asked to judge the logical correctness of conclusions. Both neutral and emotional statements were used. A high number of judgment errors, i.e. illogical conclusions being judged to be logical, in one specific type of statements is considered to be indicative of an attitude reflected by those statements (cf. Eagly and Chaiken 1993, 58).

- Cook: **Judgment of the plausibility of arguments.** Subjects are asked to rate the effectiveness of arguments. Those arguments that reflect the subjects' attitudes are rated as more effective than those that do not (cf. Eagly and Chaiken 1993, 58f).
- Sherif & Sherif: **Contrast effect.** When subjects are asked to sort statements according to favorableness, they usually judge an item to be more unfavorable if it is contrary to their own attitudes. This effect is not only interesting for use as measurement of attitudes, it is also relevant in the context of Thurstone scaling (potentially distorting judges' rating behavior)(cf. Eagly and Chaiken 1993, 59, see also Section 2.3.3).
- **Learning and retention of arguments**
 - Greenwald (amongst others): **Thought-listing.** Individuals are asked to list all their thoughts concerning an attitude object, e.g. concerning positive and negative features of the object, how one's values pertain to it, etc. In a second step, subjects judge the degree of favorableness of the thoughts they have listed. This is recommended as a measure of opinion, but may also serve as a rough indicator for attitudes (cf. Greenwald 1968, 156ff).
 - Fishbein: **Expectancy-value.** An individual holds a number of beliefs about an attitude object. Each of these beliefs has an evaluative aspect. Therefore, the attitude toward the attitude object can be calculated when (a) the beliefs and (b) the evaluative aspects of those beliefs are known: " $\sum_{i=1}^N B_i a_i$, where B_i = belief 'i' about the object, a_i = the evaluative aspect of B_i , and N = the number of beliefs" (Fishbein 1963, 233). The beliefs associated with an attitude object are collected by presenting subjects with the word representing that attitude object and asking for characteristics of this attitude object (cf. Fishbein 1963, 234f). Fishbein suggests using the semantic differential to determine the evaluative aspects of these beliefs (cf. Fishbein 1963, 235f).

Cognitive indicators can be employed in situations in which more direct approaches would yield no correct estimations of attitudes because, for example, the topic is considered taboo, or legal constraints make it difficult to elicit honest answers. It might be fruitful to combine these measurements with one of the more 'traditional' approaches in contexts where a high effect of social desirability is to be expected and it is desired to estimate the distortion between 'true' attitudes and expressed attitudes of a specific group.

2.3.2.4.2 Using affective indicators Since attitudes can trigger affective reactions, and affective reactions can be measured physiologically, such methods as assessing galvanic skin response, pupillary response or facial electromyographic activity can hint at underlying attitudes.

Galvanic skin response measures skin resistance, which is related to the amount of sweat produced. The more a person perspires, the lower his or her skin resistance is. Since perspiration can be used as an indicator of emotional reaction, the changes in galvanic skin response may indicate that an emotional reaction has taken place. Unfortunately, only the intensity and not the direction of an emotional reaction can be assessed. Therefore, a person showing strong galvanic skin response after being presented with a snake can dislike and fear snakes – or be a breeder of snakes who is pleased at seeing a rare and beautiful snake (cf. Eagly and Chaiken 1993, 60f).

Another problem, also shared by measurements via pupillary response, is the orienting reflex: Even in emotionally neutral situations, subjects will show responses of sweat glands and pupils when presented with surprising or unexpected stimulus, etc. (cf. Eagly and Chaiken 1993, 61). This needs to be taken into consideration when analyzing the data derived through this method.

Different ‘intensities’ of stimulus presentation have been experimented with. Subjects have been touched by members of different ethnicities, have interacted with members of different ethnicities, have been presented with images showing interactions between members of different ethnicities or with verbal labels for these groups. An effect was observed in all of these cases (cf. Eagly and Chaiken 1993, 61). It might therefore be possible to apply this technique to the study of language attitudes.

Unlike galvanic skin response, pupillary response is bidirectional - pupils can constrict *and* dilate. Generally speaking, pupil dilation is associated with positive affective responses and pupil constriction with negative affective responses. There are exceptions, though, which make the interpretation of results more complex. For example, pupil dilation can be observed at the initial presentation of negative stimuli, only turning into constriction on repetition of the stimulus. The orienting reflex, of course, also applies to pupillary response (cf. Eagly and Chaiken 1993, 61, Goldwater 1972).

Goldwater 1972 (343f) recommends that pupillary response not be used in combination with visual material, since physical characteristics of visual stimulus can influence pupillary response as well. This would not pose a problem in measuring attitudes toward languages and language varieties.

Measuring facial electromyographic activity means registering activities in

those muscles that are responsible for facial expression. A person experiencing a happy affective reaction might not necessarily smile, but will show some degree of activity in the muscles needed for smiling. Positive affective reactions cause reactions in the zygomatic muscles (smiling movement) and negative affective reactions cause reactions in the corrugator muscles (frowning movement) (cf. Eagly and Chaiken 1993, 62).

To my knowledge, none of these three techniques has ever been applied to the measurement of language attitudes. They might be interesting as an addition to a more conventional assessment of attitudes, e.g. for checking validity, especially in such contexts where the probability of receiving only socially desirable replies is extremely high. If the attitude questionnaire of a person indicates a neutral position but galvanic skin response, pupillary response or facial electromyographic activity show extreme reactions, then the attitude questionnaire will probably not reflect this person's attitudes. The disadvantage of these techniques is obvious. They all depend on some form of technology to record galvanic skin response, pupillary response and electromyographic activity, and on at least basic medical knowledge to interpret these responses correctly.

2.3.2.4.2.1 Self-Report of affect Subjects can also be asked to describe their emotional reaction toward an attitude object. Self-report of affect is rather untypical compared to other measurement techniques linked with specific indicator classes since it can be combined easily with many conventional scaling techniques (e.g. Thurstone, Likert, Guttman, Osgood (cf. Eagly and Chaiken 1993, 62)).

While self-report of affect does not require the amount of technical equipment the other measurements specialized on affective responses need, which is a clear advantage, it also shares the disadvantages of all self-report measurements: the possibility of conscious and unconscious distortions by the subject (cf. Eagly and Chaiken 1993, 63).

2.3.2.4.3 Using behavioral indicators Usually, it is not behavior itself that is used as a basis for attitude measurement, but verbally expressed *intentions* to act in a certain way. These declarations of intended behavior are often part of traditional scaling techniques, e.g. Likert-scaling. Instead of asking for **intended** behavior, self-report and/or other forms of assessment of **actual** behavior are also possible. Tittle and Hill did just that, using actual participation in student political activities and student elections

as indicators for attitudes (cf. Eagly and Chaiken 1993, 64)³²:

First, the voting behavior of each subject was determined by inspecting student-voting records in an election held one week prior to the administration of the questionnaire. Second, the respondent's report of his voting behavior for the previous four elections was taken as a behavioral indicator. (Tittle and Hill 1967, 206f)

Further data used:

(...) frequency of participation in meetings of a student assembly, frequency with which the individual had written to or talked with a student representative concerning an issue, (...) frequency of engagement in campaign activities of a particular candidate, frequency of reading the platforms of candidates for student political office, and frequency of discussion of student political issues in talking with friends. (...) (f)[F]requency of personal office seeking and response to an item indicating whether the respondent had ever written a letter of protest to the student newspaper. (Tittle and Hill 1967, 207)

Some of these criteria were used 'as they were' ("vote in last election", "vote over time"), some were used to create adapted versions of traditional scales: Guttman scalogram, "Likert fashion", and an "adaptation of the standard Woodward-Roper index of political participation" (cf. Tittle and Hill 1967, 206f).

A major problem with techniques of this type is that by asking for *past* behavior you can only measure *past* attitudes, not *present* ones. It is not always possible to use actions that took place only one week in the past, as Tittle and Hill did. Behavior elicited under controlled conditions (laboratory setting) might help assessing present attitudes, but will normally yield only one or two observed behaviors, which might not be enough to reliably deduce underlying attitudes. Furthermore, experimental settings tend to be rather artificial and might elicit different behavior than 'real life' situations. Therefore, such techniques are best suited for attitudes that influence frequently occurring behaviors.

It needs to be kept in mind when interpreting results of such studies that the relationship between attitudes and behavior is not as systematic as historically believed. A lack of certain behaviors does not necessarily mean a

³²Strictly speaking, Tittle and Hill 1967 did not intend to create/apply a new method of attitude measurement, but to assess common forms of measurement (Likert, Osgood and Guttman scaling).

negative attitude toward some attitude object (cf. Eagly and Chaiken 1993, 157f).

Marilyn S. Rosenthal (cf. Rosenthal 1974) used a variant of this technique in order to measure language attitudes of children aged 3-5. Children were introduced to two “magic boxes”, decorated card board boxes containing a tape recorder, one of which ‘spoke’ SAE and the other one ‘spoke’ AAVE. Children were not only asked to characterize the two boxes, but also given opportunity to act: They could choose from which of the ‘magic boxes’ they wanted to receive a gift, and to which of them they wanted to give a gift themselves (cf. Rosenthal 1974, 57ff). Due to the extremely young age of the subjects, more traditional forms of attitude measurement would not have been feasible.

2.3.2.4.3.1 Content analysis In the case of content analysis, behavior is not elicited through an experiment. Instead, evidence of past behavior manifest in texts such as newspaper articles, movies, political manifestos, soap operas, etc. (e.g. choice of terminology, expressed beliefs concerning AAVE, desired behavior as manifested in bills) is analyzed³³.

The major problem in the application of content analysis on the measurement of attitudes is the question: Whose attitudes do we measure by this technique? The attitudes of the authors/actors/directors/producers/editors involved in the production of the text? But an author might ‘cater to his/her audiences tastes’ instead of expressing his/her own attitudes.³⁴ The attitudes of society? Or of the target audience? This is a problem not easily solved. Content analysis can be a helpful tool when publicly expressed attitudes in past periods are to be measured and compared with the state of publicly expressed attitudes today; for example, attitudes toward AAVE before and after the Oakland controversy. It can also be used to learn more about the structure of beliefs held by a person or group.

³³It could also be argued that content analysis does not use behavioral indicators (the act of writing a book or directing a film constituting behavior) but cognitive indicators (the opinions stated directly or indirectly in the text constituting expressions of beliefs).

³⁴This problem is similar to that traditional methods experience with the effect of social desirability, but its effects are probably much larger.

2.3.3 The shortlist: Likert, Osgood and Thurstone scaling

Several of the measurement techniques discussed deserve to be considered for this study. Osgood scaling and Likert scaling are the most frequently used techniques in research on teachers' attitudes toward AAVE and therefore entitled to extra attention. Thurstone scaling has not been previously applied to the problem at hand, but merits being considered due to its high scale level.³⁵ Other techniques, of course, also have their specific advantages, but will not be taken into further consideration here.

The 'unique selling proposition' of Osgood's semantic differential is the option to develop a (nearly) culture-independent measurement instrument. Items that are used for developing Likert or Thurstone scales are always extremely culture-dependent. They can be used for the population of one country for a period of a few years, maybe a few decades, or only for a part of the population of one country, e.g. only for teachers. Comparing attitudes cross-culturally using a Likert- or Thurstone-scale is an experiment prone to fail. Unfortunately, this advantage also has a major disadvantage: Osgood scales are difficult to interpret. They might tell us that for American teachers, AAVE is more sour than SAE, or warmer than SAE. Only through factor analysis does this data get interpretable, e.g. if it yields an evaluative factor and a person responds differently for AAVE than for SAE on items that load high on this evaluative factor. An additional problem is the scale level of the data gathered by Osgood scaling, which is uncertain.

Since in a study on an attitude object like AAVE (an attitude object that is of little relevance in countries outside the United States³⁶) cross-cultural applicability is of little relevance, this 'unique selling proposition' has little attractiveness for this study, and therefore (for this specific study) the disadvantages of Osgood scaling weigh heavier than its advantages. This option is therefore discarded.

Likert scaling, developed as a simplification of Thurstone scaling, is relatively fast to develop and unproblematic to use. However, it is of a comparably

³⁵Steven's magnitude estimation reaches a comparable scale level (possible ratio, certainly interval). There has been so little experience with the application of magnitude estimation, though, that it was deemed inappropriate for this study.

³⁶It might be of interest to compare attitudes of teachers from different countries toward local basilectal varieties, i.e. toward different but related attitude objects, such as AAVE in the USA and British Black English in the UK, for example. There has also been some research on attitudes toward AAVE in countries such as Japan (cf. Cargile et al. 2006, who used attitudes toward AAVE as an indirect indicator of attitudes toward African Americans, or McKenzie 2008, who considered attitudes of Japanese students toward different varieties of English for purposes of improving TESOL).

low scale level, ordinal scale level, though. A higher scale level (interval) is possible when the Sigma method is used, but then the main advantage of Likert scaling – its comparable ease in scale construction – is, for the most part, lost.

Item choice is extremely difficult for Likert scales, too. When choosing items for a Likert scale, the neutral points of all items have to lie at approximately the same point, and the light/strong (dis)agreement points must have at least roughly comparable distances from the neutral point. Otherwise, it is not possible to generalize over items and to calculate any form of attitude score.³⁷

Thurstone scaling, on the other hand, is time consuming, but has a rewardingly high scale level (interval, possibly – with modifications – ratio) and avoids the ‘similar-neutral-point-similar-distances’ problem³⁸. For a study that aims at calculating, for example, correlations, a high scale level is a major advantage in itself, and clearly balances out the disadvantage of laborious scale construction. Some types of Thurstone scaling are liable to distortions through the anchor effect. This can be avoided by employing only Thurstone’s method of paired comparisons.

Therefore, I have chosen Thurstone scaling (and, among Thurstone scaling, the method of paired comparisons) as the methodical backbone of this study. A simplified form of Likert scaling, on the other hand, while not being employed within the main study, was used in a side study where analysis could be conducted in a way that avoided the potential pitfalls of Likert scales.

³⁷Of course, items can also be treated individually. Even if their neutral points and distances between extreme points are very different, each individual item will still be informative in so far as it allows judgement as to whether person A’s attitude is more or less positive than person B’s attitude.

³⁸For a comparison of Likert and Thurstone scales, see Figure 2.3.

Figure 2.3: Likert vs. Thurstone items

Chapter 3

Review of literature

In this chapter I will discuss instruments developed and studies conducted that center around teachers' attitudes toward AAVE. A plethora of studies has been conducted on attitudes toward AAVE in general, the discussion of which lies outside the scope of this thesis. Studies on individuals undergoing teacher training were included in this list. Table 3.1 gives details on the samples covered by each of these studies. In several cases studies used elements

Study	Decade	Region	Pre-service/ in-service teachers	schooltype	N
Williams: Chicago study	70s	inner-city Chicago	in-service teachers	elementary school	33
Williams: Texas study	70s	Central Texas area	in-service teachers	elementary school	175
Cecil	80s	rural Central Illinois	in-service teachers	elementary school	52
Di Giulio	70s	New York?	in-service teachers?	elementary school?	15?
Taylor	70s	U.S.	in-service teachers	all schooltypes	422
Covington	70s	inner-city Pittsburgh	in-service teachers and (partly inexperienced) assistant teachers	elementary school (individualized educational programs)	8
Ford	70s	U.S.	pre-service teachers	-	472
Bowie	90s	unspecified urban region	pre-service teachers	-	75
Hoover: Likert	-	-	-	-	-
Blake	00s	New York	in-service teachers	high school	88
Abdul-Hakim	00s	Florida	pre-service teachers	-	153
Hoover: other methods	-	-	-	-	-
Woodworth & Salzer	70s	New York	in-service teachers	elementray school	119

Table 3.1: Summary: samples

of Osgood- or Likert-scaling, but were not, in the stricter sense, Osgood- or Likert-scales. Those pseudo-Osgood or -Likert-studies are treated together with their 'stricter' 'relatives', even if they borrowed hardly more than the name from these methods.

I will begin with studies using a semantic differential, then discuss those using Likert's technique, and end with a discussion of those studies that used none of the above, nor claimed to have done so.

3.1 Osgood and pseudo-Osgood studies

In this section I will discuss all studies on teacher attitudes toward AAVE using Osgood-scaling, i.e. the semantic differential.

3.1.1 Williams

In the 1970s, Frederick Williams conducted two major studies on language attitudes: the **Chicago** and the **Texas** study. In both studies, teachers were presented with an Osgood-style questionnaire. In the Chicago study, audio records of African American and European American children were used as stimulus; in the Texas study video material was used and attitudes toward Mexican American children were also in the focus of attention. A separate scale was developed for each study, resulting in two distinct but similar questionnaires (cf. Williams et al. 1976, 9f).

3.1.1.1 The Chicago study

The Chicago study measured reactions to audio recordings of African American and European American elementary school children.

3.1.1.1.1 The subjects “33 primary school teachers, all from schools in inner-city Chicago, who were attending a summer institute in speech and language” (Williams et al. 1976, 26) served as subjects. This choice is slightly problematic since the participation in this summer institute might have influenced the rating behavior of the subjects, especially since rating took place during a period of four days (cf. Williams et al. 1976, 26) and on each day the subjects might have gained new knowledge. Additionally, subjects were not informed that the exercise they were participating in had the nature of an experiment (cf. Williams et al. 1976, 28)¹. This might raise the question whether subjects explicitly agreed to their data being used for research purposes.

Three subjects were male, 30 were female; 12 were African American, 21 were European American (cf. Williams et al. 1976, 26).

¹“The teachers were told that the listening sessions were a part of the institute's program” (Williams et al. 1976, 28)

3.1.1.1.2 The audio samples Audio samples were taken from a selection of sound tapes used in an earlier study (“Detroit study”). They contained recordings of African American and European American children of different socioeconomic backgrounds. The children attended fifth- and sixth grade at the time of taping (Williams et al. 1976, 25).

Only a part of the original recordings were used. Ten African American and ten European American children with similar socioeconomic backgrounds were matched. Half the sample consisted of male, the other half of female speakers. Each language sample consisted of an interview, performed at the children’s homes, on the topic of games and TV. The average length of recordings was 250 words (Williams et al. 1976, 24f).

Data on the linguistic features of each of these recordings existed from earlier studies (cf. Williams et al. 1976, 28). Assessments of language variety spoken were also available (cf. Williams et al. 1976, 25). This is relevant since the assumption that African American children automatically speak AAVE (and European American children automatically do *not*) might lead, if it is incorrect, to considerable distortions.

3.1.1.1.3 The questionnaire Samples of the tapes were played to subjects to elicit descriptions, from which, after some amount of editing, prototype scales were developed (Williams et al. 1976, 6f & 26). No pretesting of these prototype scales is mentioned. These are the opposition pairs used in the Chicago-study:

- WORD USAGES ARE: consistently incorrect – consistently correct
- THE CHILD IS: highly fluent – highly disfluent
- THE CHILD SOUNDS: male-like – female-like
- THE MEANING OF THE MESSAGE IS: very unclear – very clear
- PRONUNCIATION IS: nonstandard – standard
- SENTENCES ARE: complex-elaborated – simple-unelaborated
- THE CHILD USES LANGUAGE: effectively – ineffectively
- THE CHILD’S FAMILY IS PROBABLY: low social status – high social status
- THE AGE OF THE CHILD IS: seven, eight, nine, ten, eleven, twelve, thirteen, fourteen

- THE CHILD'S SPEECH INDICATES: a poor education background – a good one
- VOCABULARY IS: sophisticated – unsophisticated
- THE MESSAGE PERSPECTIVE IS: seldom tied to speaker – solely tied to him
- THE OVERALL MESSAGE IS: disorganized – organized
- SENTENCES ARE: fragmentary – complete
- THE CHILD SOUNDS CULTURALLY: disadvantaged – advantaged
- THE MESSAGE IS: rich in detail – sparse in detail
- THE CHILD SOUNDS: confident – unsure
- THE LANGUAGE SHOWS A: standard American style - marked ethnic style
- PRONUNCIATION IS: unclear-indistinct – clear-distinct
- THE GRAMMAR IS: quite bad – quite good

In the actual test booklets, different permutations of these items were used in the attempt to avoid order effects (cf. Williams et al. 1976, 26).

Williams deviates in his choice of opposition pairs from standard Osgood scaling. Most importantly, he does not use merely opposition pairs, but opposition pairs completing a sentence. This is highly untypical of Osgood-scaling. Additionally, in most cases, the opposition pairs are not adjective pairs: “a poor education background”, “seldom tied to speaker”. In some cases these pairs could have easily been rephrased in adjectival form, e.g. “rich in detail” as “detailed”. In one pair, the intermediate steps have been labeled individually, i.e. not “young” - “old”, but “seven” - “eight” - “nine” - “ten” - “eleven” - “twelve” - “thirteen” - “fourteen”.

In the traditional Osgood-approach, scales are developed to compare the semantic differential of different objects. Here, only the reactions of different subjects to a class of very similar objects (i.e., different language varieties) are to be measured. As a consequence of this approach the pairs used are tailored to the attitude object, i.e. language centered, or, concerning pairs such as “standard American style” – “marked ethnic style”, even specific to varieties of American English.

The subjects were divided into five subgroups. Each subgroup was presented with four voice recordings on four consecutive days, so that each group listened to and evaluated a total of 16 recordings. Since every group was given different tapes, this totals 80 different recordings, each evaluated by at least six judges (cf. Williams et al. 1976, 26).

3.1.1.1.4 Results The questionnaires were scored by assigning a value from one to seven to each item, the most ‘negative’ extreme being scored as seven² (cf. Williams et al. 1976, 28). The raw data is neither presented nor discussed.

3.1.1.1.4.1 The factors Factor analyses were conducted, one for the data from African American teachers, the other for data from European American ones. Data from both groups yielded four factors, two major and two minor ones. The two minor ones were omitted from further analysis and discussion (cf. Williams et al. 1976, 29ff). Factor I reflected a dimension dubbed “confidence-eagerness”, Factor II “ethnicity-nonstandardness” (cf. Williams et al. 1976, 31). The make-up of these two factors is not identical across groups. For African American teachers, for example, Factor I loaded higher on message detail, while for European American teachers fluency seems to be more important for Factor I (cf. Williams et al. 1976, 31).

In his report on the Texas study (in which similar factors were found), Williams suggests the following interpretation of the factors:

(...) ratings of confidence-eagerness seem to reflect perception of fluency in a situation. Ratings of ethnicity-nonstandardness may be a direct reflection of the grammatical characteristics exhibited in the child’s language. (Williams et al. 1976, 68)

No “evaluation” factor as such was found.

Status-judgment Three items, items 8 (family status), 10 (educational background) and 15 (culturally advantaged/disadvantaged), were considered to reflect a social status judgment (they did not, though, yield a factor). Those ratings of the children’s social status influenced both factors, ethnicity-nonstandardness *and* confidence-eagerness. The relation between

²The process of determining an opposition pair’s ‘orientation’ (positive vs. negative end) is not being discussed by Williams.

status-judgment and the two factors was greater for the European American than for the African American teacher group (cf. Williams et al. 1976, 31f). Status-judgment and actual status were related: The mean ratings of the high status group were higher than those of the lower status group. This did not apply to all subgroups of children, though:

Statistically reliable judgment differentiations of a child's actual social status were found mainly in the case of Negro male and female children rather than White children. (Williams et al. 1976, p. 39)

The data on male European American children showed a relationship, but not a statistically significant one, the data on female European American children even contradicted this trend (cf. Williams et al. 1976, 38).

In short, then, it could be said that reliable status differentiations were made for Negro children, although by negative implication the lack of further significant interactions was evidence of the generality of such differentiations across the further parameters of teacher race and speech topic. (Williams et al. 1976, 38)

Comparison linguistic data - evaluation Next, it was tested whether the data concerning the actual linguistic characteristics of the language samples had predictive power for the status-judgments and the two major factors:

[...] it was hypothesized that if individual speech characteristics are cues to a person's social status, then a set of such cues in quantified form should serve as salient predictor variables for the prediction of the status-judgments in the present data. Moreover, if the present interpretations regarding the two-factor judgmental model were valid, then the predictor variables for status-judgments should also relate in an interpretable fashion to the dimensions of the model. (cf. Williams et al. 1976, 32f)

The 82 original language measures were reduced (using statistical operations) to 17 variables that possessed predictive power. As an 18th variable, children's ethnicity was added (cf. Williams et al. 1976, 33ff):

- Production phenomena
 1. Silent pauses
 2. Filled pauses
- Amounts of production

1. Juncture total (number of clause terminals per topic)
 2. Word total (per topic)
- Syntactic elaboration
 1. Frequency of clause fragments
 2. Sentence length
 3. Mean number of immediate constituent division in verb constructions
 - Functional characteristics
 1. Reticence index (relative frequency of instances where the field worker's probes for elaboration were not followed by elaboration in the child's response)
 2. Introductory interjections (e.g. *now, anyhow, well, oh*)
 - Nonstandard characteristics
 1. Pronominal apposition
 2. Variance (in contrast to SAE) in verb construction
 3. Variance (in contrast to SAE) in [s] and [z] in word final position
 4. Variance (in contrast to SAE) in [ð] and [p] in medial and final position
 5. Variance (in contrast to SAE) in [m] in medial and final position
 6. Variance (in contrast to SAE) in [n] in medial and final position
 7. Variance (in contrast to SAE) in [ŋ] in word final position
 - Ethnicity (cf. Williams et al. 1976, 35ff)

All 18 variables had predictive power for status-judgment. In many cases, they could also be identified with one of the two factors (cf. Williams et al. 1976, 35ff). These results were interpreted as support for the two-factor judgmental model outlined before (cf. Williams et al. 1976, 37).

Effects of ethnicity It was mentioned before that a relationship between actual status and status-judgment was found only for African American (male) children. Other relationships between children's and/or teachers' ethnicity that were discovered by Williams et al. 1976 are the following:

- “the child’s race (...) did show a salient relation with status-judgments in the White teachers’ ratings but not in the data from the Negro teachers. Although the child’s race was related to the factor scores in the Negro teachers’ ratings, it was independent of their status-judgments” (Williams et al. 1976, 37f).
- “ratings of a child’s race were more of a central correlate of factor II (ethnicity-nonstandardness) for White teachers than for Negro teachers” (Williams et al. 1976, 39)
- “White teachers’ ratings of race were more correlated with status-judgments (-0.55) than were those of Negro teachers (-0.27). Even the influence of the child’s actual race appeared greater in the White teachers’ ratings than in those of their Negro counterparts.” (Williams et al. 1976, 39)

Generally, the influence of children’s (true or perceived) ethnicity appeared to have more influence on the judgment behavior of European American than on African American teachers:

These differences could be seen in the distributions of children as gained by dichotomizing the data on the status-judgment and race scales. White teachers, for example, placed 17 of the 20 White children in the upper half of the status distribution, as against only 12 White children so rated by the Negro teachers. All White children, whether rated high or low in status, or by Negro or White teachers, were rated as being White. As for Negro children, nine (of 20) were located in the high category by White teachers, but six of these children were also rated as being White (all, accidentally, were from the original H.S [= high status, addition mine] sample). Of the eight Negro children placed in the high category by Negro teachers, only two were rated on the White half of the race scale, and only barely so in terms of the magnitude of the ratings. In short, the bias in the White teachers’ ratings might be summarized as: sounding White is equated with high status. (Williams et al. 1976, 40)

Williams explains these differences between African American and European American teachers by differences in the amount and type of contact with the two varieties: African American teachers would be more likely to have had outside-of-school contact with AAVE and to have learned to differentiate effectively within this variety along the lines of effectiveness and status (cf. Williams et al. 1976, 41).

3.1.1.1.5 Evaluation Major departures from standard Osgood-scaling can be observed. The most relevant difference to Osgood's methodology is that Williams uses the complete scale as an indicator of attitude instead of only the evaluative factor (which Williams does not find in his factor analysis). He is, effectively, using a different definition for "attitude" than the developer of this technique did - and than I do here.

Other departures from standard Osgood-scaling, e.g. different methods of items choice and presentation, constitute only a minor variance.

3.1.1.2 The Texas study

The Texas study used an Osgood scale and video stimulus to study the attitudes of Texan elementary school teachers toward the language of European American, African American and Mexican American children.

3.1.1.2.1 The subjects The data of 175 in-service elementary school teachers from the Central Texas area was used in this study. Around 60% were from towns with less than 35,000 inhabitants, the remainder from larger cities. They belonged to a total of 15 different schools. They were sorted into five groups based on their teaching experience: 0-4 years, 5-9 years, 10-19 years, 20-29 years and more than 30 years of experience. Each of these groups consisted of 25 "Anglo" and 10 African American teachers (cf. Williams et al. 1976, 58).

The subjects did not know that they were participating in an experiment. As in the Chicago study, the questionnaires were presented to them as part of an in-service training program (cf. Williams et al. 1976, 59).

3.1.1.2.2 The video samples The video samples consisted of filmed semi-guided interviews conducted by a European American, female, mid-twenty interviewer in a comfortable atmosphere. The interviewees were 41 male fifth- and sixth-grade children of European American, African American and Mexican American background. Part of the interviewees belonged to a low status, the other part to a middle status socio-economic level, as defined by neighborhood/school attended and fathers' occupation (cf. Williams et al. 1976, 52f).

From these recordings, 24 clips of two minutes length each were extracted (the recordings of 17 children were discarded at this step). These were assembled into four sets of six clips each, each set containing one clip for each combination of ethnicity and status (cf. Williams et al. 1976, 58).

3.1.1.2.3 The questionnaire Again, stimulus material on video tapes was played to the subjects in order to elicit descriptions, from which, after some editing, prototype scales were developed (Williams et al. 1976, 54 & 26). This time, pretesting is mentioned, but not discussed in detail (cf. Williams et al. 1976, 54).

The questionnaires were administered in groups of around ten individuals at the site of the subjects' schools, usually at the schools' library (cf. Williams et al. 1976, 60). They were given a test booklet containing general explanations, a sheet for demographic data, three sheets for eliciting reactions toward certain labels, and six sheets for use with the video clips.

Before watching the video clips, they were presented with labels "representing children of a particular ethnic group" (cf. Williams et al. 1976, 60) and "asked to provide evaluations of their average or anticipated experiences with children of these groups" (cf. Williams et al. 1976, 60).

After that, subjects were shown one of the four sets of six video clips each and asked to fill out a questionnaire sheet for each clip they saw. For each video clip, the test booklet contained a separate sheet consisting of two parts:

The first part was a seven-options "semantic differential"³, containing the following items (items marked by an asterisk are filler items):

- The child seems: unsure - confident
- The language of the child's home is probably: standard American - marked ethnic style
- *The child seems: intelligent - unintelligent
- *The child probably spends: little - much time away from home
- *The child seems to be: sad - happy
- The child sounds: Anglo-like - non-Anglo like
- The child's home life is probably: like - unlike yours
- The child is: active - passive
- The child seems: reticent - eager to speak
- The child's family is is probably: low- high social status
- *The child is: determined - not determined in school

³This "semantic differential" shows the same diversions from traditional semantic differentials as that used in the Chicago-study (cf. 3.1.1.1.3).

- The child seems: hesitant - enthusiastic
- The child seems culturally: disadvantaged - advantaged
- The child seems to: like - dislike talking
- *The child seems: non-competitive - competitive (cf. Williams et al. 1976, 59)

Additionally, in the second part of these sheets, teachers were asked to assign the child to graded classes (remedial, below average, average, above average, far above average) for the subjects grammar, physical education, social studies, mathematics, spelling, music, composition and reading (cf. Williams et al. 1976, 59).

3.1.1.2.4 Results Most importantly, when submitted to factor analysis, the Texas study yielded, just like the Chicago study, a two factor model along the lines of (a) confidence-eagerness and (b) ethnicity-nonstandardness (cf. Williams et al. 1976, 60)⁴.

In the low status group, European American ('Anglo') children were perceived as the most confident-eager, followed by Mexican Americans, African Americans being perceived as the least confident-eager. This looked different for the middle status group, where European American children led in confidence-eagerness, followed closely by African Americans. For this group, Mexican Americans were perceived as the least confident-eager (cf. Williams et al. 1976, 63).

Concerning the ethnicity-nonstandardness factor, trends are somehow clearer. Here, European Americans always take the lead, appearing least ethnic-nonstandard. For the low social status group, Mexican Americans are perceived as more ethnic-nonstandard than European Americans, and African Americans as most ethnic-nonstandard of all three groups. In the middle status group, the ratings for Mexican and African American children are not significantly different from each other (cf. Williams et al. 1976, 65).

In all cases, the ethnic groups were rated as less ethnic-nonstandard and more confident-eager when they belonged to the middle as opposed to the low social status group.

Stereotyping The "evaluations of (...) average or anticipated experiences with children of these groups" (cf. Williams et al. 1976, 60) that had been elicited from the teachers were correlated with their judgments of individual

⁴Again, no raw data was presented or discussed.

children. A “statistically significant but negligible” (Williams et al. 1976, 65) correlation between confidence-eagerness and expressed stereotypes could be found for low-status children, but none for middle-status children. For ethnicity-nonstandardness, on the other hand, both socio-economic groups showed moderate correlations (cf. Williams et al. 1976, 65).

Teacher characteristics The teaching experience of subjects had no influence on their rating behavior (cf. Williams et al. 1976, 63), while teachers’ ethnicity had some influence: “minority group children [were, addition mine] being rated generally less ethnic-nonstandard by Black teachers than by Anglo teachers” (Williams et al. 1976, 67).

Teacher expectancy The nine school subjects included contained in the questionnaire were subsumed into three main categories, based upon intercorrelations found in the data (cf. Williams et al. 1976, 66). These three categories were (a) Language arts (grammar, spelling, composition, reading), (b) Language arts-related (social studies and mathematics), and (c) Non-language arts (music, arts, physical education) (cf. Williams et al. 1976, 66). For all subjects, the ratings on the Osgood scale had predictive power for teacher expectancy, though this relationship was more profound for those school subjects involving more language arts. Additionally, the more language arts involved in a school subject, the larger the predictive power of ethnicity-nonstandardness ratings as contrasted with confidence-eagerness ratings got.

3.1.1.2.5 Evaluation The strength and problems of the Texas study are basically the same as those of the Chicago study. The correlation between language ratings and teacher expectations for that student is interesting, since it supports the Rosenthal hypothesis (cf. Section 1).

3.1.2 Cecil

Cecil 1988 is listed in this section because Nancy C. Cecil refers to her questionnaire as a “modified semantic differential format” (Cecil 1988, 35). Even though I am going to show that this study does not constitute an Osgood study, I have subsumed it under the heading of Osgood studies for simplicity’s sake.

3.1.2.1 The subjects

52 white teachers from rural Central Illinois served as subjects. They were randomly assigned to one of two groups (cf. Cecil 1988, 35).

3.1.2.2 The audio samples

The second-grade teachers and principal of a Southeastern Missouri elementary school were given a three-hour training session on AAVE to ensure their ability to correctly identify speakers of this variety (cf. Cecil 1988, 35). They then identified 27 children (assumedly from within their school population) that spoke AAVE. Five AAVE-speaking and five SAE-speaking⁵ children were then selected for taping (cf. Cecil 1988, 35). Efforts were made to create two groups as homogeneous as possible:

From these two pools, two groups of five children for each dialect were selected who were as similar as possible with regard to age and tested intelligence (all children in the study were in the “average” range on the Stanford-Binet, varying no more than eight IQ points and varying no more than four months in age).

Each of the two groups was comprised of three girls and two boys and all ten children were in the second-grade in a suburban elementary school in South-eastern Missouri. All were from similar lower middle-class backgrounds. None had been retained. (Cecil 1988, 35)

Interviews centering around a stuffed animal were performed. The interviews were structured: All children received the same set of questions. Each interview lasted around five minutes (Cecil 1988, 35).

3.1.2.3 The questionnaire

Cecil describes her questionnaire as a “modified semantic differential format” (Cecil 1988, 35).

A total of three questions were asked:

- What do you think this child’s chances are of successfully completing second grade?
- What would you imagine to be the IQ of this child?

⁵The selection process for the SAE-speaking children is not being discussed. Probably they were picked from the group of children not identified as speakers of AAVE.

- How would you predict this child might perform in reading? (Cecil 1988, 36)

Teachers could respond on a scale from “1” (“Very Low”) to “5” (“Very High”) (Cecil 1988, 36).

This questionnaire might be called “modified semantic differential format” insofar as the outer appearance slightly resembles that of a semantic differential (Osgood) questionnaire. It does not, however, share any other characteristics of the semantic differential. It does not use adjective pairs⁶, and the total number of items is very low for an Osgood scale, but not prohibitively low (cf. Eagly and Chaiken 1993, 56). The collection of items is not discussed, nor does it seem that any of the pretesting customary for the development of a semantic differential was conducted.

This questionnaire does not, of course, constitute an Osgood-scale, but I would also refrain from labeling it a pseudo-Osgood-scale, since the authors borrowed so little from the semantic differential that it is equally justified to call it a pseudo-Osgood-scale as it would be to label it as a pseudo-Likert-scale.

In each group subjects listened to the audio tapes and then filled out the questionnaire (Cecil 1988, 35).

3.1.2.4 Results

A significant result (simple t-test) was found for the following observations: “teachers held higher overall expectations for the Black children who spoke SE than for the group that spoke BE”, “the teachers in the study thought that the children who spoke SE were more intelligent than those who spoke BE”, and “the teachers surveyed expected greater reading success from the children who spoke SE than from the BE speakers” (Cecil 1988, 36).

This study therefore confirms the frequently made assumption that teachers hold lower expectations for those students that speak AAVE than for those that speak SAE.

⁶Two items, though, could be changed in order to create a pseudo-Osgood-scale:
This child is:

unsuccessful - - - - successful

unintelligent - - - - intelligent

The third one would be difficult to adapt, since no adjective describing reading performance is available.

Questions	Williams (Ch)	Williams (T)	Cecil
Number of adjective pairs	20 opposition pairs (no adjective pairs)	15 opposition pairs (no adjective pairs)	not applicable (three questions)
Item selection process according to Osgood?	yes	yes	no
Item choice/item phrasing according to Osgood?	no	no	no
Strictly speaking an Osgood scale?	problematic	problematic	no

Table 3.2: Summary Osgood scales

3.1.2.5 Evaluation

Even though labeling the instrument used in this study as “modified semantic differential format” has been infelicitous, this should not make one discard the data collected by it too easily. The danger that low expectations lower school performance of students (Rosenthal effect) has been discussed frequently, and is one of the main arguments for attitude change programs targeted at teachers. This study is relevant in so far as it has brought forward additional evidence for the assumption that speaking a basilectal variety such as AAVE does indeed lower teachers’ expectations.

3.1.3 Summary

Table 3.2 summarizes some of the most important points concerning the scales used in the three studies discussed in this section.

3.2 Likert and pseudo-Likert studies

In this section I will review four Likert- and pseudo-Likert scales and the studies they were used in: The questionnaires by Di Giulio, by Taylor, by Hoover and by Blake. I will call something a Likert-scale that follows most of Likert’s rule for the construction of scales. A pseudo-Likert scale resembles a Likert scale in its outer appearance, but does not share the methodological premises of Likert-scaling. It could also be called a “Likert-type” scale (cf. Eagly and Chaiken 1993, 86).

3.2.1 Di Giulio

I will begin this section by reviewing Robert C. Di Giulio's work from the 1970s.

3.2.1.1 Subjects

Di Giulio stresses the exploratory nature of his questionnaire (cf. Di Giulio 1973, 25). While he discusses the development and pre-testing of his questionnaire in detail, we learn little if anything about the subjects the test was applied to. The best guess is that the results presented by him spring from the results of the last pre-testing, which was done using 15 elementary school teachers "selected by simple random sampling" (Di Giulio 1973, 26), presumably from New York.

3.2.1.2 The questionnaire

A pool of 100 items (cf. Di Giulio 1973, 26) was collected from "professional literature in the areas of speech, linguistics, education, and other related disciplines" (Di Giulio 1973, 26), and from "a series of fifty personal interviews with New York City teachers and administrators" (Di Giulio 1973, 26). Di Giulio assumes that attitudes change frequently and that opinion statements ("statements of attitudes") depend on social milieu (cf. Di Giulio 1973, 26), which of course implies that the best way of compiling opinion statements for a pool of items is directly from members of the socioeconomic group that is to be studied by the questionnaire and to do this 'just in time'.

The items were subjected to eight rounds of revisions. They were, in the customary Likert-format, presented to a group of teachers (no details concerning this group are offered, possibly it consisted of the same teachers that had been interviewed before), and these pre-test results were then used to check for ambiguity (items in response to which many individuals selected "I don't know"), ambivalence (extreme items that were rejected by almost every subject, even if the subjects are in concord with the orientation (positive or negative) of the item, merely for being too extreme⁷) and discrepancy (items that "were not in correlation with the teacher-subject's total attitude, and,

⁷This is one of the major shortcomings of the Likert-scale: Likert-scaling only works with items expressing moderate opinions. If you assume that "I think that AAVE is the best language in the universe" is an indicator of a positive attitude, then any person disagreeing with that statement would be, in the Likert-framework, considered to express a negative attitude toward AAVE. In Thurstone scaling, such a person might be recognized correctly as an individual holding positive, but not extremely positive, attitudes toward AAVE.

specifically, were seen to be diametrically opposed to other related items”⁸ (Di Giulio 1973, 26).

An equal number of positively and negatively stated items (24 each) was finally selected, plus five ‘neutral’ items (cf. Di Giulio 1973, 26). Why Di Giulio included five “items that reflected neither a positive nor a negative attitude” (Di Giulio 1973, 26) is not quite clear. Ideal items for a Likert scale are moderately positive or negative. Neutral items are not well suited for scaling using the Likert technique. It is not possible to score neutral items as you would score moderately positive or negative items. Disagreement with a neutral item indicates a non-neutral attitude, but the actual orientation of this attitude, positive or negative, cannot be deduced from it.

The selection was pre-tested once more, using 15 elementary school teachers. The coefficient of correlation between item score and total score of each subject was calculated for each item to determine the attitude-predicting ability of the items. Items with low correlation were discarded in order to ensure reliability (cf. Di Giulio 1973, 26). The final version of the questionnaire contained 23 of the original 100 items. The resulting questionnaire in its entirety is not presented in Di Giulio’s paper, nor does he offer a complete list of the 23 items.

3.2.1.3 Findings

Even though Di Giulio suggests the standard scoring of replies using numbers one to five, he does not present his findings in such a fashion, but rather gives a qualitative summary of the results.

The results imply a basically negative attitude toward AAVE on the side of the teachers studied: Teachers usually disagreed with the following (positive) items:

- Teachers should encourage the use of Black English (by both teachers and children) in the classroom.
- Black English is a logical predecessor to the use of Standard English in the classroom.

And accepted the following statements:

- Black English should not be used in the schools either as a ‘starting point’ for Standard English, or as a co-language with Standard English.

⁸This poses a problem for attitudes possessing an ambivalent internal structure. Since items that may function as indicators for ambivalent attitudes within a subject are excluded during scale creation, the resulting attitude measurements must give the appearance of non-ambivalent attitudes (cf. sections 2.2.6 and 6.2).

- The (spoken) use of BE in the classroom should be discouraged (by the teacher).
- The (New York City) Board of Education should not set guidelines with regard to BE.
- It is not important for a black child to be fluent in Black English.
- Teachers should not be required to take in-service training in BE.
- ‘Suburban’ (white) children should not be exposed to BE.
- ‘Urban’ (black) children should speak as the majority speaks.
- Black children should ‘replace’ their dialect with Standard English (cf. Di Giulio 1973, 26 & 49)

They disagreed, though, with some ‘classic’ statements concerning AAVE:

In an apparent contradiction of some recent research, teachers in this study neither perceived nor characterized their black students as being “non-verbal.” They do not conclusively describe black children as having “sloppy speech <sic>; nor do they feel black children have “lazy lips and lazy tongues” as previously reported under teacher attitudes. (Di Giulio 1973, 49)

Additional positively-oriented opinions shared by the teachers were:

1. Black English is not “slang”.
2. Teachers should not prohibit a child’s use of BE in the classroom, but neither should it be “encouraged”. Inherent in this apparent contradiction is both the feeling that black children should not be embarrassed by a correction and that their BE should be replaced (discouraged) in a “gentle” way, with a seemingly “better” form of language.
3. Although they agreed that BE should be discouraged, they did not feel that “every effort” should be made to do so. Similarly, they felt that “variations of Standard English” should be “allowed” in school, but without supporting the use of AAVE in schools. (Di Giulio 1973, 49)

To sum these findings up: Teachers were found to have negative attitudes, but did not endorse all negative statements included in the questionnaire.

3.2.1.4 Evaluation

Di Giulio's "Teacher Attitude Scale" is labeled a "Likert-type scale" (Di Giulio 1973, 25), but I would not hesitate to call it a Likert-scale. Items were selected carefully, and pre-tested accordingly. The only problematic decision made during scale construction was the inclusion of 'neutral' items.

3.2.2 Studies employing the Language Attitude Scale

The Language Attitude Scale (LAS), originally developed by Orlando Taylor during the 1970s, has been used by several researchers throughout the following decades.

3.2.2.1 Taylor

In this section, I will discuss the design of the LAS and the research conducted with it by Orlando Taylor.

3.2.2.1.1 Subjects 422 teachers served as subjects. 80 subjects were male, 342 female. They came from seven out of nine Federal Census districts⁹ (cf. 3.3). Taylor lists "North East" as a Census district. Usually, "Northeast" is not considered a Census Division, but a Census Region, consisting of the Census Divisions "New England" and "Middle Atlantic". Since the "Middle Atlantic" Census Division is listed separately, I assume that here, "North East" refers to the "New England" Census District (cf. U.S. Department of Commerce Economics and Statistics Administration U.S. Census Bureau, Geography Division nd).

The total sum of subjects according to region is 423 (not 422), so a typo must have occurred at some point.

111 subjects were African American, 281 European American, and 17 Asian American.

Besides region, sex and race, also teaching field, years of teaching experience, grade taught, the racial composition of the school the subjects worked at, and the educational level of subjects' parents were recorded.

3.2.2.1.2 The questionnaire From an initial pool of 117 items (no details are offered concerning how these were compiled), the final items were

⁹I refer to these as *Census Divisions*, following current usage at the U.S. Census Bureau.

Region	Number of Subjects	
	Urban Setting	Rural Setting
Middle Atlantic	36	0
South Atlantic	46	37
East North Central	84	13
East South Central	30	0
West North Central	0	21
Pacific	150	0
North East	0	6
Total	346	77

Table 3.3: Subjects according to region

selected based on a pre-test with 186 teachers across the United States. Based on assumptions on whether or not an item is to be evaluated as positive or negative (standard scoring was applied here), 25% of teachers with the most positive results were compared with those 25% of teachers with the most negative results, using the t-test. This test is used to identify items with low item sensitivity,¹⁰ which were then discarded (cf. Taylor 1973, 175). From the resulting list of items, two sets of 25 items were selected for the two final forms. Items were chosen in order to cover four “content categories” that will be discussed in more details under “findings” (cf. 3.2.2.1.3).

Form 1:

1. Black English sounds as good as Standard English.
2. Black English is cool.
3. Black English is an inferior language system.
4. Black English is too imprecise to be an effective means of communication.
5. The encouragement of Black English would be beneficial to our national interests.

¹⁰An item with which everybody agrees, whether his/her other statements indicate an overall positive or an overall negative attitude, does not help in differentiating subjects with positive from subjects with negative attitudes. This, of course, is only feasible if the orientation (positive or negative) of the majority of items has been established correctly. Additionally, the t-test is problematic when the attitudes measured are ambivalent. If most items tested reflect a positive attitude, those which reflect a negative attitude may be deleted for showing low item sensitivity.

6. Societal acceptance of Black English is important for development of self-esteem among black people.
7. When teachers reject the native language of a student, they do him great harm.
8. If use of Black English were encouraged, speakers of Black English would be more motivated to achieve academically.
9. It would be detrimental to our country's social welfare if use of Black English became socially acceptable.
10. The continued use of a nonstandard dialect of English accomplishes nothing worthwhile for an individual.
11. Allowing and accepting the use of nonstandard English in the classroom will retard the academic progress of the class.
12. A decline in the use of nonstandard English dialects would have a positive influence on social unity.
13. There is much danger involved in accepting Black English.
14. Widespread acceptance of Black English is imperative.
15. A child should not be corrected by teachers for speaking his native nonstandard dialect.
16. We should encourage the continued use of nonstandard English.
17. It is ridiculous to encourage children to speak Black English.
18. One of the goals of the American school system should be the standardization of the English language.
19. Teachers have a duty to insure that students do not speak nonstandard dialects of English in the classroom.
20. Black English should be discouraged.
21. A black child's use of Black English thwarts his ability to learn. (Taylor 1973, 177f)

Form 2:

1. Black English is a clear, thoughtful, and expressive language.

2. Nonstandard English is as effective for communication as is Standard English.
3. Black English is a poorly structured system of language.
4. Complex concepts cannot be expressed easily through nonstandard dialects like Black English.
5. Black English should be encouraged because it is an important part of black cultural identity.
6. Acceptance of Black English by teachers is vitally necessary for the welfare of the country.
7. To reject Black English is to reject an important aspect of the self-identity of black people.
8. Attempts to eliminate Black English in schools results <sic> in a situation which can be psychologically damaging to black children.
9. Continued usage of nonstandard dialects of English would accomplish nothing worthwhile for society.
10. Allowing Black English to be spoken in schools will undermine the school's reputation.
11. The scholastic level of a school will fall if teachers allow Black English to be spoken.
12. The elimination of nonstandard dialects is necessary for social stability.
13. In a predominantly black school, Black English as well as Standard English should be taught.
14. Nonstandard English should be accepted socially.
15. Teachers should allow black students to use Black English in the classroom.
16. Teachers should avoid criticism of nonstandard dialects of English.
17. The sooner we eliminate Black English, the better.
18. The sooner we eliminate nonstandards dialects of English, the better.
19. The possible benefits to be gained from approval of Black English do not alter the fact that such approval would be basically wrong.

20. A teacher should correct a student's use of Nonstandard English.
21. Children who speak only Black English lack certain basic concepts such as plurality and negation. (Taylor 1973, 179f)

These are, of course, only 21 items each, eight items are missing. On page 181f., Taylor presents a "(R)[r]andomized presentation order for LAS [Language Attitude Scale, addition mine], Form 1", which does contain a total of 25 items. Seven of its items are not contained in the previous lists of items: item 2: "Black English is a misuse of Standard English", item 8: "Black English must be accepted if pride is to develop among black people.", item 11: "Black English should be considered a bad influence on American culture and civilization.", item 12: "Black English sounds sloppy.", item 15: "Black English has a faulty grammar system.", item 21: "Acceptance of nonstandard dialects of English by teachers will lead to a lowering of standards in school." and item 25: "One successful method for improving the learning capacity of speakers of Black English would be to replace their dialect with Standard English.". Possibly, the seven 'new' items were added to the selection without pre-testing. Possibly, they were only accidentally left out from the lists of t-test results. The other 18 items consist of a mixture of form 1 and form 2 items. It is not possible to actually reconstruct the scale used from the information offered in the article.

The items seem to center around different attitude objects. Items such as "Teachers should avoid criticism of nonstandard dialects of English" may serve as a better indicator for attitudes toward nonstandard dialects in general.

3.2.2.1.3 Findings Standard scoring was used. For analysis, the items were collapsed into four "content categories". These content categories were not derived from statistical analysis, but used as a basis for item choice. Category 1: "Structure of Nonstandard and Black English" (Taylor 1973, 183), category 2: "Consequences of using and accepting Non-standard English" (Taylor 1973, 194), category 3: "Philosophies concerning use and acceptance of Nonstandard English" (Taylor 1973, 195) and category 4: "Cognitive and intellectual abilities of speakers of Black English" (Taylor 1973, 196).

For an overview of findings, see Table 3.4 (cf. Taylor 1973, 184ff).

Taylor summarizes these findings:

[E](e)xcluding topics dealing with the structure of Nonstandard and Black dialects, the majority of teachers throughout the country tend to reveal positive to neutral opinions. In the category

Content category	answer option				
	neg.	mildly neg.	neither	mildly pos.	pos.
1	20.7	18.3	20.5	17.9	22.3
2	13.3	12.1	17.5	24.6	32.5
3	16.7	17.6	17.9	20.5	24.8
4	13.0	17.3	17.8	21.8	30.2

Table 3.4: Percentage of negative, mildly negative, neutral, mildly positive and positive options chosen

pertaining to structure, they are about evenly distributed. (Taylor 1973, 197)

He interprets these results the following way:

[T](t)eachers' attitudes relating to various topics of Nonstandard and Black English vary from topic to topic. Thus, teachers do not appear to have a single, generic attitude toward dialects, but, rather, differing attitudes depending upon the particular aspect of dialect being discussed. (Taylor 1973, 197)

Since this scale is only of ordinal scale level, this conclusion is somewhat problematic. Only if all items are comparably 'extreme' (have same scale distance, and, more importantly, the same zero point), can they be compared this way. Strongly disagreeing with the positive oriented "I love AAVE more than anything else in the world" and agreeing with the equally positive oriented "I quite like AAVE" does not imply that these two items are independent of each other and "love of AAVE" and "liking AAVE" reflect two different attitudes. Only if item choice and pretesting have been done extremely carefully can it be hoped that zero point and scale distances are similar enough to make conclusions of this type safe.

Assuming that these methodological problems do not apply here, there are two ways that this data can be analysed. First, as Taylor did, it can be assumed that instead of having one attitude toward one attitude object, teachers hold different attitudes for different aspects of that attitude object, i.e. we have one attitude concerning "the structure of AAVE", one attitude concerning "AAVE use and acceptance", etc. Alternatively, it can be hypothesized that the attitudes observed are ambivalent, i.e. that people associate positive and negative evaluations with the same attitude object. The difference between these two interpretations lies in the assumed internal structure of the underlying attitude(s). Since my study created data that

pointed toward an ambivalent structure of attitudes, I will discuss this question in more detail in 6.2.

As a result of ‘data-mining’, Taylor finds:

- Teachers from predominately European American schools tend to hold more negative attitudes (cf. Taylor 1973, 198).
- Teachers with three to five years of teaching experience show more positive attitudes than their colleagues with more or less experience (cf. Taylor 1973, 197).
- Teachers from the Pacific urban region demonstrated more positive attitudes than those from other regions (cf. Taylor 1973, 199).

Since these findings are the result of data-mining, they should be considered as merely explorative.

3.2.2.1.4 Evaluation Taylor’s study stands out as one of the few studies with an N over 100, and with any attempt to create a balanced sample, especially the attempt to cover various regions of the U.S.

Unlike many other studies, Taylor’s study found relatively positive attitudes toward AAVE. His data-mining results are highly interesting and clearly deserve to be retested on new data. His discovery that attitudes toward AAVE might not be homogenous is very interesting, and, if proven correct, should influence the conception of future attitude change programs. I will discuss this question in more detail in Section 6.2.

It is of some interest here that Taylor misspelled the name “Likert” as “Lickert” (cf. Taylor 1973, 174). This does not necessarily serve as a criticism against this study (after all, typos simply *do* happen), but the fact that this typo was repeated by at least one other study employing the LAS developed by Taylor (Ford 1978; Blake and Cutler 2003 made the same mistake, but did not use the LAS) may indicate that (at least) this study picked this instrument only out of convenience, and without methodological consideration.

3.2.2.2 Covington

Covington 1975 tries to determine whether a relationship between teachers’ attitudes toward AAVE and their students’ educational attainment can be measured.

3.2.2.2.1 Subjects Four teachers and four assistant teachers from the third grade of a Pittsburgh inner-city school served as subjects. Six were African American, two were European American, three were male, five were female. One of the teachers had one year of teaching experience, the other three teachers had more than two years of experience in the classroom. Most assistant teachers had no teaching experience, only one of the assistant teachers had any teaching experience. The school they were employed at was at that time engaged in the development of individualized educational programs, i.e. forms of teaching by which students could progress through content matter at their individual speed (cf. Covington 1975, 43).

3.2.2.2.2 The questionnaire Several measures were combined in this study:

- “a modified teacher interview form from Taylor and Hayes” (cf. Covington 1975, 45):
 - Section A: dealing “with the teachers’ characterizations of the students’ speech. The teacher was requested to rate the speech of each student on a one to six point-scale” (Covington 1975, 45). The stimulus used consisted of audio recordings of semi-structured interviews with third-grade boys from the same school as the teachers (cf. Covington 1975, 43f)
 - Section B: “designed to measure teachers’ attitudes toward dialectal differences” (Covington 1975, 45).
 - Section C & H <sic>: “designed to measure the consistency of the teachers’ ratings” (Covington 1975, 45).
- the Language Attitude Scale as discussed above.¹¹. This questionnaire was used one month after the interview. It is important to note that scoring was done exactly *opposite* to convention: “1 point for strong agreement with a positive statement; (...) 5 points for strong agreement with a negative statement” (Covington 1975, 47).

Several measures were taken on the children whose voice recordings served as stimulus: reading achievement according to the “Metropolitan Test of Reading”, the progression through their individualized mathematics program, and an evaluation of their speech by linguists (cf. Covington 1975, 46f)

¹¹Covington 1975 notes “Taylor & Hayes, 1971” as source, but since she, unfortunately, does not include this title in her bibliography, I could not check whether there might be minor differences between this version and that presented in Taylor 1973. The description is detailed enough, though, to recognize that it is, in all important regards, identical to the instrument discussed in Taylor 1973.

3.2.2.2.3 Findings The mean scores of the LAS were computed: 1.0, 1.4, 1.5, 2.0, 2.0, 2.1, 2.8, 3.6; Mean: 2.0. The most positive score possible would be a 1.0, the most negative a 5.0. Teachers tended to be nearer the positive than the negative end.

Ann J. Covington correlated these with “(a) students’ language as perceived by teachers; (b) students’ language as perceived by judges; and (c) students’ reading achievement scores” (Covington 1975, 47).

Covington found the following statistical relationships:

- students’ language as perceived by teachers – student’s language as perceived by judges – reading achievement
- students’ language as perceived by teachers – reading achievement
- student’s language as perceived by judges – reading achievement

She did *not* find relationships between the following:

- students’ language as perceived by teachers – student’s language as perceived by judges
- the score at the LAS – progress in mathematics (cf. Covington 1975, 48)

In how far the interrelationship between perception of student language and reading achievement is due to teacher behavior based on this perception, or on interference of dialectal features with SAE learning-to-read materials cannot be deduced based on this data alone.

3.2.2.2.4 Evaluation The extremely small N of Covington’s sample reduces the generalizability of these findings. It is, though, an interesting application of Taylor’s questionnaire, and the fact that Covington could not find an interrelationship between teachers’ attitudes and students’ progress in a less language oriented subject such as mathematics should lead to further research. If this finding can be reproduced in studies with a larger sample, this would speak against the applicability of the Rosenthal assumption to non-language oriented subjects.

The fact that speech samples from students from the same school as the subjects were used might have distorted results. Some teachers may have recognized the voices of individual students. In that case, their judgment of students’ speech would likely be influenced by other factors, such as their knowledge about the students’ progress at school, the outer appearance of students, etc.

3.2.2.3 Ford

Ford assumes that foreign language teachers, since they are “less ethnocentric and more aware of, and sensitive to, cultural and linguistic differences in general” (Ford 1978, 382), have more positive attitudes toward nonstandard English. AAVE has been included in his study as an example of nonstandard English.

3.2.2.3.1 Subjects 472 prospective teachers of foreign languages [N=116], English [N'=114], mathematics [N'=97], and social studies [N'=145] (enrolled in corresponding classes in 1973/74) served as subjects (cf. Ford 1978, 384 & 387). Five teacher training institutions were covered, from “five of the eight national regions designated by the Modern Language Association for the purpose of its foreign language enrollment surveys” (Ford 1978, 384) (Great Lakes: 163 subjects; Mideast: 95 subjects; Plains: 73 subjects; Southeast: 58 subjects; Southwest: 83 subjects (cf. Ford 1978, 385)). The majority of subjects was European American and female (cf. Ford 1978, 384). Other data that were elicited were: urban/nonurban background, experience abroad and the teacher training institution enrolled in (cf. Ford 1978, 383).¹²

3.2.2.3.2 The questionnaire Ford 1978 used Taylor’s Language Attitude Scale:

The LAS was deemed most appropriate for the present study since it was designed specifically to measure teachers’ attitudes toward nonstandard English, and because of its applicability to a survey method of gathering data from relatively large groups of subjects. (Ford 1978, 384)

Ford also copied Taylor’s “Lickert”-typo (cf. Ford 1978, 384). Ford’s version of the LAS differs from the version presented by Taylor in the use of “nonstandard English”/“nonstandard dialects of English” instead of “Black English”, and the variant “Widespread use of nonstandard English is imperative” for “Widespread acceptance of Black English is imperative” (Ford 1978, 390, Taylor 1973, 181f). During the administering of the questionnaire, “at appropriate points in the instructions, brief audiotaped samples specifically identified as Black English and Appalachian English were played in order to orient the subjects as to what was meant by the term nonstandard

¹²The number of students at each teacher training institution can be deduced from the numbers according to the five regions. Numbers for urban/nonurban background and experience abroad are not available.

English” (Ford 1978, 384f).

The overview over “content categories of the LAS” (Ford 1978, 390), i.e. the list of actual items used, presents a mixture of the questions from Form 1, Form 2, and the “Randomized presentation order for LAS, Form 1” (cf. Taylor 1973, 177-182).

3.2.2.3.3 Findings Standard scoring was used: answers indicating a positive attitude received a score of five, those indicating a negative attitude a score of one (cf. Ford 1978, 385). For each content category, the scores of a subject were summed, and the subjects’ attitudes were, on the basis of these sums, evaluated as either “strongly positive, mildly positive, no opinion, mildly negative, [or] strongly negative” (Ford 1978, 385). The evaluation as “no opinion” implies that a person with neither a positive nor a negative attitude has no attitude at all (as opposed to a neutral attitude), which is a concept not shared by all attitude models. Additionally, the “neither agree nor disagree”-options of each single item might indicate entirely different points on an attitude continuum (e.g. neither agreeing nor disagreeing with “AAVE is the best thing in the universe” does not express the same attitude as neither agreeing nor disagreeing with “AAVE is a terrible abomination”). Even if all “neither agree nor disagree”-option should reflect the same point on an attitude continuum, this point can still be rather on the positive or negative end of the attitude continuum instead of at a truly neutral point.

The five gradations were reduced to three:

Because of the small frequencies in the ‘strongly negative’ category of all the tables, it was deemed advisable to collapse the tables by combining the ‘strongly negative’ and ‘mildly negative’ designations into one labeled ‘negative’. Accordingly, the designations ‘strongly positive’ and ‘mildly positive’ were combined into one labeled ‘positive.’ (Ford 1978, 385f)

For the average scores across content categories see Table 3.5.

Content Category	Positive	No Opinion	Negative
I	220	176	76
II	374	77	21
III	263	148	61
IV	324	76	72

Table 3.5: LAS scores

Each of the four demographic variables (subject, urban/nonurban, experience abroad, teacher training institution) were correlated with each of the four content categories. Additionally, for the subgroup of foreign language teachers, the data was tested for a relationship between content category on the one hand and urban/nonurban, experience abroad and teacher training institution on the other hand. The significance level was set at .05 (cf. Ford 1978, 386).

Ford found the following correlations:

- In three of the content categories (I, III and IV), a correlation between subject and attitude score could be found. Prospective teachers of English tend to have a more positive attitude, prospective teachers of social studies a more negative attitude. Prospective teachers of foreign languages and mathematics are positioned in between these two extremes (cf. Ford 1978, 386-388).

Of equal significance are the correlations he did *not* find:

- The variables urban/nonurban, experience abroad and teacher training institution seem to have no influence on the attitudes of foreign language teachers as a subgroup (cf. Ford 1978, 386).

It must therefore be concluded that the original assumption that the reduced ethnocentrism and heightened awareness of cultural and linguistic differences that were assumed to be found within foreign language teachers have a direct bearing on attitudes toward basilectal language varieties has not been verified. Instead, teachers of English were shown to have a more positive attitude to these varieties. Additionally, the type of contact with different cultures and languages created by traveling also had no influence on language attitudes.

3.2.2.3.4 Evaluation The interrelationship between the school subject an individual is preparing to teach and attitudes is a very interesting discovery. Whether prospective English teachers approach their studies already with more positive attitudes or whether they were influenced by their university training is a factor deserving further research, especially since the training of prospective English and foreign language teachers must be considered to be similar in many regards (e.g. contact with linguistics, study of literature, etc.). It may be assumed, though, that their training will be dissimilar in as many regards (i.e. elements of linguistics stressed).

It would also be very interesting to replicate this study for in-service teachers.

3.2.2.4 Bowie

Bowie and Bond 1994 used the LAS to research the question whether contact with the topic ‘AAVE’ during teacher training would influence prospective teachers’ attitudes toward it.

3.2.2.4.1 Subjects 75 pre-service-teachers served as subjects. Three quarters were undergraduate elementary majors, the rest were students at the graduate level. All students were “nearing the end of their degree and licensure program” (Bowie and Bond 1994, 113). Their teacher training institution was a large urban university. 86% were European American, 92% were female (cf. Bowie and Bond 1994, 113).

Even though a relevant part of these pre-service teacher were expected to teach at predominately African American schools after graduation (cf. Bowie and Bond 1994, 114), “only 63% of the respondents reported having received even minimal exposure to the topic of Black English in their preservice education, typically in the form of a single class discussion which reviewed some of the research on Black English and implications for teaching Black English poetry.” (Bowie and Bond 1994, 114). Only 19% (already included in the number of 63%) of subjects “said that the issue had been addressed substantially in their preservice training” (Bowie and Bond 1994, 114).

3.2.2.4.2 The questionnaire The questionnaire used was “an adapted version of the Language Attitude Scale” (cf. Bowie and Bond 1994, 114). Based on the items quoted in the article, we can assume that the questions of the “Randomized presentation order for LAS, Form 1” (cf. Taylor 1973, 181f) were used. Unlike the study by Taylor 1973, in this study subjects were offered only three answer-options: “agree, no opinion, disagree” (Bowie and Bond 1994, 114). Additionally, an open-ended question was used, in order “to elicit a more qualitative determination of attitudes” (Bowie and Bond 1994, 114). To ensure every student understood the term “Black English” correctly, the topic was introduced by the instructor (cf. Bowie and Bond 1994). The exact way this was done is not explained in detail.

3.2.2.4.3 Findings No invalid statistical operations were used to assess the data collected. The percentages of positive, neutral and negative replies for the complete group were collected: 41% negative, 32% positive, 27% indifferent responses were given (Bowie and Bond 1994, 114). This contrasts with earlier studies using the LAS which found predominately positive or neutral replies.

Several questions are reviewed separately:

- 76% disagree with the statement that Black English sounds as good as Standard English.
- 61% believe that Black English operates under a faulty grammar system.
- 60% support standardization of the English language as a general goal in the schools.
- 63% believe that it can harm a student if teachers reject his/her native language.
- 39% agree that attempts to eliminate Black English can be psychologically damaging to the African American student.
- 52% support the idea that Black English should be accepted socially.
- 39% believe that teachers should allow African American students to speak Black English in the classroom (cf. Bowie and Bond 1994, 114).

To assess the influence of contact with the topic during teacher training on language attitude, the reply behavior of non-contact and contact-group were compared. Four items were found that are more often disagreed with by members of the contact-group than by members of the non-contact-group (relevance was tested using the Pearson Chi Square test):

- “Black English is an inferior language system”
- “One of the goals of the American schools should be the standardization of the English Language”
- “Black English should be discouraged”
- “The sooner we eliminate Black English, the better” (Bowie and Bond 1994, 114)

Interestingly, the changes in reply behavior covered mostly “educational policy” items. This might have to do with the type of contact with the topic of AAVE during teacher training (cf. description of “exposure” situations in Bowie and Bond 1994, 114.).

The analysis of the open-ended question confirmed the general findings of the LAS-part of the questionnaire:

The most negative responses to this question came from respondents who reportedly had not been exposed at all to the topic in their coursework. (Bowie and Bond 1994, 115)

3.2.2.4.4 Evaluation The influence teacher training might have on language attitudes is highly interesting and relevant and deserves retesting with larger and more diverse samples, including in-service teachers. Statistically significant differences on 4 out of 25 measures between contact and no-contact group are too many to be easily discounted. Contact with a scholarly treatment of AAVE obviously influenced the beliefs of the contact-group, and thereby, probably, also the underlying attitude. It would theoretically be possible that change occurred mostly in teachers' beliefs concerning teaching practice, while their evaluations of AAVE remained largely unchanged ("AAVE is a terrible thing, but there is research that says that doing X in the classroom is better than doing Y, so I'll do X, even though I dislike AAVE nonetheless."). The change concerning the item "Black English is an inferior language system" and the reactions to the open-ended question do support Bowie and Bond's interpretation of the data, though. For a related study that came to opposite conclusions, see Section 4.3).

3.2.3 Hoover

In this section I will discuss the "African American English Teacher Attitude Scale". This scale is discussed alongside the "African American English Speech Varieties Attitudes Test" in Hoover et al. 1996. That scale will be treated in the section on studies employing other methods (see Section 3.3). I separated discussion of these two sub-scales because Abdul-Hakim 2002 (see 3.2.3.1) used only one sub-scale of the two-part questionnaire, and because the two parts are of an entirely different methodological make-up. The instrument consists of 46 items, most of which "were paraphrased from actual statements made by educators and laypeople" (Hoover et al. 1996, 386). The items reflect three different perspectives on AAVE: the 'excellence perspective', "which sees AA culture as multifaceted, ranging from the artistic to the academic (...) [and] assumes that African American children are capable of learning anything, and that their culture has adequate resources to facilitate such learning" (Hoover et al. 1996, 385), is contrasted with the 'deficit perspective' (that considers AAVE to be inferior) and the 'extreme difference perspective' (that treats AAVE culture not just as 'different' but

as ‘exotic’) (Hoover et al. 1996, 385f)¹³.

Items used:

1. “Most African American people’s major potential is in music, art, and dance.”
2. “African Americans should try to look like everybody else in this country rather than wearing Bubas and Afros.”
3. “African Americans need to know both standard and African American English in the school in order to survive in America.”
4. “African American English is a unique speech form influenced in its structure by West African languages.”
5. “The reason African Americans aren’t moving as fast as they could is that the system discriminates against them.”
6. “African American English is a systematic, rule-governed language variety.”
7. “African American English should be eliminated.”
8. “African American English should be preserved to maintain oral understanding and communication among African Americans of all ages and from all regions.”
9. “The African American community concept of discipline involves not letting children ‘do their own thing’ and ‘hang loose.’”
10. “African American kids have trouble learning because their parents won’t help them at home.”

¹³Generally, a two-perspective dichotomy dominates the linguistic literature: the contrast between the deficit perspective and the difference perspective. The deficit perspective interprets differences between a nonstandard and a standard lect as deficiencies. Typically, the nonstandard lect is considered to be a restricted code, borne out of an environment of ‘cultural deprivation’. A nonstandard lect is less good, less logical, less suitable for abstract thought and higher order learning. The difference perspective interprets differences as variance. The standard and the non-standard lect differ in grammar, pronunciation, lexicon, etc., but this cannot be interpreted as a deficiency on the side of either lect (cf. Fasold 1972, Eller 1989). The traditional difference perspective is not identical to Hoover et al.’s extreme difference perspective, which represents an approach in which differences are considered to be ‘exotic’.

11. "It is racist to demand that African American children take reading tests because their culture is so varied that reading is an insignificant thing."
12. "African American English should be promoted in the school as part of African American children's culture."
13. "Standard English is needed to replace African American English to help with worldwide communication."
14. "It is not necessary for African American children to learn anything other than their own dialect of African American English in school."
15. "The reason African American people aren't moving as fast as they could is that they're not as industrious as they should be."
16. "There is no such thing as African American English."
17. "The use of African American English is a reflection of unclear thinking on the part of the speaker."
18. "African American children's language is so broken as to be virtually no language at all."
19. "African Americans should talk the way everybody else does in this country."
20. "African American English is principally a Southern form."
21. "When a child's native African American English is replaced by standard English, s/he is introduced to concepts which will increase his/her learning capacity."
22. "The home life of African American children offers such limited cultural experiences that the school must fill in the gaps."
23. "African and African American hair and dress styles are very attractive."
24. "African American kids would advance further in school without African American English."
25. "African American English has a logic of its own, comparable to that of any other language."

26. "African American children can't learn to read unless African American Vernacular English is used as the medium of instruction in the schools."
27. "African American people have their own distinctive speech patterns which other people in this country should respect."
28. "African American English was produced by its history in Africa and this country and not by any physical characteristics."
29. "African American English can be expanded to fit any concept or idea imaginable."
30. "The home life of African American people provides a rich cultural experience directly connected to African origins."
31. "The reason African American children have trouble learning in school is that they are not taught properly."
32. "African American English is basically talking lazy."
33. "African American children can be trained to pass any test written."
34. "African American children can learn to read in spite of the fact that most readers are written in standard English."
35. "African American people have the same potential for achievement in math and science as any other people."
36. "African American kids are advantaged through African American English; it makes them bidialectal just as Chicanos are bilingual."
37. "African American English is a misuse of standard language."
38. "African American children should be allowed to choose their own course of study and behavior in school from an early age and should not be directed by the teacher."
39. "Standard English is superior to nonstandard English in terms of grammatical structure."
40. "African American English should be preserved because it creates a bond of solidarity among the people who speak it."
41. "Acceptance of nonstandard dialects of English by teachers would lead to a lowering of standards in school."

42. "African American English should be preserved because it helps African Americans to feel at ease in informal situations."
43. "African American English enhances the curriculum by enriching the language background of the children."
44. "African American English expresses some things better than standard English."
45. "Since only standard English is useful in getting jobs, it should always be preferred over African American English."
46. "African American English should be abandoned because it does not provide any benefits to anybody." (Hoover et al. 1996, 391ff)

The most remarkable characteristic of these items is that many seem to be without any relevance to language attitudes, e.g. "African Americans should try to look like everybody else in this country rather than wearing Bubas and Afros.". I assume that these items are used as indicators for a certain perspective on African Americans in general, and that this perspective is considered to necessarily influence language attitudes as well. Since separating language attitudes from attitudes toward groups is always problematic, using such items may compound the problem.

Some items could be argued to be double-barreled (cf. section 2.3.2.3.1), e.g. "The home life of African American people provides a rich cultural experience directly connected to African origins." A person might agree with the first part of the sentence "The home life of African American people provides a rich cultural experience" without agreeing with the second "It is directly connected to African origins" and vice versa. Likert recommended avoiding such double-barreled items if possible.

The items are presented using a four-point format. Subjects choose between "Agree Strongly", "Agree Mildly", "Disagree Mildly", "Disagree Strongly" (cf. Hoover et al. 1996, 386 & 391). Standard scoring, adapted to a four-point instead of a five-point format, is recommended for this scale (cf. Hoover et al. 1996, 386).

The reader is informed that "(I)[i]n several test administrations, the reliability of the scale varied from Cronbach alpha 0.89 to 0.93" (Hoover et al. 1996, 386). No other information, though, is given about the testing procedures applied to this scale before application, or the details about those test administrations whose Cronbach alpha's were listed.

3.2.3.0.5 Evaluation The instrument is interesting – not only due to its relative length, but also because of its combination with a non-traditional scale (cf. Section 3.3.1). Nonetheless, the steps that led to the creation of this scale, and the scale’s pretesting (which is to be expected although not mentioned) are not discussed in enough detail to allow a qualified evaluation of this questionnaire.

Additionally, the highly unusual and controversial choice of items poses the question of validity: whether or not this questionnaire measures attitudes toward AAVE and *only* toward AAVE. In any case, the selection criteria for items that Likert proposes are definitely stretched here.

3.2.3.1 Abdul-Hakim

Abdul-Hakim used Hoover’s “African American English Teacher Attitude Scale” to study preservice teachers’ attitudes in Florida.

3.2.3.1.1 Subjects 153 preservice teachers from two Florida universities served as subjects. 58.3% of subjects belonged to Florida State University, 41.2% to Orange University¹⁴ (cf. Abdul-Hakim 2002, 87 & 100). Florida University is the larger institution of the two, with a (predominantly European American) student population of 33,971. Orange University has 12,126 students and is predominantly African American (cf. Abdul-Hakim 2002, 87).

80 subjects were African American, 64 subjects were European American, 9 individuals belonged to other ethnic groups. The majority, 116 subjects, were female, 37 individuals were male (cf. Abdul-Hakim 2002, 100).

Besides university attended, ethnicity and sex, the following demographic data were also collected: hometown size (rural, urban or suburban), age, variety/varieties of English identified as “home language”, estimated family household income, length of residence in the U.S. or U.S. citizenship, contact with the topic of AAVE during high school and during college courses (cf. Abdul-Hakim 2002, 127f).

3.2.3.1.2 The questionnaire Abdul-Hakim used Hoover’s “African American English Teacher Attitude Scale” with minor modifications: the items were rearranged in order to put those items dealing with culture at the end

¹⁴Since the questionnaire targeted only students from these two institutions, it is not quite clear why 0.5% of subjects are not accounted for by these numbers.

of the questionnaire (cf. Abdul-Hakim 2002, 79). Abdul-Hakim foresaw potential problems in the use of this scale due to its, by 2002, “outmoded terms” (Abdul-Hakim 2002, 79), but decided to use it with unaltered wording.

3.2.3.1.3 Findings Standard scoring (adapted to a four-point instead of a five-point scale) was used.

Abdul-Hakim summed the attitude scores for each questionnaire, and calculated standard deviation, mode, median and mean for the total sample. Frequencies, percentages, means, standard deviations and sums were calculated for each item (cf. Abdul-Hakim 2002, 89). Multiple regression analysis was used to check associations between demographic factors and language attitudes (cf. Abdul-Hakim 2002, 89).

Since a total of 46 questions were used, 46 was the lowest theoretically possible score, 184 the highest possible one. The actual mean score was 127.34, the median 129 and the mode 132. The lowest score lay at 85, the highest at 170 (cf. Abdul-Hakim 2002, 94). These attitude scores were divided into three categories¹⁵: “high”, “middle” and “low”, contrasting with the system suggested by Hoover: “deficit”, “difference” and “excellence”. Table 3.6 gives an overview over these two trichotomies and the boundaries between their units (Abdul-Hakim 2002), 99).

A total of 7 subjects scored in the “high” range, 126 in the “middle” range

Hoover		Abdul-Hakim	
Deficit	under 120	Low	under 110
Difference	120-159	Middle	110-153
Excellence	160 or above	High	154 or above

Table 3.6: Trichotomies Hoover / Abdul-Hakim

and 20 in the “low” range.

The results of Abdul-Hakim’s multiple regression analysis are especially interesting. He found associations between language attitudes and two demographic factors: language (variety) used at home and hometown population. Subjects who grew up bidialectal or as monodialectal speakers of AAVE tend to have more positive attitudes toward AAVE than individuals with a monodialectal SAE background. Individuals whose hometown was suburban tended to express more positive attitudes than individuals with a rural or urban background (cf. Abdul-Hakim 2002, 107).

¹⁵Range, standard deviation and mean were used to establish the boundaries between these units (cf. Abdul-Hakim 2002, 98)

Abdul-Hakim found *no* association between language attitudes and ethnicity, university attended, sex, age, socioeconomic status, contact with AAVE during high school courses and contact with AAVE during college courses (cf. Abdul-Hakim 2002, 104ff.).

3.2.3.1.4 Evaluation Concerning methodology, Abdul-Hakim's approach carries with it the same merits and problems that have already been discussed in section 3.2.3.0.5.

Abdul-Hakim's research is relevant for the controversy surrounding attitude change programs. Does teaching about AAVE at school or at college influence attitudes? In this study, no association between contact with the topic of AAVE at high school or college could be found. While this does not demonstrate that no such effect is possible, it shows that more research is needed on which *form* of contact with this topic (if any) is best suited to achieve measurable attitude change.

3.2.4 Blake and Cutler

Blake and Cutler 2003 studied the interrelationship between teachers' attitudes and school philosophy.

3.2.4.1 Subjects

88 teachers from five New York high schools served as subjects. Schools were selected based on availability, subjects were "self-selected", i.e. questionnaires were handed out to all teachers and handed back on a voluntary basis. At each school, at least ten percent of the faculty participated (cf. Blake and Cutler 2003, 168).

67% of subjects were European American, 13% African American, 8% Latino, 2% Asian American and 10% 'other'. Slightly more than half of the respondents (55%) were male, 42% were female, 3% declined to indicate their sex. 10% had one year or less teaching experience, 17% between one and two years, 23% between three and five years, 5% six to ten years, and 45%, the largest group, ten or more years of teaching experience (cf. Blake and Cutler 2003, 169).

Composition of the student body differed between schools. 66% of subjects identified "African Americans as one of the two largest groups" (Blake and Cutler 2003, 169) at their school.

3.2.4.2 The questionnaire

Blake and Cutler 2003 cite two sources for their items:

The questions for the survey were culled from the questionnaire developed by Hoover et al. (1996, pp. 391-393). In developing the questions, we were also influenced by Taylor's (1973) study (...). (Blake and Cutler 2003, 166)

Few items were directly copied from these sources. Most items collected this way have been heavily edited, and several items have been added that were not included in Taylor 1973 or Hoover et al. 1996.

1. "People speak differently in different situations."
2. "In every language, there are always variations in the way people from different age, class and ethnic backgrounds speak."
3. "*Bilingual* education is the right of every child who does not speak the dominant language."
4. "Federal funds should be used to support bilingual education."
5. "Some children do poorly at school because they do not speak Standard English."
6. "*Bidialectal* education is the right of every child who does not speak the dominant language or dialect."
7. "Federal funds should be used to support bidialectal education."
8. "African American English (Ebonics) is a form of English."
9. "African American English is lazy English."
10. "African American English is subject to its own set of rules."
11. "African American kids would advance further in school without African American English."
12. "Using AAE as a tool to teach subjects to African American students would hurt their chances to learn Standard English."
13. "There are settings outside the classroom where African American English is appropriate."

14. “It is educationally sound to use a student’s first language as a way of teaching that student the standard language of the community.”
15. “AAE would be inadequate for teaching subjects such as social studies or math.”
16. “African American students have language problems similar to those of students learning English as a second language.”
17. “African American students should be taught in classrooms alongside English as a Second Language (ESL) students.”
18. “Standard English is dominant in schools and business because it is the best form of English.”
19. “One purpose of school is to make certain that all students graduate proficient in Standard English.”¹⁶ (Blake and Cutler 2003, 166f)

No further explanations are made about the collection or selection of these items.

Blake and Cutler 2003 label their questionnaire as a “Lickert-type scale” (note the ‘classic’ typo), not as a “Likert-scale”. Strictly speaking, they do not claim to have created a Likert-scale, but merely to have adopted the outer appearance of one, which is, actually, the case.

No form of pretesting is mentioned. Several items (e.g. “African American English (Ebonics’) is a form of English.”) in my opinion do not fulfill the ‘minimum requirement’ of Likert-items, i.e. possessing an ‘orientation’ (expressing a positive or negative evaluation, but not both). Like Hoover et al. 1996, the authors decided to include questions which do not possess direct relevance for the attitude object AAVE. Mixing items focused on AAVE and on bilingual education may be infelicitous should subjects hold different beliefs concerning items like “It is educationally sound to use a student’s first language as a way of teaching that student the standard language of the community.” when applied to bidialectal as opposed to bilingual settings.

3.2.4.3 Findings

Blake and Cutler 2003 decided to collapse their five categories into two, but fail to explain which way they did this (cf. Blake and Cutler 2003, 167f).

¹⁶Occasionally, several variants of the same item can be found in the research article. Question 9, for example, is also quoted as “African American English (Ebonics) is a lazy form of English.”

Two options exist: Either “Agree Strongly” and “Agree Mildly” were summarized on the one hand, and “No opinion”, “Disagree Mildly” and “Disagree Strongly” on the other, creating an ‘agreement’ versus ‘non-agreement’ dichotomy, or the “No Opinion” category fell into the first group, creating a ‘non-disagreement’ and a ‘disagreement’ dichotomy. I consider the first option to be the more likely one (especially since the authors label their graphs as “% Teachers in Favor” (cf. e.g. Blake and Cutler 2003, 174), but this is merely an assumption, not made explicit by the research article itself.

For most items no total numbers are available – merely results split up according to school and according to ethnic makeup of the student population. Using the numbers of subjects at each school, the original percentages can be reconstructed, but rounding mistakes, etc. cannot be entirely excluded. Those numbers in Table 3.7 that are printed in italics were extracted directly from the text. All other numbers were reconstructed by me. Blake and Cut-

Number of item	Number of teachers in favor	Percentage of teachers in favor
1 & 2		<i>95</i>
3	46	52.3
4	52	59.1
5		<i>90</i>
6	18	29.5
7	21	23.9
8	60	68.2
9		<i>14</i>
10	48	54.5
11	42	47.7
12	63	71.6
13		-
14	38	43.2
15	48	54.5
16	35	39.8
17		<i>14</i>
18		<i>28</i>
19		<i>93</i>

Table 3.7: Blake and Cutler 2003 results

ler 2003 report two major findings: Firstly, they found no interrelationship between “having taken at least one linguistics course” (Blake and Cutler 2003, 187) and teachers’ attitude. Secondly, they found an interrelationship between school philosophy and attitudes:

The most compelling trend in this study is that teachers' language attitudes appear to be influenced by the philosophies, or lack thereof, of the schools in which they teach. (...) While we resist making correlations due to the size of this study, we are struck by the observation that the two schools with proactive language and dialect programs are the ones with the most positive language attitudes among its teachers. It is important to reiterate that more in-depth attitudinal studies in the schools are needed to fully appreciate generalizations and idiosyncracies. (Blake and Cutler 2003, 186f)

3.2.4.4 Evaluation

Blake and Cutler's research question, the relationship between school policy and teacher attitudes, is highly relevant. Unfortunately, only a relatively small number of teachers from each school volunteered to participate. It might be the case that more of those teachers who stood in direct contradiction to their school's policy as compared to those teachers in perfect agreement with their school's policy refused to participate, which would explain the influence of school policy on measured attitude. Additionally, it is conceivable that school policy does not influence language attitudes but merely the employment strategies of a school or its appeal to teachers with certain attitudes. Teachers with attitudes not in harmony with school policy might look for another employer, taking their unchanged attitudes to another school. These questions definitely deserve further attention.

The fact that in this study, contrary to Bowie and Bond 1994, no relationship between contact to linguistics and language attitudes could be found, highlights the importance of further research on this topic.

3.2.5 Summary

The following Table serves as a summary of the Likert and pseudo-Likert scales discussed so far:

3.3 Studies employing other methods

Studies on teachers' attitudes toward AAVE that use neither Osgood- nor Likert-scaling – or, at least, that do not claim to use these techniques – are rare. Both Likert- and Osgood-scaling are valuable techniques, but the

Questions	Di Giulio	LAS	Hoover	Blake
Number of questions	23	ca. 25 (varies)	46	19
Carefull selection of items?	yes	unknown	unknown	unknown
Item choice according to Likert?	yes	mostly	problematic	problematic
Carefull pretesting?	yes	yes	unknown	unknown
Strictly speaking a Likert scale	yes	yes	problematic	problematic

Table 3.8: Overview over studies employing (pseudo-)Likert scales

degree to which they dominate this aspect of language attitude research is noteworthy.

What can occasionally be found are addenda to Likert- and Osgood-scales that are independent of these, e.g. Williams et al. 1976, for example, who, in addition to their pseudo-Osgood-scale, asked subjects to assign children to graded classes. Addenda are usually used to enable a further analysis of the main (in this case: the pseudo-Osgood scale) data, and are little valued in themselves (e.g. their results are not displayed, merely their correlations with the main data are discussed).

In this section, I will discuss the second sub-scale presented by Hoover et al. 1996, and the method used by Woodworth and Salzer 1971.

3.3.1 Hoover

Hoover et al. 1996 use two different instruments for their “African American English Attitude Measures for Teachers”: the “African American English Speech Varieties Attitudes Test” and the “African American English Teacher Attitude Scale” (cf. Hoover et al. 1996, 383). The second scale is of Likert-type and is discussed in that section. The first test is an ‘independent’ measurement and will be dealt with here.

3.3.1.1 The audio samples

The most notable fact here is that in this study, AAVE was not contrasted with SAE, but with “AASE” (“African American Standard English” (Hoover et al. 1996, 384). It is defined as follows:

AASE is distinguished by its “standard” grammar and varying degrees of Africanized intonation, pronunciation, and style. (Hoover et al. 1996, 384)

AAVE and AASE as understood here resemble my differentiation between ‘core’ speakers and ‘periphery’ speakers (see section 2.1).

The AAVE samples were designed to contain two specific grammatical features that were absent in the AASE samples: absence of the third person singular *-s* and negative concord. (cf. Hoover et al. 1996, 384). A direct comparison between two text samples illustrates the contrast; the numbers indicate parallel structures.

AAVE text:

Wiletta *can’t hardly never* (1) sit still. She say she just *have* (2) to keep busy. So she work around the house, or she *go* (3) out and shop, or else she get her girlfriend to show her the latest dance. She laugh when we tease her and say at least *nobody can’t* (4) call her lazy. Cause no matter what she *do* (5) she keep busy. (Hoover et al. 1996, 384)

AASE text:

Sharon King *has* (2) to cook and iron and keep house and she *almost never* (1) gets to go out and play anything. Sometimes when she gets tired, she tries to get through in a hurry or she asks her little sister to help her. And sometimes she just gets mad. But no matter what Sharon *does* (2), she still has to work and can’t play. That’s why Sharon *frowns* (3) so much. Nobody *wants* (4) to do all that. (Hoover et al. 1996, 384)

A total of 16 different texts were used (Hoover et al. 1996, 390f)

Four “bidialectal African American children” (Hoover et al. 1996, 384) were used as speakers. Every child read both AASE and AAVE passages. The speakers’ age, sex, socioeconomic background, etc. are not mentioned.

3.3.1.2 The questionnaire

The questionnaire contains five items. The five items are:

- A person who talks like this usually is:
Very uneduc’d Slightly Uneduc’d Can’t Decide Slightly Educated Very Educated

- This speech is most appropriately used in the following place:
No place
Playing in the Streets
Playing in the Living room
On the Playground at School
No Opinion
Eating at Thanksgiving Dinner
On a Church Program
On a School Program
Any Place
- A student who talks like this is most likely to be a:
Very Poor Achiever
Slightly Poor Achiever
Makes No Difference
Slightly Good achiever
Very Good Achiever
- Do you like this speech?
Dislike Strongly
Dislike Mildly
No Opinion
Like Mildly
Like Strongly
- This kind of speech is:
Very Nonstan'd
Slightly Nonstan'd
Can't Decide
Slightly Stan'd
Very Stan'd (cf. Hoover et al. 1996, 391)

Hoover et al. 1996, 385 discourage the use of 'attitude scores' for this part of their instrument:

Ratings given to the five dimensions should not be combined, because each dimension measures a different aspect of attitude. (Hoover et al. 1996, 385)

In a way, every questionnaire always measures a multitude of different aspects of an attitude. This in itself should not hinder a researcher from creating sums or averages of his/her scores (as long as the appropriate scale level is present). What is measured are always only indicators for a hypothetical construct that we cannot access directly. Yet refraining principally from combining and connecting these different indicators can be a major barrier in gaining an improved understanding of the attitude itself, as opposed to e.g. individual beliefs connected with it. In this context, though, the recommendation is perfectly acceptable, especially since the scale level is only ordinal.

3.3.1.3 Evaluation

The idea to contrast AAVE and AASE is highly interesting and improves the chances to find speaker truly ‘at home’ in both varieties. Having speakers read text samples improves comparability between speech samples, but also makes the recordings sound unnatural and possibly stilted. Especially using younger speakers who might not be perfect readers could result in very artificial language samples. The fact that the samples differed only in three grammatical features reduces naturalness even more. On the other hand, once again, it raises the comparability between the two guises.

Hoover et al. 1996 (387f) warn that the instrument should not be used without an accompanying workshop or other means of information on AAVE, since the makeup of the questionnaire might lead to wrong conclusions on the nature of AAVE (e.g. by implying that you can judge whether a speaker is educated or uneducated merely on the basis of his/her language variety, or by giving the impression that AAVE differs from SAE only in a very few grammatical features). This warning is legitimate – this is a factor that is relevant for a great majority of questionnaires.

We learn nothing about pretesting, nor about the item selection process. Therefore, little can be said about the inherent quality of this instrument.

3.3.2 Woodworth & Salzer

Woodworth and Salzer 1971 presented teachers with identical compositions read aloud by African American and European American children. They were asked to evaluate the quality of compositions using a variant of Braddock’s rating sheet for written composition (cf. Braddock et al. 1963). If teachers evaluate the same text differently when spoken aloud by children of different

ethnicity, this can point toward differing attitudes toward the language use of these children, or toward the children themselves.

3.3.2.1 Subjects

119 elementary school teacher undergoing professional education in five different classes at a Western New York college served as sample (Woodworth and Salzer 1971, 168).

3.3.2.2 The stimulus

The same social studies texts were read by an African American and a European American male sixth-grader and taped. A second European American male sixth grader read the distractor texts. At two sessions, three weeks apart, six audio samples were presented to the subjects, two of those being distractor recordings. Those two texts read during the first session by an African American child were spoken by a European American child in the second session, and vice versa.

Woodworth and Salzer 1971 find that the same texts were evaluated more positively when spoken by a European American child as compared to an African American child (cf. Woodworth and Salzer 1971, 169ff). How this is to be interpreted, though, is an open question. The authors state that

the black child read the material as written in standard English. That is, he did not alter syntax nor did he substitute black-dialect lexical variants. Neither did he appear to deviate substantially from conventional pronunciation of words. It seems likely that the teachers involved in the study picked up sufficient paralinguistic clues (e.g., intonation pattern) to identify the race of the child. (Woodworth and Salzer 1971, 171)

The authors state that (a) there was no **substantial deviation** from standard pronunciation, and (b) that there might have been different intonation patterns used. It is not certain whether the linguistic cues allowed an identification of ethnicity, or whether the voice quality of the speaker made a specific ethnicity plausible. Furthermore, since only two children's voices were compared, one cannot be certain that teacher evaluations were influenced by assumed ethnicity, as compared to, e.g. fluency differences in reading, or other non-ethnicity-related differences in speech.

3.3.2.3 Results

Woodworth and Salzer compared the ratings for both children on the variables introduction, variety, logical sequence, unity, transition, clarity, coherence, significance, critical thinking and overall grade (cf. Woodworth and Salzer 1971, 170).

Examination of the data revealed a consistent statistical difference between teachers' evaluations of material presented orally by black and white children. For each of the variables the white student received a substantially higher rating than the black child for identical material. (...) (s)[S]even of ten differences on the subscales reached levels of high significance. (Woodworth and Salzer 1971, 169f)

3.3.2.4 Evaluation

Even though its methodological problems are obvious, I have included this study in my review of literature because it employs an indirect technique of attitude measurement (measuring the effects that negative or positive attitudes toward a language variety/toward a speaker may have on the perceived quality of work presented by this speaker), which is rarely found in the literature on attitudes toward AAVE.

Further research on the question of whether ethnic cues or subtle linguistic cues triggered the different evaluations would be very relevant.

3.4 Overview

Table 3.9 gives a short overview over the results of the studies discussed here. General results are indicated using a seven-option method: +++, ++, + for very positive, positive and fairly positive attitudes, x for neutral attitudes, -, --, --- for slightly negative, negative and very negative attitudes in the population studied. The Osgood studies are marked by a question mark, since their results cannot easily be translated into degree of favorability unless an E factor has been established for several different attitude objects to warrant comparability. It is not possible to directly compare results across different instruments, therefore these evaluations are, of course, only rough estimates. Even though direct comparison between these studies is not a straightforward task – too multifold are the differences in methods, sample, region covered and time (see also Table 3.1) – it is obvious enough that the results of these studies have not produced identical results. In how far

these disparities in results reflect differences in attitudes between the different samples, and in how far they can be attributed to different methods and/or methodological mistakes, cannot be ascertained here. It is sufficiently clear, though, that there is still a research gap to be filled.

Study	General results	Method
Williams: Chicago study	?	Osgood (problematic)
Williams: Texas study	?	Osgood (problematic)
Cecil	- - -	free format (labeled as Osgood)
Di Giulio	- -	Likert
Taylor	x/+, exception: structure of AAVE (x)	Likert
Covington	++	Likert
Ford	+++, including structure of nonstandard English ($\neq$ AAVE!)	Likert
Bowie	-	Likert
Hoover: Likert	no data	Likert (problematic)
Abdul-Hakim	++	Likert (problematic)
Blake	+	Likert (problematic)
Hoover: other methods	no data	free format
Woodwort & Salzer	-	other

Table 3.9: Summary all methods

Chapter 4

Changing attitudes

4.1 The theory of attitude change

Within social psychology, attitude change has been a field of active and multi-faceted research. Just as an illustration: Eagly and Chaiken 1993, in their textbook on attitudes, use 444 out of a total of 794 pages, or 9 out of 14 chapters, to discuss the topic of attitude change and attitude formation. On the other hand, merely one chapter, or 65 pages, are dedicated to the structure of attitudes and beliefs¹. The sheer amount of space attributed to this topic in a survey-like textbook demonstrates powerfully that a detailed discussion of this question cannot be achieved within the scope of a work like this. Instead of detailing theoretical approaches, I will merely list **factors** that have been considered as contributing to attitude change. This, naturally, means ripping single elements out of complex theoretical frameworks and is accompanied with all defects that such an approach must be attributed with.

The effect of persuasive messages on attitudes depends on three clusters of factors. Firstly, there is a **message**. Then, there is the **sender** of the message and last, there is the **receiver** of the message. The message can be positive or negative, easy to understand or highly complex, short or long, made explicit or only implied, be frequently voiced or very rare. The sender of a message can be trustworthy or shady, can be an expert or a layperson, can have high prestige or practically no prestige at all, can be a friend, a political opponent, an actor just reading a script, or a person giving a highly emotional account of something very important to him/her. The receiver can

¹As discussed under 2.2.7, sociolinguistics tends more toward a somewhat atheoretical approach toward attitude. Sociolinguistics' stance on attitude change will therefore be discussed separately under 4.2

be bored and tired, or awake and fascinated by the topic. He/she can have prior attitudes or not, can have attitudes in accordance with the message content or directly opposed to them, can have thought a lot about this topic on previous occasions, or have never cared much about it, may have strong or less strong cognitive abilities, etc. All these factors, according to one theory or the other, have a bearing on attitude change.

This also means that when conducting an experiment on attitude change employing persuasive messages, all these factors should, at least ideally, be controlled for.

4.2 Popular claims

Linguists often encounter attitudes via the linguistic beliefs expressed by laypeople. They realize that these beliefs are often contrary to non-folk linguistic knowledge, and, therefore, assume that they – and, most importantly, their underlying attitudes – may be ‘remedied’ by ‘corrections’ through linguists.

Programs in linguistics have been suggested as a ‘remedy’ for negative attitudes – usually targeted at teachers – by a number of linguists. Some stress the role of linguistics for attitude change rather strongly, such as, concerning attitudes toward African American Vernacular English (AAVE), Abdul-Hakim 2002, and Okawa 2000;² others mention it only as one measure out of a whole catalog of suggestions, or recommend it while cautioning against its assumed limited effectiveness, e.g. Taylor 1973, 199; Shuy 1970, 176ff; Bowie and Bond 1994; Baugh 2001; Smitherman 2006, 138ff. The type of program suggested of course differs greatly.

Similar assumption have been made concerning other languages and language varieties (e.g. Löffler 1980, 102).

What all these suggestion have in common is that they focus on the role of the message in persuasive communication. Teaching linguistics is seen as a means to influence the belief system, and by which attitude change can be effected.

²Wolfram 1999 urges the establishment of dialect awareness programs, but centers his argumentation for them more on the presence of beliefs (that is: a specific type of evaluative response to an attitude object which is connected with an underlying attitude (cf. Eagly and Chaiken 1993, 11)) than on that of attitudes. Strictly speaking, therefore, he does not belong into this category, though being, of course, closely associated with it.

4.3 Language attitudes and linguistics: a side study

This side study examines the impact that two introductory courses in synchronic English linguistics had on German students' attitudes to German dialects. It was designed to test the often reported 'general knowledge' that contact with linguistics (often not further specified) produces more positive attitudes to basilectal language varieties. The fact that these course taught *English* and not *German* linguistics should be seen as an advantage to the research design: The questionnaires would have been perceived less as attitude questions than as knowledge tests if they had been used within an introductory course for German linguistics, a danger that any attitude questionnaire faces, but especially those used in a teaching context.

A questionnaire containing demographic questions and twelve language attitude questions in a simplified Likert format was distributed among two introductory courses in synchronic English linguistics at two different universities: Aachen University and Siegen University. The first version of the questionnaire was handed out the beginning of the term, the second version toward its end³. Participation was voluntary. Sufficient time to complete the questionnaire during class was given. Students remained anonymous.

Questionnaires filled out by the same person could be matched by a identification code consisting of a combination of letters and numbers based on the subjects' date of birth and his or her parents' first names. Only those pre-test questionnaires that could be matched with their post-test equivalents were retained.

The Aachen data was collected during the winter term 2005/2006 in the lecture of Prof. Meyer, the Siegen data during the winter term 2006/2007 in the class of Dr. Arndt-Lappe. In Aachen, the questionnaires were distributed by the researcher, in Siegen by the class's teacher.

4.3.1 The subjects

For the Aachen group, a total of 66 valid questionnaire pairs forms the basis of the analysis. In some cases, a person refused to note down his or her identification code; in some cases, he or she might have misinterpreted the instructions on one occasion and interpreted it correctly on the other occasion, making matching his or her questionnaires impossible; several students that

³In the Aachen group, it was distributed at the first and the last lecture, in the Siegen group at the third and at the last two meetings.

were present in the first week might have missed the lecture in the last week, and vice versa. All this resulted in this relative low number of complete sets, compared to total number of questionnaires (169 for the pre-test, 131 for the post-test). Since the questionnaire was in German and centered around German language dialects, subjects with native languages other than German were eliminated from the list, further reducing the number of analyzed questionnaires. Questionnaires by individuals indicating bi- or multilingualism and identifying German as one of their native languages were retained. The Siegen sample analyzed here consists of 19 questionnaire pairs, selected from the total number of questionnaires (102 for the pre-test, 30 for the post-test) in an identical fashion.

For the Aachen sample, the average age of subjects was 20,7 years, with a modus of 19 years. The youngest subject was 18, the oldest 30 years old (for details see Table 4.1).

49 subjects were female, 17 male.

Age	18	19	20	21	22	23	24	26	27	30
Number	1	24	17	9	5	5	1	1	1	2

Table 4.1: Age of subjects (Aachen)

22 labeled themselves as dialect speakers, two of those with some hesitation. All stated that their dialect was German (as opposed to Austrian or Swiss); one person stated that she spoke both a German and a non-German dialect. 52 out of 66 students participated in a teacher training program (“Lehramtsstudenten”). 14 subjects did not offer details of their combination of studied subjects. All subjects that indicated their course of studies studied English literature and/or linguistics, usually in combination with other subjects within or outside the humanities.

Only 17 subjects indicated prior contact with linguistics.⁴

The age structure of the Siegen group was similar. The average age was 20.7 years, with a modus of 20 years, the youngest subject being 19 and the oldest subject 27. For details, again, please see Table 4.2.

For this sample, sex was better balanced: 11 subjects were female, 8 were

⁴If contact with linguistics improves attitudes toward basilectal varieties, it is very unlikely that repeated contact (as is the case with those 17 subjects) worsens attitudes. Either attitudes should be further improved, or, more likely, remain stable. The number of subjects with prior contact with linguistics, therefore, is relevant only in so far as a very high number of such individuals in a sample might reduce the overall effects when assessed using the Bortz modification, discussed below.

Age	19	20	21	22	23	27
Number	4	7	4	2	1	1

Table 4.2: Age of subjects (Siegen)

male.

8 identified themselves as speakers of a German dialect, 11 did not.

The great majority, 17 out of 19, participated in the teacher training program. Two students did not indicate the subjects they studied, all other studied a combination of subjects including English language/English linguistics.

Four subject had had prior contact with linguistics, 15 had not.

4.3.2 The questionnaire

The twelve attitude questions were designed in a simplified Likert format. They use the usual five-gradation answer scale, but were created with less pre-testing than is normally recommended for a Likert scale.

The statements were selected from a pool of intuitively chosen opinion statements. Each class of evaluative responding is represented by four statements. Affective responding by statements 1, 2, 7 and 10:

- 1: “A person speaking dialect is directly unappealing to me.”
- 2: “I am glad to be able to speak dialect/unhappy to be unable to speak dialect.”
- 7: “Dialects are nice”
- 10: “I think that Standard German sounds nicer than dialects.”

Behavior/behavior intention by statements 3, 5, 9 and 11:

- 3: “I would refuse to read a book written in dialect.”
- 5: “If I could watch a theater play in dialect or in Standard German, I would chose the one in dialect.”
- 9: “As an employer I would always hire the most qualified candidate, even if this person spoke dialect.”
- 11: “If I were a teacher, I would reprimand students speaking dialect.”

And cognitive responding by statements 4, 6, 8 and 12:

- 4: “Dialects and standard language do not differ in their degree of complexity.’
- 6: “Whether somebody speaks dialect or not implies nothing about his degree of education.”
- 8: “Dialects are a broken form of German.”
- 12: “Dialects are more efficient than Standard German since they require less movements of lips and tongue for their pronunciation.”

Two statements in each dimension are considered to express a positive attitude to dialects, and two a negative attitude (see Table 4.3 for details). The twelve questions were randomly sorted.

Number	Statement	Dimension	Orientation
1	unappealing person	affective	negative
2	glad to speak dialect	affective	positive
3	book	behavior	negative
4	complexity	cognitive	positive
5	theater play	behavior	positive
6	education	cognitive	positive
7	nice	affective	positive
8	broken form of German	cognitive	negative
9	employer	behavior	positive
10	Standard German sounds nicer	affective	negative
11	teacher	behavior	negative
12	movements of lips and tongue	cognitive	negative

Table 4.3: Overview questionnaire items

As it is generally the case with Likert-style questionnaires, this scale is ordinal. The point of origin (the point in the scale of the perfectly neutral statement) and the exact differences between statements are unknown. Ordinal scale level means that all mathematical operations involving addition, subtraction, multiplication and division are problematic, and so are statistical methods such as calculating the arithmetic middle (see also section 2.3.1.4). I will attempt an interpretation of this data using only unproblematic methods.

4.3.3 Results

During scoring, a “strong agreement” with a question expressing a positive attitude would receive a score of five, with questions expressing a negative attitude a score of one (Standard scoring). “strong disagreement” would lead to a score of one in positive attitude questions, and a score of five in negative attitude questions. The three responses in between were graded accordingly with scores from 2 to 4, a score of three signifying the “neither nor” reply (see Table 4.4 for examples).

Positive attitude questions Example: “Dialects are nice”				
Strong agr.	Slight agr.	Neither nor	Slight disagr.	Strong disagr.
5	4	3	2	1
Negative attitude questions Example: “I think that Standard German sounds nicer than dialects.”				
Strong agr.	Slight agr.	Neither nor	Slight disagr.	Strong disagr.
1	2	3	4	5

Table 4.4: Coding

Occasionally, a subject left a question blank. This was then coded as “NA”. As can be seen in Figures 4.1 to 4.6, the evaluative behavior for each item on the first and second questionnaire were rather similar for most items.

I tested the following hypotheses:

H_1 : The introductory course in linguistics did improve attitudes toward German dialects.

H_0 : The introductory course in linguistics did not improve attitudes toward German dialects.

I decided to use a directed H_1 since I wished to test the frequently made claim that contact with linguistics improves language attitude, while the opposite claim, that contact with linguistics has a negative influence on language attitudes, is not normally made.

Since I had a dependent sample in ordinal scale level, I decided to use the exact binomial test. All calculations were performed using R: A Language and Environment for Statistical Computing (binom.test, option: greater (directed hypothesis)). All calculations were done twice, first neglecting cases in which no change occurred, and second using the modification suggested by Bortz and Lienert to account for those no-change-cases (cf. Bortz and Lienert 1998, 175).

Only one question showed a significant change on the 5% level: question 4,

Figure 4.1: Cognitive Dimension - Aachen

Figure 4.2: Affective Dimension - Aachen

Figure 4.3: Behavioral Dimension - Aachen

Figure 4.4: Cognitive Dimension - Siegen

Figure 4.5: Affective Dimension - Siegen

Figure 4.6: Behavioral Dimension - Siegen

and even this question only for the Aachen group.⁵ In all other cases, changes between pre- and posttest did occur, but were not significant (see Tables 4.5 and 4.6).

Question	positive	negative	no change	NA	binomial	Bortz
1	24	15	26	1	0.0998	0.1605
2	11	21	33	1	0.975	0.9157
3	21	18	26	1	0.3746	0.4022
4	26	12	22	6	0.01678	0.04623
5	17	14	35	0	0.3601	0.4022
6	13	12	41	0	0.5	0.5
7	17	19	30	0	0.6911	0.6439
8	21	17	24	4	0.3136	0.3518
9	7	13	43	3	0.9423	0.813
10	21	11	32	2	0.05509	0.1302
11	20	16	30	0	0.3089	0.3561
12	16	17	33	0	0.6358	0.6043
Total	214	185	375	18	0.08045	0.1569

Table 4.5: Binomial test results Aachen

Question 4 was one of the cognitive dimension questions: “Dialects and standard language do not differ in their degree of complexity.” From a linguistic point of view, of course, the case is clear: There is no difference in complexity between Standard German and the various German dialects. This is a fact that is usually highlighted by the instructor using examples of grammatical constructions present in his native dialect that Standard German lacks. The reactions to these demonstrations are usually hilarious laughter – not necessarily because students consider this dialect to be ridiculous, but rather stemming from their surprise at hearing a professor of linguistics speaking in dialect during a formal lecture. The form of reaction should warrant that students keep this part of the lecture in memory. Therefore, it would have been very surprising indeed, if no change had taken place here at all. It is difficult to establish, though, whether the change is due to a change of the underlying attitude - or even of the individuals’ personal beliefs – or whether the question was perceived and answered as a mere exam-like test-question. Figure 4.7 illustrates the changes that occurred. The vertical indicates the evaluative behavior before the lecture, the horizontal after the lecture. The

⁵Due to the extremely small sample size of the Siegen group, minor effects would be extremely difficult to prove. The absence of statistically relevant effects does not exclude the possibility of there being an effect, but it makes it plausible that if such an effect exists, it would be rather small.

Question	positive	negative	no change	NA	binomial	Bortz
1	3	2	14	0	0.5	0.5
2	3	3	13	0	0.6563	0.5927
3	7	3	8	1	0.1719	0.2403
4	8	7	4	0	0.5	0.5
5	6	3	10	0	0.2539	0.3238
6	2	4	13	0	0.8906	0.7597
7	6	2	11	0	0.1445	0.2403
8	5	9	5	0	0.9102	0.881
9	2	4	13	0	0.8906	0.7597
10	2	8	8	1	0.9893	0.9519
11	2	5	12	0	0.9375	0.8204
12	4	4	11	0	0.6367	0.5927
Total	50	54	122	2	0.6879	0.6302

Table 4.6: Binomial test results Siegen

black highlighted fields indicate those subjects that did not change their evaluation. The gray highlighted fields indicate subjects whose answer behavior underwent major changes, i.d. from a negative evaluation (“1” or “2”) to a non-negative (neutral “3” or positive “4” and “5”) evaluation.

It can be seen that a relatively large group (21 individuals) changed from a negative to a neutral-positive result, while only very few individuals (8) changed from a neutral-positive to a negative reply.

If you compare the results concerning question 4 with those of question 8 (“Dialects are a broken form of German.”), the difference is stunning. Question 8 is, from a linguistic point of view, just as straightforward as question 4. It is also as likely to be covered in an average introduction to linguistics (though probably not applied to the question of *German* dialects) as question 8. But nearly as many students changed their opinion to the negative as to the positive end of the scale. Why there was an effect concerning question 4 but none concerning question 8 is an open question. I assume that without the ‘fun presentation’ that is regularly part of this lecture (and that, probably, was not employed by the lecturer of the Siegen group), the effect concerning question 4 would have been nil as well. The significant change in item 4 cannot be attributed to attitude change, but merely to ‘studied knowledge’ not integrated (or integrated without significant effect on the overall attitudes) into the belief system associated with this attitude object. If such an attitude change had occurred, a more widespread change in answer behavior should be observable.

After Before	1	2	3	4	5	<i>n/a</i>
1	0	3	0	0	1	<i>1</i>
2	0	7	16	3	2	<i>1</i>
3	1	4	7	3	2	<i>1</i>
4	0	3	4	5	6	<i>0</i>
5	0	0	0	0	3	<i>0</i>
<i>n/a</i>	<i>0</i>	<i>1</i>	<i>1</i>	<i>0</i>	<i>1</i>	<i>0</i>

Figure 4.7: Question 4

To summarize the results: There was no significant change in any affective dimension question. There was no significant change in any behavior dimension question. There was significant change in one out of four cognitive dimension question for one of the two samples. Significant change in one out of twelve indicators of a construct does not justify assuming change in the underlying construct (in this case, change in the underlying attitude). I therefore do not reject H_0 : The introductory course in linguistics did not improve attitudes toward German dialects.

It must be concluded that if the questionnaire did indeed measure attitudes toward German dialects, even if only in a ‘quick and dirty’ way, that mere contact with linguistic knowledge – in the form of a general introductory course – is not an effective means for improving attitudes toward basilectal language varieties.

4.4 Language attitudes and linguistics: outlook on the main study

In this chapter, I have highlighted the multitude of factors that, according to different theories of attitude change, influence attitudes. I have also shown that for many linguists, sending a message with a specific content, that is, teaching linguistics, is seen as an effective or at least feasible way to create attitude change in regard to basilectal language varieties. I have conducted

a preliminary study to test this hypothesis, and could not find a direct relationship between studying introductory linguistics and improved attitudes toward basilectal varieties. In the following chapter, I will present the main study in which I, again, tried to check the effectiveness of this approach to attitude change.

Chapter 5

Study

So far, I have shown that for a sample of German students of English Studies, contact with linguistics had no influence on their attitudes toward German dialects. In this study, I will look at American teachers and their attitudes toward AAVE. Besides the obvious differences in sample and attitude object, the methodology for this study also differs significantly. Instead of using a pretest-posttest format, an ex-post-facto format was used, i.e. the currently held attitudes of subjects were viewed in relation to past experience (e.g. linguistics classes). Instead of using a simplified Likert scale, a Thurstone scale was developed. The main focus lay on studying the interrelationship between contact with linguistics and attitudes toward AAVE, but the data collected surprisingly revealed highly relevant characteristics of the inner structure of those attitudes measured, which I will address in detail in the following sections as well.

5.1 Methods

In this study, data was collected using online questionnaires, comprising a Thurstone scale and a scale for the self-assessment of degree of contact with linguistics, plus questions that elicited demographic data and one question that asked whether the subject had perceived an effect of contact with linguistics on their attitudes toward AAVE.

5.1.1 The questionnaire

In this section, I will outline the steps that led to the creation of the questionnaire, especially the creation of the Thurstone attitude scale.

Label	Frequency
African American English	32 (62.7%)
African American Vernacular English	28 (54.9%)
Black English	38 (74.5%)
Black English Vernacular	27 (52.9%)
Black Speech	26 (52%)
Broken English	38 (77.6%)
Ebonics	48 (92.3%)
Non-Standard English	41 (80.4%)
Slang	48 (92.3%)
Street Slang	30 (58.8%)
Vernacular Black English	25 (50%)

Table 5.1: Labels recognized by at least 50% of teachers

5.1.1.1 Choice of terminology

For my purposes, the ideal label:

- Is understood by all subjects.
- Carries roughly the same denotative meaning for all subjects.
- Is not considered offensive by any subject.
- Should not be in itself highly evaluative.
- Should be neutral in its wording and not create tautologies or contradictions in the finished items (e.g. *Black Slang* is not suitable if you use items such as “Black Slang is slang”, or *African American English* is problematic when used together with the item “African American English is a form of English”.).
- And is, ideally, short enough to allow for elegant and simple wording of items.

Concerning the first criterion, I based my decision on the side study on usage/recognition of synonyms for AAVE by teachers b discussed in section 2.1.3. I include all terms that were recognized by at least 50% of the sample of 87 U.S. teachers, teaching assistants and students undergoing teacher training (cf. Buendgens-Kosten 2006). For a list, see Table 5.1.

Concerning criterion two, I first deleted those terms from the list that can refer to more varieties of English than AAVE, and those that refer to only

a subgroup of what is subsumed by *AAVE*. For the most part, this was an intuitive choice. Only for *Ebonics* did I back this decision up with empirical data¹. I deleted the following terms: *Broken English*, *Ebonics*, *Non-Standard English*, *Slang*, *Street Slang*.

The next criterion, inoffensiveness of terms, leads to the deletion of all terms including *Black*. *Black* is not deemed an offensive term by everybody, but since *African American* seems to be considered more neutral, I preferred this term. *Black English*, *Black English Vernacular*, *Black Speech*, *Vernacular Black English* were discarded.

Broken English, *Slang*, *Street Slang* are highly evaluative terms (*slang* is not an evaluative term for linguists, but it is for non-linguists), and were therefore likewise discarded. The aforementioned study on the use of *Ebonics* by laypeople also raised the question whether *Ebonics* has, for many individuals, highly negative connotations. It was not, though, discarded on this reason alone.

The next criterion had no effect on choice of labels, since items were chosen in such a way that tautological (or contradictory) questions such as “Black slang is slang” or “African American English is not a form of English” were avoided.

The only labels that were not discarded during this process were *African American English* and *African American Vernacular English*. Of these, the first one is slightly shorter, which is a small advantage over the alternative *African American Vernacular English*.

I decided to favor *African American English* over *African American Vernacular English*, first, because it was recognized slightly more often by teachers, and secondly, because the absence of *vernacular*, a term defined in varying fashions by linguists which may be more confusing than elucidating for some laypeople, was considered an advantage.

It was decided, though, that even though *African American English* was to be the term used for the actual items, the questionnaire would receive a short introduction referring to differing language use, and helping individuals who might use a different term themselves to interpret *African American English* correctly.

Using such a short explanatory text is a double-edged sword. In this study, we wish to study the influence of contact with linguistics on pre-existing attitudes. Theoretically it is possible that individuals receive this questionnaire who did not previously identify an entity like “AAVE”, and to whom this

¹I am concerned that in the non-linguistic community, *Ebonics* is used almost exclusively to refer to an AAVE-based slang (cf. Buendgens-Kosten 2007), and less often used to refer to the full range of AAVE registers.

concept is entirely new. They might understand the explanation, and they may form an attitude concerning this new attitude object instantaneously. In such a case, we would not be measuring pre-existing attitudes, but creating attitudes in our subjects².

Barnes 1998 (discussed before in Section 2.1.2) found that in a sample of Southern U.S. students, more than 90 percent of subjects were aware of the existence of “Ebonics”. Only nine out of 420 students asked were not aware in any way whatsoever of “Ebonics” (cf. Barnes 1998 12). These numbers may be slightly lower today because of the time passed since the “Oakland controversy”, but, for the absolute majority of subjects studied here, the mere awareness of an entity “Ebonics”, “Black English”, “AAVE”, etc. can safely be assumed to be present; the data in Table 5.1 also supports this assumption. Therefore, the risk of creating the attitude object by asking for attitudes toward it is extremely low in this specific context. The advantage, i.e. giving subjects who customarily use another term to refer to AAVE the certainty that they have interpreted the label used in the questionnaire correctly, is believed to outweigh this risk.

5.1.1.2 Constructing the scale

5.1.1.2.1 Selection of items Thurstone compiled a list of five criteria that potential items have to fulfill in order to be considered for a Thurstone scale:

1. “The statements should be as brief as possible (...)” (Thurstone and Chave 1929, 22)
2. “The statements should be such that they can be endorsed or rejected in accordance with their agreement or disagreement with the attitude of the reader.” (Thurstone and Chave 1929, 22)

²Of course, this problem is irrelevant for researchers with a strong constructionist view: “According to this perspective, people often lack preconsolidated general evaluations. When asked to report attitudes, people consider readily available information and integrate this information into an overall attitudinal judgement.” (Fabrigar et al. 2005, 80). The only difference from a constructionist point of view would be that there is no pre-existing knowledge structure for this newly introduced idea. While the concern discussed here is irrelevant for representatives of a constructionist view of attitudes, my solution will not create any problems for this approach, while avoiding problems if the more traditional approach to attitudes holds.

3. “Every statement should be such that acceptance or rejection of the statement does indicate something regarding the reader’s attitude about the issue in question.” (Thurstone and Chave 1929, 22)
4. “Double-barreled statements should be avoided (...).” (Thurstone and Chave 1929, 23)
5. “One must insure that at least a fair majority of the statements really belong on the attitude variable that is to be measured.” (Thurstone and Chave 1929, 23)

Additionally, there is one more characteristic potential items must have that differentiates them e.g. from those suitable for other types of scales. In a Thurstone scale, items of different scale value are needed, that, if arranged in the order of degree of positiveness or negativeness, should ideally look like a string of beads: You have one extremely negative item, one item that is very negative, but slightly less negative than the most negative one, then you have an item that is still quite negative, but not as negative as the very negative one, etc. Likert scales (Likert’s Simple method, that is), for example, need the opposite type of items: items with a similar or, ideally, identical neutral point and comparable extremity of the different other answer options.

This characteristic of potential Thurstone items is one of the main reason why items from other scales can rarely be ‘recycled’ for Thurstone scales. Additionally, it is one of the reasons why the collection of items from only one source, e.g. only interviews or only books on this topic, is problematic. While moderately extreme opinions are regularly expressed, truly extreme opinions are expressed less often (this is especially true for formal or semi-formal situations, such as interviews). Neutral attitudes may be expressed in casual conversations, but are rarely quoted in books, articles, or other media. The collection of items of the neutral part of the attitude spectrum proved to be so difficult at times, that Thurstone explicitly permits violating the rule to avoid double-barreled statements if this violation yields neutral items (cf. Thurstone and Chave 1929, 23).

Potential sources for Thurstone scale items are:

1. Interviews/questionnaires eliciting opinion statements (cf. Thurstone and Chave 1929, 22).
2. Search of current (non-linguistic) literature (cf. Thurstone and Chave 1929, 22). This might be e.g. newspaper articles about AAVE, posts on online bulletin boards or blogs, etc.

3. Linguistic literature.
4. To a very limited degree: ‘recycling’ older questionnaires.
5. Intuitive choice.

The first two options seem to be the preferable ones, but for parts of the attitude continuum that are difficult to cover otherwise, the other three methods might also prove helpful additions.

The items used in this study were, in the main, extracted from an internet source: UrbanDictionary.com.³ UrbanDictionary is an interactive website where users define – often in a humorous way – terms they consider to be ‘urban’. It contains a wide variety of slang terms and common usage terms defined from an ‘urban’ perspective. All definitions available for “Ebonics” at January 2007 (a total of 70 definitions – two definitions appearing twice in the data set, two definitions discussing two separate meanings each) were reviewed in order to extract suitable opinion statements of presumably different scale value. The informal style used at UrbanDictionary.com, while making it an ideal source for more extreme items, made major rephrasing necessary.

The majority of belief statements gathered from this source (presumably) belongs to the AAVE-critical part of the attitude continuum, therefore, further material was sought to guarantee a more balanced selection of opinion statements for scaling. Opinion statements reported in linguistic literature discussing the attitudes and beliefs of laypeople⁴ were collected; the opinion statements were extracted and tested for suitability for Thurstone scaling. Again, statements were rephrased.

A list of potential items had been compiled by collecting items from previous studies (mostly Likert- and Osgood-scales)⁵, but in the end, none of the items

³Definitions were downloaded at 1/25/2007, from the following URLs: www.urbandictionary.com/define.php?term=ebonics; www.urbandictionary.com/define.php?term=ebonics&page2; www.urbandictionary.com/define.php?term=ebonics&page3; www.urbandictionary.com/define.php?term=ebonics&page4; www.urbandictionary.com/define.php?term=ebonics&page5; www.urbandictionary.com/define.php?term=ebonics&page6; www.urbandictionary.com/define.php?term=ebonics&page7; www.urbandictionary.com/define.php?term=ebonics&page8; www.urbandictionary.com/define.php?term=ebonics&page9; www.urbandictionary.com/define.php?term=ebonics&page10.

⁴The following publications supplied items included in the finished scale: Shuy 1970, Lippi-Green 1997, Lippi-Green 2000, Green 2002, Niedzielski and Preston 2003

⁵Items from the following studies were taken into consideration: Blake and Cutler 2003 (cf. Hoover et al. 1996; Taylor 1973); Bowie and Bond 1994; Cecil 1988; Covington 1975;

thus collected were included in the final version of the questionnaire. The selected process of scaling selected made it necessary to strictly limit the number of items scaled. The total number subjected to scaling was 13. The items so selected, including their sources, are listed in Table 5.3.

5.1.1.2.2 Paired comparisons A program was written by the author with which the items could be matched into pairs. In order to avoid order effects, the ‘orientation’ of item pairs (e.g. whether you ask a judge to compare item X with item Y, or item Y with item X) was randomized, using both orientations equally (or nearly equally in the case of odd numbers of pairs) often. Additionally, the item pairs were shuffled (i.e. randomized), so that “item 1 vs item 2” was not always followed by “item 1 vs. items 3”. For each judge, a newly randomized combination of item pair orientations and item pair positions was used, to further minimize the danger of group-wide order effects⁶.

A second tool, developed by Robert Kosten, was used to automatically transfer the randomized item pairs into questionnaires, using an HTML/PHP site and a MySQL database. An example of such a questionnaire can be found in the appendix.

Volunteers for this questionnaire were recruited at www.livejournal.com,⁷ and at various online bulletin boards (forums.atozteacherstuff.com, www.teacherfocus.com, www.proteacher.net, www.theteacherscorner.net). Only teachers (or individuals employed by a school whose job description included teaching, i.e. substitute teachers and some teaching assistants) resident in the United States were chosen as subjects for the scaling process. Figure 5.1 shows the flowchart of the online questionnaire. Subjects were offered an incentive of a \$2 donation on their behalf to one of three charities.

5.1.1.2.3 Scale construction The data collected was analyzed according to Allen Edwards’s description of case V of the law of comparative judgment. Case V makes certain plausible (though unproven) assumptions, e.g. that all standard deviations are equal (cf. Edwards 1957, 27). I have chosen

Robinson 1996; Taylor 1973; White et al. 1998; Abrams and Hogg 1987; Williams et al. 1976

⁶This program is available, under the GNU General Public Licence, from www.judith.buendgens-kosten.de/e_downloads.php.

⁷Advertising for the questionnaire was posted in the following interest groups: 1st_yr_teachers, 4thgrade, arted, art_teachers, educators, disabledteacher, elem_ed, fl_teachers, frenchteacher, history_teacher, history_teachers, lesson_plans, math_teachers, mdlshltchr, music_education, ne_teachers, real_teaching, scienceteachers, substitutes, teach_electives, teachergoths, teaching_math, teaching, teaching_sp_ed, urban_teachers

No	Item	Item source
1	African American English is broken English.	"Broken english used by African-Americans (ebo-americans?) and white wanna-be gen-X-ers." (Urban Dictionary, Definition 18)
2	Speaking African American English is an indicator for laziness.	"When backed into a corner, you can always claim that it has something to do with a sort of symbolism or is a defining trait that makes your race great, versus own up to the fact that it is essentially laziness at it's<sic> finest." (Urban Dictionary, Definition 1); "Has a relation to poor education but leans more towards laziness, if educated" (Urban Dictionary, Definition 42)
3	Speakers of African American English do not express complete thoughts.	"They don't realize that they aren't making a complete thought." (Shuy 1970)
4	African American English is not cool.	"Simply put,ebonics is not cool, at all." (Urban Dictionary, Definition 26)
5	African American English has very little grammar.	"It is generally spoken in the " hood" (neighborhood) and has almost no defined syntactical structure. Also of note is the almost complete lack of conjugation of verbs ("I be", "she be", "thems be", etc) and the mixing of pronouns. (Urban Dicionary, Definition 5)
6	Speaking African American English can cause difficulties on the job market.	"In corporate America, if you want to put an extra burden, a yoke on your neck, then speak slang, speak incorrect English and grammar, because youre not going to get the job."; "speaking correctly is an indication (...) to the person who is going to hire you that perhaps maybe you can do the job." (Lippi-Green 1997)
7	African American English is a perfectly normal thing in American media.	"An evolving form of American English popularized in multimedia." (Urban Dictionary, Definition 17); "White America use black dialect on commercials every day." (Lippi-Green 1997)
8	African American English differs from other forms of speech.	"they have a different slang than white people do" (Niedzielski and Preston 2003)
9	African American English is very structured.	"And that was the point at which everybody suddenly discovered that Black English actually is a very structured dialect of its own. That it has grammatical patterns." (Niedzielski and Preston 2003)
10	African American English has a place at the home of its speakers.	"But where does it have a place then." "Hey at home." (Niedzielski and Preston 2003)
11	African American English is used to strengthen a common identity between its speakers.	"So we gotta have our survival mechanism within our community. And our language is it. It lets us know that we all in this thing together" (Lippi-Green 2000)
12	It would be a terrible thing to lose African American English.	"The worst of all possible things that could happen would be to lose that language." (Lippi-Green 1997)
13	African American English can add flavour and soul to a poem, song or novel.	"Regardless of the 'genuineness' of the dialect, regardless of how remarkably it may add flavor and soul to a poem or song or novel"(Green 2002)

Table 5.3: List of items selected for scaling

Figure 5.1: Flowchart of the paired comparisons questionnaire

case V because only case V provides a method of analysing realistic (i.e. not textbook-perfect) datasets efficiently. After the actual paired comparisons were conducted, an F-matrix was assembled. An F-matrix indicates with which frequency each column stimulus was judged more favorably than the row stimulus. In a second step, a P-matrix was created by multiplying the cell entries of the F-matrix with the reciprocal of N (=number of judges). In a third step, a Z-matrix was created. Each value in the P-matrix was looked up in a table of normal deviates for normal distributions contained in the appendix of Edwards 1957 to determine the Z-values for the P-matrix entries. In principle, the arithmetic mean of all entries in each column of the Z-matrix constitute the scale values of the item represented by this column (cf. Edwards 1957, 31-37). If the Z-matrix is incomplete, which is usually the case, a slightly different approach has to be used.

Thurstone's technique of scale construction is based on the fact that sometimes people err about which stimulus is more positive than the other. If you hold two stones, one stone being 500 g (stone A), the other being 550 g (stone B), and try to estimate which one is heavier, you might judge correctly or incorrectly. If you ask N judges to estimate which one is heavier, it is likely that more individuals judge the second one as heavier, though not all. If you give the same number of judges two stones, one weighting 500 g and the other 600 g (stone C), the ratio of judges who judge correctly will grow larger. If you learn that, in the first case, 40% think that A is heavier than B, and 60% that B is heavier than A, and, in the second case, that 30% think that A is heavier and 70% that C is heavier, you not only learn that B is heavier than A and C is heavier than A, the difference in proportion of judges that decided one way or the other tells you that the difference between A and C is larger than between A and B, and that, therefore, C is heavier than B, which again is heavier than A. If you, though, use stones weighting 500 g and 1000 g, all judges will agree that the second stone is heavier than the first. Their reaction to the difference between 500 g and 1.000 g will be the same as to 500 g and 10,000 g: 100% agreement amongst judges. From this rating behavior, no conclusion can be drawn. The 1,000 g stone is heavier than the 500 g stone, and the 10,000 g stone is heavier than the 500 g stone, but whether the 1,000 g stone or the 10,000 g stone is heavier cannot be deduced from the estimations of judges. Edwards 1957 recommends that, with an N around 100, cell values below 0.02 and above 0.98 should be discarded; that is, you should allow for around 2% invalid judgments (due to e.g. uncooperative judges) in your data.⁸ Each time, therefore, that a very high or very

⁸According to Edwards 1957, with an $N=100$, cell values below 0.02 and above 0.98 should be discarded. With an $N=20$, I choose to discard all data below 0.1 and above 0.9,

low value appears in the F-matrix, it is discarded, and an incomplete matrix is the consequence. This is no sign of faulty item choice. It is merely the consequence of items of very different ‘weight’ (a necessary prerequisite for measuring attitudes in a community with potentially very diverse attitudes) having been included in the pool for scaling.

With incomplete data sets, you estimate missing cells based on the data provided by adjacent columns (cf. Edwards 1957, 40ff).

All matrices can be found in the appendix (6.3).

5.1.1.2.4 Internal consistency check I applied the internal consistency check suggested by Edwards 1957 (37ff.). Basically, I created a hypothetical P-matrix, a P’-matrix, based on the scale values I had calculated, and compared this theoretical P’-matrix with the P-matrix that served as a basis for calculating these scale values in the first place. First, a Z’-matrix, that is a matrix of theoretical Z-values, was calculated from the scale values. This Z’-matrix was then converted into a P’-matrix (using the same table that had been used to convert the P-matrix into the Z-matrix). In a third step, the cell values of the P’-matrix were subtracted from the corresponding cell values of the P-matrix. The arithmetic mean was calculated for each column of the resulting matrix, and a mean across columns was calculated (Z’ matrix, P’ matrix and matrix of differences between P and P’ have been included in the appendix (6.3).). A relatively high mean, i.e. a relatively low measure for internal consistency, was to be expected. A value of around 0.025 to 0.03, as recommended by Edwards, is difficult to reach if the smallest gradations in the data used are on the level of 0.05, as is the case with an N=20. The final mean was 0.23.

The internal consistency test (cf. Edwards 1957) yielded a higher than recommended value. The scale values constitute merely rough estimations of relative distance between items and should not be overinterpreted. The scaling is reliable enough to determine a rough ordering of items, but the exact order of items very close to each other on the attitude continuum may be open to debate.

5.1.1.2.5 The finished scale 5.5 contains an overview over the items and their scale values.

Items in order of scale value (from most positive, to most negative):

making allowances for one uncooperative judge in my sample of twenty judges.

No	Item	Scale value
1	African American English is broken English.	216
2	Speaking African American English is an indicator for laziness.	360
3	Speakers of African American English do not express complete thoughts.	239
4	African American English is not cool.	265
5	African American English has very little grammar.	190
6	Speaking African American English can cause difficulties on the job market.	174
7	African American English is a perfectly normal thing in American media.	49
8	African American English differs from other forms of speech.	125
9	African American English is very structured.	90
10	African American English has a place at the home of its speakers.	62
11	African American English is used to strengthen a common identity between its speakers.	0
12	It would be a terrible thing to lose African American English.	87
13	African American English can add flavour and soul to a poem, song or novel.	41

Table 5.5: Scale values of items

- 11: African American English is used to strengthen a common identity between its speakers.
- 13: African American English can add flavour and soul to a poem, song or novel.
- 7: African American English is a perfectly normal thing in American media.
- 10: African American English has a place at the home of its speakers.
- 12: It would be a terrible thing to lose African American English.
- 9: African American English is very structured.
- 8: African American English differs from other forms of speech.
- 6: Speaking African American English can cause difficulties on the job market.
- 5: African American English has very little grammar.
- 1: African American English is broken English.
- 4: African American English is not cool.
- 3: Speakers of African American English do not express complete thoughts.
- 2: Speaking African American English is an indicator for laziness.

This ordering of items based on scale values derived through Thurstone scaling agrees, for the most part, with an intuitive ordering of items. One surprising thing, perhaps, is that item 12 is only the 5th most positive item. If viewed together with its scale values, though, this makes more sense. Item 12 has a scale value of 87 on a scale from 1 (most positive) to 360 (most negative). It is only 12 scale points from item 10 (scale value 62), and 49 scale points from item 13 (scale value 41). Due to the relatively high number of positively evaluated items in this scale, item 12 may be perceived as further away from the most positively evaluated item than it really is when only looking at the relative ordering of items, not at the scale values of items. They are, as their scale values reveal, fairly close to each other, especially keeping the rather rough scaling done here in mind.

5.1.1.3 Additional elements of the questionnaire

The questionnaire consisted of a short introduction, a block where general demographic data was collected, a block where teaching-specific demographic data was collected, a self-assessment of contact with linguistics, the main part of the questionnaire, where subjects were asked to endorse or reject the items of the attitude scale, accompanied by a second short introduction, and a final block where individuals were asked whether they had perceived a relationship between contact with linguistics and language attitudes and were invited to make any comment they liked. A short test question checked whether the questionnaire had been manually completed, to exclude bot-generated replies.

On a second page, individuals could enter their e-mail to receive further information, and could indicate whether they might be contacted with further questions. The data derived through this second page was stored separately from that of the first page to protect subjects' anonymity.

On a third page, subjects were thanked for their participation, and their choice whether or not they would receive further information was confirmed (see 6.3 for details).

5.1.1.3.1 Self-evaluation of contact with linguistics It was decided that an objective measurement of linguistic knowledge was not feasible. A person might have had a lot of contact with one linguistic discipline (e.g. second language acquisition/psycholinguistics), without any exposure to other branches of linguistics. Therefore, a test of linguistic knowledge would have to cover a wide range of topics. Additionally, a person might have had significant exposure to certain topics, but might have difficulties to recall details of what was learned. Such a person might answer test questions incorrectly, or be unsure whether he/she really remembers a certain term, but might still have experienced attitude change through the exposure to linguistics he/she had. Instead of 'measuring' the gain in content knowledge, I have limited myself to measuring intensity of exposure.

Self-assessment took place along two dimensions. The first dimension addresses intensity of contact. A guided format with five gradations was used: Contact with linguistics

- I have never had contact with linguistics in any form.
- I have had a little contact with linguistics (e.g. read an article in a journal, watched a documentary)

- I have had some contact with linguistics (e.g. attended at least one seminar on linguistics at university or attended a teacher training workshop, read one or more books on linguistics)
- I have had a lot contact with linguistics (e.g. attended several seminars on linguistics at university, etc.)
- I am a linguist (e.g. degree in linguistics)

The second dimension is one of context of exposure:

Overall, I had contact with linguistics

- in my freetime (hobbies and entertainment)
- through my job (e.g. on-the-job training, books and journals for teachers)
- during university training

Subjects could, of course, have been exposed to linguistics through a variety of means, therefore they were able to choose more than one option here.

5.1.1.4 Technical characteristics of the questionnaire

The frontend of the questionnaire consists of a HTML/PHP site. Data was stored in a MySQL database.

5.1.2 Conducting the study

Volunteers were recruited through blogs, online bulletin boards and Yahoo groups catering to American teachers. A certain number of school principals were contacted with information on the research project and asked to distribute information to teachers at their school. Two volunteers were obtained through personal contacts.

5.1.2.1 Subjects

A total of 193 responses to the questionnaire were collected. From this collection, I excluded responses based on the following criteria:

1. If, based on the time of posting and the content of the response, a double posting could be assumed, the older version was deleted. Subject 77, Subject 94

2. Individuals who had teaching experience only at university or only at private teaching institutions that did not constitute schools (e.g. private music teachers) were excluded.
college: Subject 23, Subject 39, Subject 44, Subject 86, Subject 152
non-school: Subject 14, Subject 36, Subject 49, Subject 52, Subject 65
3. Individuals with purely administrative functions who do not nor ever did teach, were excluded. Individuals undergoing teacher training were retained, as they are, in a way, future teachers. The questionnaires from retired or unemployed teachers were retained for similar reasons.
administration: Subject 52, Subject 129, Subject 142, Subject 181
unset: Subject 111, Subject 159
4. Responses marked as “test” were deleted. Subject 111
5. Responses not indicating a state were excluded. Subject 4, Subject 159
6. Individuals who did not answer the ‘bot test’ question, or answered it incorrectly, were checked ‘manually’ for bot-influence. They were obviously human-made and were retained in the sample.
7. Individuals who did not reply to the “understood” question, i.e. to the question that followed the short explanation of what is meant by “African American English” in this questionnaire, asking whether the person understood the term, or answered they did not understand it, were excluded from the sample. Subject 92, Subject 144, Subject 155

171 questionnaires were retained in the sample.

5.1.2.1.1 Make-up of the sample 154 (90.0%) subjects are female, 17 (9.9%) are male.

Volunteers were encouraged to describe their ethnicity⁹ in a free format. No rubrik was imposed on them. I did, though, summarize subjects’ ethnic self-identifications according to a simplified version of the U.S. Census format for ethnicity/race outside of Hawaii (cf. U.S. Census Bureau nda). The majority of subjects (87.2%) are of European American origin. For details, see Table

⁹I do not follow the U.S. Census’ distinction between the notions of *ethnicity* and *race* (cf. U.S. Census Bureau 2001), and use *ethnicity* in its widest possible sense.

5.7¹⁰.

	White non- Hispanic	Hispanic	Black non- Hispanic	AIAN non- Hispanic	Asian non- Hispanic	NHOPI non- Hispanic	Other/ Mixed	unset
Total	141	3	7	1	4	0	4	10
%	87.6	1.9	4.4	0.6	3.1	0	2.5	-

Table 5.7: Sample: Ethnicity

Subjects came from 37 different U.S. states: 2x AK, 3x AL, 6x AZ, 16x CA, 1x CO, 1x CT, 4x FL, 10x GA, 9x IL, 6x IN, 1x KS, 3x KY, 1x LA, 4x MA, 3x MD, 8x ME, 8x MI, 4x MN, 2x MO, 1x MS, 1x MT, 4x NC, 2x NE, 2x NH, 8x NJ, 6x NV, 11x NY, 4x OH, 2x OR, 9x PA, 1x RI, 1x SC, 11x TX, 2x UT, 7x VA, 6x WA, 1x WI.

All Census Regions and Divisions are reflected in this sample (see Table 5.8 and the accompanying map 5.2¹¹).

West 42		Midwest 37		South 48			Northeast 44	
Pacific	Moun- tain	West North Cen- tral	East North Cen- tral	West South Cen- tral	East South Cen- tral	South At- lantic	Middle At- lantic	New Eng- land
26	16	9	28	12	7	29	20	24

Table 5.8: Sample: Census Regions and Divisions

The average age is 36.6 years, with ages ranging from 21 to 65.

The younger teacher population is more strongly represented in the sample than the older one. 72 subjects are in their 20's, 35 in their 30's, 28 in their 40's, 30 in their 50's, 5 in their 60's. One person did not indicate his/her age.

Subjects had an average of 9.5 years of teaching experience, ranging from 0 to 40 years. Ten subjects had no teaching experience.

¹⁰The categories for ethnic origin are a simplified version of those used by the U.S. Census (cf. U.S. Census Bureau nda). *NHOPI* stands for 'Native Hawaiian and Other Pacific Islander', *AIAN* for 'American Indian and Alaska Native'.

¹¹This map is based on an image created by the U.S. Census Bureau (U.S. Department of Commerce Economics and Statistics Administration U.S. Census Bureau, Geography Division nd).

Figure 5.2: Sample: Census Regions and Divisions

All types of schools were covered. 6 subjects worked at/had worked at/were undergoing teacher training for kindergarten, 54 at/for elementary school, 41 at/for middle school, 5 at/for junior high, 42 at/for high school, and 23 at other forms of schools (including combinations of the above mentioned). 17 of these subjects had a special education professional background. 143 were currently employed as teachers, 5 were retired or unemployed teachers, and 10 were undergoing teacher training. 10 were working in school administration, but either taught or had taught in the classroom as well. One subject worked as an assistant teacher. Two subjects did not answer the question concerning their current job status, but gave sufficient information in other parts of the questionnaire to determine that they fulfilled all conditions necessary for being retained in the sample.

More important than the question of how many men and how many women, or how many African Americans and how many European Americans answered the questionnaire is the question in how far this sample is representative of the American teacher population as a whole.

Data concerning the teacher population was collected from the U.S. Census of 2000 (cf. U.S. Census Bureau *nda*). I combined the data of four different Census Occupation Codes: “Elementary and Middle School Teachers”, “Secondary School Teachers”, “Special Education Teachers”, and “Teaching Assistants”. I did not include “Preschool or Kindergarten Teachers” and “Education Administrators”. Both groups intersect with my sample. Kindergarten teachers were included, preschool teachers were not. Education administrators that actively teach or taught at a school were included, administrators without current or former teaching duties were not. Students undergoing teacher training and retired teachers are included in my sample, but not in the data from the U.S. Census.

Occupation	Male	Female
Elementary and Middle School Teachers	656,145	2,469,180
Secondary School Teachers	318,685	453,775
Special Education Teachers	23,480	151,715
Teacher Assistants	77,315	838,605
Total	1075625	3,913,275
Percentage	21.6	78.4

Table 5.9: Teacher population: Sex

Of course, the sample here does not perfectly reflect the make-up of the teacher population as a whole. The number of women in my sample (90.0%)

Occup.	Total	White non- Hispanic	Hispanic	Black non- Hispanic	AIAN non- Hispanic	Asian non- Hispanic	NHOPI non- Hispanic	Other/ Mixed
Elem. & Middle School T.	3,125,320	2,572,850	171,655	283,600	14,440	47,845	1,930	33,005
Secondary School T.	772,460	664,025	37,035	48,215	3,165	10,625	505	8,890
Teacher Assis- tants	915,915	625,700	123,195	123,185	9,365	20,440	1,100	12,935
Special Ed.	175,190	151,060	7,155	12,125	980	1,775	210	1,890
Total	4,988,885	4,013,635	339,040	467,125	27,950	80,685	3,745	56,720
Percentage	100	80.5	6.8	9.4	0.6	1.6	0.1	1.1

Table 5.10: Teacher population: Ethnicity (simplified)

is slightly higher than in the U.S. Census (78.4%) (see also Table 5.9), as is the number of White non-Hispanic subjects (87.6% as compared with 80.5%) (see also Table 5.10). For the size of the sample, though, as well as for the sampling method used, the sample reflects the overall population fairly well. In addition to sex and ethnicity, it reflects all Census Districts and Regions. See Table 5.11 for an overview on state level (Census data vs. sample data). It must be assumed that the medium chosen, internet-based questionnaires, influenced the age structure of the sample. According to the U.S. Census, 17% of all public school, and 19% of all private school teachers are under the age of 30 (cf. U.S. Census Bureau ndb). In my sample, which also included students undergoing teacher training, 42.4% of subjects fell into that age group. Younger individuals responded to the questionnaire much more often than older individuals. Still, the relevant age group has been completely covered. The distortion in the age range that was to be expected can be observed, but it is less extreme than might have been feared.

Comparing make-up of sample according to school type is slightly more difficult, since the categories used here and those used by the U.S. Census differ significantly. First, the U.S. Census treats special education teachers and assistant teachers as an occupation independent from school type. It also subsumed elementary school and middle school into one category, and has only one category for high school teachers. Kindergarten teachers, which I treated as one category, are subsumed in one category together with

State	Census	Percentage (Census)	Sample	Percentage (Sample)
Alabama	71560	1.43	3	1.75
Alaska	14280	0.29	2	1.17
Arizona	78610	1.58	6	3.51
Arkansas	43485	0.87	0	0
California	554160	11.11	16	9.36
Colorado	72845	1.46	1	0.58
Connecticut	72515	1.45	1	0.58
Delaware	13280	0.27	0	0
District of Columbia	7800	0.16	0	0
Florida	226695	4.54	4	2.34
Georgia	149305	2.99	10	5.85
Hawaii	22335	0.45	0	0
Idaho	23295	0.47	0	0
Illinois	216340	4.34	9	5.26
Indiana	99755	2	6	3.51
Iowa	55315	1.11	0	0
Kansas	58095	1.16	1	0.58
Kentucky	68080	1.36	3	1.75
Louisiana	81710	1.64	1	0.58
Maine	27265	0.55	8	4.68
Maryland	97425	1.95	3	1.75
Massachusetts	127930	2.56	4	2.34
Michigan	169205	3.39	8	4.68
Minnesota	97915	1.96	4	2.34
Mississippi	51155	1.03	1	0.58
Missouri	97290	1.95	2	1.17
Montana	18615	0.37	1	0.58
Nebraska	33445	0.67	2	1.17
Nevada	26260	0.53	6	3.51
New Hampshire	25930	0.52	2	1.17
New Jersey	173095	3.47	8	4.68
New Mexico	35265	0.71	0	0
New York	398845	7.99	11	6.43
North Carolina	140545	2.82	4	2.34
North Dakota	13105	0.26	0	0
Ohio	179800	3.6	4	2.34
Oklahoma	59180	1.19	0	0
Oregon	57500	1.15	2	1.17
Pennsylvania	201475	4.04	9	5.26
Rhode Island	20755	0.42	1	0.58
South Carolina	68905	1.38	1	0.58
South Dakota	15505	0.31	0	0
Tennessee	89860	1.8	0	0
Texas	411935	8.26	11	6.43
Utah	36990	0.74	2	1.17
Vermont	15845	0.32	0	0
Virginia	130320	2.61	7	4.09
Washington	95850	1.92	6	3.51
West Virginia	29870	0.6	0	0
Wisconsin	101390	2.03	1	0.58
Wyoming	10940	0.22	0	0
	4988870	100	171	100

Table 5.11: Teacher population: States

Occupation	Number	Percentage
Elementary and Middle School Teachers	3.125.325	80.18%
Secondary School Teachers	772.460	19.82%

Table 5.12: Teacher population: School type

preschool teachers. Again, the U.S. Census does not include students undergoing teacher training or retired teachers. A comparison of my data with that of the U.S. Census is therefore extremely rough. Keeping these problems in mind: My sample contains 98 (66.9% as compared to 80.18% in the U.S. Census) subjects with a background in elementary or middle schools, and 47 (33.1% as compared to 19.82% in the U.S. Census) subjects with a high school background (leaving all additional categories aside) (see also Table 5.12).

5.2 Findings

In this section, I will first discuss the general findings concerning the attitude scale, then those of the self-assessment of contact with linguistics, and finally, the interrelationship between these two measures.

5.2.1 Overall attitudes

Table 5.13 shows the frequency with which items were endorsed or rejected by the subjects. Table 5.14 shows the same data, arranged in order of item scale value.

Item	1	2	3	4	5	6	7	8	9	10	11	12	13
Agree	68	25	141	153	72	139	63	68	77	73	163	31	110
Agree (%)	39.8	14.6	82.5	89.5	42.1	81.3	36.8	39.8	45	42.7	95.3	18.1	64.3
Disagree	94	144	29	16	96	26	105	97	91	92	7	138	57
Disagree (%)	55	84.2	17	9.4	56.1	15.2	61.4	56.7	53.2	53.8	4.1	80.7	33.3
Unset	9	2	1	2	3	6	3	6	3	6	1	2	4
Unset (%)	5.3	1.2	0.6	1.2	1.8	3.5	1.8	3.5	1.8	3.5	0.6	1.2	2.3

Table 5.13: Agreement and disagreement with items

Item	11	13	7	10	12	9	8	6	5	1	3	4	2
Scale value	0	41	49	62	87	90	125	174	190	216	239	265	360
Agree	163	110	63	73	31	77	68	139	72	68	141	153	25
Agree (%)	95.3	64.3	36.8	42.7	18.1	45	39.8	81.3	42.1	39.8	82.5	89.5	14.6
Disagree	7	57	105	92	138	91	97	26	96	94	29	16	144
Disagree (%)	4.1	33.3	61.4	53.8	80.7	53.2	56.7	15.2	56.1	55	17	9.4	84.2
Unset	1	4	3	6	2	3	6	6	3	9	1	2	2
Unset (%)	0.6	2.3	1.8	3.5	1.2	1.8	3.5	3.5	1.8	5.3	0.6	1.2	1.2

Table 5.14: Agreement and disagreement with items in order of scale value

More than 80% of subjects believe that AAVE was not cool, that speakers of AAVE do not express complete thoughts, that speaking AAVE could cause difficulties on the job market. At the same time, more than 90% of subjects also believe that AAVE is used to strengthen a common identity between its speakers.

Endorsement is usually more informative than non-endorsement. If a person endorses “I like AAVE”, we know that this person likes AAVE. If he/she does not endorse this item, he/she might adore AAVE and consider ‘like’ too weak a term, or dislike AAVE. In many instances, a person may also consider a wording infelicitous, and decide not to endorse an item that corresponds, or nearly corresponds, to his/her attitude. The existence of ‘gaps’ within a range of accepted items is therefore nothing surprising.

5.2.1.1 Latitude of acceptance

Many subjects demonstrated an extremely high latitude of acceptance¹². I define *Latitude of acceptance* as the breadth of items on a (at least ordinal scale level) scale agreed with by an individual. An extremely high latitude means that subjects agree with items that are relatively far apart on the attitude continuum, in several cases endorsing the complete range from scale value 0 to 360.

See Figure 5.3 for an overview of the endorsements of some subjects. Each

¹²The notion of ‘latitude of acceptance’ originally comes from social judgement theory (cf. Eagly and Chaiken 1993, 93ff). I use it here without committing to other theoretical notions of this theory.

0	41		62			125	174			239	265	
0		49	62			125	174			239	265	
0	41	49	62			125	174		216	239	265	
		49	62		90	125	174				265	
0	41	49	62				174			239	265	
0	41		62				174			239	265	360
0	41		62			125	174		216	239	265	
0	41	49	62	87	90		174	190		239	265	360
0	41	49			90		174	190	216	239	265	
0	41				90					239	265	
0	41	49	62			125	174			239	265	
0	41			87	90		174	190	216	239	265	
0	41	49	62			125	174			239	265	
0		49	62			125	174		216	239	265	
0	41		62	87			174	190	216	239	265	360

Figure 5.3: Extract from data

row represents the responses of one subject. The latitude of acceptance is highlighted in grey.

Table 5.15 shows an overview of the latitude of acceptance found for the sample. I operationalized latitude of acceptance as scale value of the most negative item endorsed minus the scale value of the most positive item endorsed. Gaps between endorsed items were ignored. The highest possible latitude of acceptance score is 360.

Latitude	0-90	91-180	181-270	270-360
	1	1	146	23

Table 5.15: Latitude of acceptance

A certain latitude of acceptance is normal. A person with a positive attitude will endorse a number of items from the positive end of the attitude continuum, a person with an attitude between positive and negative will endorse

Figure 5.4: Latitude of acceptance

a number of items from the neutral middle of the attitude continuum. An extremely small latitude of acceptance is often an indicator for high involvement with an issue (cf. Eagly and Chaiken 1993, 153f, but also Eagly and Chaiken 1993, 379f), i.e. often a radical position. An extremely wide latitude of acceptance, on the other hand, leads to the acceptance of evaluatively contradictory items, i.e. a person with a very wide latitude of acceptance agrees with items whose evaluative dimensions contrast starkly (see Figure 5.4).

The larger the latitude of acceptance gets, the larger the probability for ambivalent attitudes becomes: If a scale spans items that represent positive evaluations, neutral evaluations and negative evaluations, and a subject agrees with the full breadth of the scale, this person endorses items with contradictory evaluative loading.

A set of data may only **appear** to have a wide latitude of acceptance if the items used for measuring attitude are very similar in their evaluative meaning. If the items presented cover only the range from “I like this product a little” to “I like this product very much”, individuals might endorse the complete measured range of attitudes without having an unusual high latitude of acceptance. Thurstone scaling can never cover a complete range of possible attitude-reflecting items, because the more extreme items become, the lower the possibility of deriving comparison data with a sufficient degree of disagreement between judges to calculate distance between items becomes (cf. discussion at Section 5.1.1.2.3). It does allow, though, for the creation of scales that, if endorsed in their entirety, would reflect a self-contradictory attitude.

Attitudes with a very large latitude of acceptance as observed in this study are referred to in social psychological research as *ambivalent attitudes*:

Attitudinal ambivalence occurs when there is evaluative tension associated with one’s attitude because the summary includes both positive and negative evaluations (...). (Fabrigar et al. 2005, 84)

Simply speaking, “(A)[a]ttitudinal ambivalence is the simultaneous existence of positive and negative beliefs or emotions with regard to the same object in an individual’s attitude base.” (Jonas and Ziegler 2007, 31). Priester and Petty 1996 explains ambivalence through a literary example:

Hamlet both longs for and at the same time fears his own death. Clearly, a “neutral” or “slightly positive” response form Hamlet toward suicide on a traditional bipolar attitude scale would lose a great deal of information concerning the conflict and indecision underlying his overall attitude. (Priester and Petty 1996, 431)

This can be illustrated by some of the freetext comments:

While I have an appreciation for African American English as a cultural identity thing, and while it can make literature more relatable and realistic, I am too much of a stickler for correct spelling and grammar in the English language to think that it has much of a place in society anymore. (...) (Subject 10)

This person holds both positively evaluated (“cultural identity thing”, “can make literature more relateable and realistic”) and negatively evaluated opinions (not much of a place in society anymore, not “correct”) towards the attitude object AAVE.

The concept of ambivalence has been part of the social psychological debate on attitudes since the 1960s and has been playing an important role in attitude research since the 1990s (cf. Jonas and Ziegler 2007, 31).

Ambivalent attitudes are associated with what Clark et al. 2008 dub “controversial and consequential issues” (565). The examples he gives are “capital punishment, legalized abortion, U.S. involvement in Iraq”. The debate surrounding, for example, the Oakland resolution may be taken as an indicator for the controversy surrounding this topic. Other researchers have used social and ethnic minorities (“stigmatized groups”) as their focus of research, African Americans being ‘bread and butter’ examples of attitude objects evoking ambivalent reactions, especially during the early years of research on ambivalence (e.g. Katz and Glass 1973, Haas and al. 1991). Since attitudes towards a language or language variety are usually heavily influenced by the attitudes towards the people using them, it would be plausible that if attitudes toward African Americans are ambivalent, that attitudes towards African American Vernacular English are also ambivalent.

5.2.1.2 Contact with linguistics

The majority of teachers and future teachers in this sample have had at least some contact with linguistics. Only 12.2% have had no contact with linguistics at all. (For details, see Table 5.16.)

Nearly half the subjects (46.8%) had contact with linguistics at university. Around one third (35.7%) had contact with linguistics through their jobs, e.g. through on-the-job training, books or journals specifically aimed at teachers. A surprisingly high number (29.8%) have had contact with linguistics in their freetime, during activities such as reading, watching TV, cultivating hobbies, etc.

intensity of contact	never	little	some	a lot	linguist	unset
totals	20	64	62	20	3	2
percentage	11.8	37.8	36.7	11.8	1.2	-

Table 5.16: Contact with linguistics

5.2.2 Contact with linguistics and effect on language attitudes

The fact that most subjects expressed ambivalent attitudes means that summarizing the results of attitude measurement in the form of “attitude scores” (i.e. the arithmetic mean of scale values of endorsed items), even if it were permissible by the scale level (which it is not in the case of interval Thurstone scales), is not possible. The attitude score of an individual with ambivalent attitudes and of an individual with an attitude within the neutral range may be identical. The similarity of arithmetic mean would cover the gross dissimilarities between the answer behavior of subjects. Instead, answer behavior between low-contact and high-contact groups was compared at the level of individual items. Difference between frequency of agreement between groups was tested using an undirected chi-square test with Yates’ continuity correction and a significance level of 5%. For those items with an expected cell value below 5, Fisher’s exact test was used, using the same significance level (cf. Bortz 2005, 169f). All calculations were performed using R: A Language and Environment for Statistical Computing.

First, I checked for significant differences between subjects with no contact with linguistics, and those with a lot of contact with linguistics. If contact with linguistics has an effect on language attitudes, this difference should

be most pronounced when comparing these two groups. In a second step, I collapsed those subjects who indicated having had no contact with linguistics and those who indicated having had little contact with linguistics into one group (“no contact group”), and those with some and a lot of contact with linguistics into another group (“contact group”). While the difference between these two groups might not be as pronounced as between the never contact and lots contact groups, the higher number of subjects in each group (84 no contact group, 62 contact group, as compared to 20 never group, 20 lots group) makes it easier for them to yield significant results.

In the comparison of the never contact with the lots contact groups, four items reached the significance level of 5%. For the comparison of the contact with the no contact groups, one additional item reached the significance level. (See Tables 5.17 and 5.19 for an overview over the chi-square results, and 5.20 and 5.20 for an overview over the Fisher’s exact test results. Significant results are marked by italics.)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13
Scale value	216	360	239	265	190	174	49	125	90	62	0	87	41
X-squared	2.6389	-	1.2	-	6.8267	-	0.989	3.3933	12.6042	12.1	-	-	0
p-value	0.1043	-	0.2733	-	0.009	-	0.32	0.0655	0.0004	0.0005	-	-	1

Table 5.17: Chi-square results never contact vs. lots contact group (rounded)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13
Scale value	216	360	239	265	190	174	49	125	90	62	0	87	41
p-value	-	0.3416	-	0.2308	-	0.3436	-	-	-	-	1	0.0436	-

Table 5.18: Fisher’s exact test results never contact vs. lots contact group

Item	1	2	3	4	5	6	7	8	9	10	11	12	13
Scale value	216	360	239	265	190	174	49	125	90	62	0	87	41
X-squared	1.4416	1.0158	0.5569	3.1164	12.7538	1.3713	0.68	9.023	12.4415	14.1553	-	9.4678	0.0027
p-value	0.2299	0.3135	0.4555	0.0775	0.0004	0.2416	0.4096	0.0027	0.0004	0.0002	-	0.0021	0.9584

Table 5.19: Chi-square results contact vs. no contact group (rounded)

This shows that individuals who had contact with linguistics differ in answer behavior from those who did not¹³. It is interesting to look at the kind of

¹³The observed difference between these groups is, of course, no evidence that contact with linguistics is the cause of that difference. While this interpretation is likely, other factors might be at play here besides contact with linguistics. The self-perception of the

Item	1	2	3	4	5	6	7	8	9	10	11	12	13
Scale value	216	360	239	265	190	174	49	125	90	62	0	87	41
p-value	-	-	-	-	-	-	-	-	-	-	0.6819	-	-

Table 5.20: Fisher’s exact test results contact vs. no contact group

Figure 5.5: Types of attitude change

differences observed, and at how they would relate to an assumed attitude change through contact with linguistics.

Figure 5.5 illustrates two ways in which the answer behavior of a person may change. The line stands for the attitude continuum, the beads stand for those points of the attitude continuum represented by a specific item (in ‘real world’ questionnaires, these items would of course not be as evenly spaced as in this illustration). The arrows stand for change. Arrows pointing up mean endorsement where items were not previously endorsed, arrows pointing down mean items are not endorsed which were previously endorsed. If you assume that positive items lie on the left side of the continuum and negatively worded items lie on the right side of the continuum, then change type 1 means an improvement of attitudes. A person endorses more positive items and less negative items. Such a systematic change in answer behavior can be interpreted as indicator for general attitude change.

The situation looks different for change type 2. The person endorses more positive items, but also more negative items. The person has furthermore stopped to endorse some positive items, and also stopped to endorse some negative items. There has been change, but the change cannot be summarized as a “more positive” or a “more negative” attitude toward the attitude

influence of linguistics on opinions (Section 5.2.2.1) and the linguistic epiphanies (Section 5.2.2.2) offer some support for the interpretation that linguistics is at least one of the factors here, though.

object.

Looking at the differences between the contact and the no-contact group, a pattern similar to that of change type 2 can be observed, while the differences between the never contact and lots contact group seem to resemble type 1. This requires further explanation.

While it is easy to determine which is the “positive” and which the “negative” side of the continuum, settling on a “neutral” point is not as easy. If a clear pattern of type 1 emerges, the point where increased agreement is replaced by lessened agreement (or vice versa) may be assumed to be a neutral point. This is not possible for change of type 2, and Thurstone scaling in its traditional form, as done here, yields interval scales, i.e. does not come with a natural zero point.

If the differences between the never contact / lots contact groups form a pattern of type 1, then the neutral point should lie somewhere between the scale values of 62 and 87, i.e. **agreement** with “African American English has a place at the home of its speakers.” (scale value 62) would indicate a positive attitude, as does **disagreement** with “It would be a terrible thing to lose African American English.” (scale value 87). This is counterintuitive. Scaling only served to establish the relative ordering and the distances between items; therefore, scaling itself does not contradict this analysis. On the other hand, it is difficult to accept that disagreement with items such as “It would be a terrible thing to lose African American English.” (scale value 87) or “African American English is very structured.” (scale value 90) express a positive attitude. If we assume that disagreement with item 12 does **not** express a positive attitude, then it must be assumed that we have change of type 2, not of type 1. This interpretation is supported by the results of the significance tests on differences between the contact and the no contact group. Here, the pattern of differences makes change of type 2 plausible (See Tables 5.21 and 5.22.).

5.2.2.1 Perception of attitude change

Subjects were asked about their own perception concerning their attitudes (or, more precisely, their opinions) toward AAVE and their contact with linguistics: “Do you believe that what you learned in linguistics changed your opinions regarding African American English?” Subjects were given four answer options: “Yes”, “No”, “Don’t know”, and “Not applicable” (for subjects without contact with linguistics).

50 subjects answered “yes”, 64 “no”, 26 “don’t know”, 30 “not applicable” and one person did not answer this question. Obviously, a large number of

Item	10	12	9	5
	“African American English has a place at the home of its speakers.”	“It would be a terrible thing to lose African American English.”	“African American English is very structured.”	“African American English has very little grammar.”
Scale value	62	87	90	190
Increase/ Decrease in agreement	inc.	dec.	dec.	dec.

Table 5.21: Significant difference in answer behavior lots contact vs. never contact group

Item	10	12	9	8	5
Scale value	62	87	90	125	190
	“African American English has a place at the home of its speakers.”	“It would be a terrible thing to lose African American English.”	“African American English is very structured.”	“African American English differs from other forms of speech.”	“African American English has very little grammar.”
Increase/ Decrease in agree- ment	inc.	dec.	dec.	inc.	dec.

Table 5.22: Significant difference in answer behavior contact vs. no contact group

subjects have the perception that their contact with linguistics did influence their opinions on AAVE¹⁴. This supports the observations made before: For at least 50 out of 117 subjects (42.7%), linguistics is likely to have had an impact on opinions on AAVE. This fact, though, does not solve the question of how this change is to be interpreted, as improvement of attitudes, or as a more complex form of change. Both interpretations remain possible.

5.2.2.2 Linguistic epiphanies

In the online questionnaire, the list of items was followed by a freetext field, into which subjects could enter any comments they wanted to make. Some used it to comment on the items, or to give more information on their biography. Three subjects discussed linguistics and what influence it had on their responses to the items.¹⁵ Two of those discuss something I call *linguistic epiphanies*:

I really wish more people were educated on this topic because reading up on it myself did open my eyes to Black English Vernacular as a “legitimate” and structured dialect. I do feel it has its limitations socially, as it is not the mainstream way of speaking in the United States. Therefore, if one interviews for a job using BEV, it will most likely hurt their chances of obtaining said job, in a similar way that speaking with a strong southern dialect might. I think an ability to code switch in order to fit into a variety of speaking situations is a very valuable skill for anyone to have and students should be allowed/encouraged to speak however they feel most comfortable with their friends, but also be taught and encouraged to use formal register in order to be their most effective in society at large. (Subject 55)

Coming from Alaska where African American English is spoken rarely, my perspectives have changed since reading several articles about the linguistics of African American English. (Subject 90)

¹⁴It is to be duly noted that the subjective perception of change does not necessarily reflect objective change, nor does the absence of such perception guarantee the absence of change.

¹⁵Actually, four subjects did, but one of them interprets *linguistics* as language use: “I have found that linguistics can change from town to town or from city to city even within the same region of the state. I am married to a man who is Indian, my sister is married to someone who is African-American and I see various differences in linguistics between the races and my own race.” (Subject 63).

For these two subjects, linguistics was perceived to have had a profound impact on their attitudes. It functioned as an “eye-opener” for them, which they happily acknowledge as the source of opinion change. The third person (Subject 97) discusses the opinions on AAVE of a linguist in her family, without describing a linguistic epiphany.

These linguistic epiphanies are evidence that linguistics **can** work as a “magic remedy” that influences how people think about a language or language variety and judge its speakers. They do not show, though, that this is automatically the case for all people brought in contact with linguistics. Linguistic epiphanies do occur, but their frequency is uncertain.

5.2.2.3 Latitude of acceptance and contact with linguistics

As discussed above, the answer behavior of subjects is characterized by an unusually high latitude of acceptance, indicative of ambivalence. Ambivalence can have several origins. According to the PAST-model (which I will discuss in more detail in the next chapter), this may actually be an effect of attitude change. A person who has undergone attitude change will demonstrate ambivalent attitudes in some situations – sometimes acting upon their old attitudes, sometimes upon their new attitudes. If contact with linguistics effects attitude change, it would be plausible to assume that subjects with contact to linguistics have a higher latitude of acceptance than those without such contact.

Even if you do not share the assumptions of the PAST-model, the influence of linguistics on latitude of acceptance is interesting. Does contact with linguistics reduce ambivalence, or increase it? Both would be possible. If contact with linguistics creates new beliefs or shakes old beliefs, this can certainly increase or decrease ambivalence.

Degree of contact	never contact	little contact	some contact	a lot of contact	linguist
Average latitude	282.7	237.7	274.1	269.8	176
Degree of contact	no contact		contact		linguist
Average latitude	276.5		273		176

Table 5.23: Latitude of acceptance across contact groups

Calculating an arithmetic mean for the linguist group is problematic. The linguist-subsample consists of only three individuals, one of whom did not endorse a single item, resulting in a latitude of acceptance of 0.

The latitude of acceptance of subjects is fairly similar between the different degree-of-contact groups (ignoring the linguist group). No pattern, such as, for example, increased latitude of acceptance with increased contact with linguistics, can be observed (cf. Table 5.23).

This observation does not provide evidence **against** the assumed relationship between contact with linguistics and attitudes toward basilectal varieties. It only demonstrates that contact with linguists seems to have no impact on the latitude of acceptance.

5.2.3 Summary

The attitudes of subject with and without contact with linguistics differ significantly. This may be interpreted as an influence of linguistics on attitudes toward AAVE. This influence, though, is of a complex nature. Teaching linguistics is not a ‘dead sure’ remedy of negative attitudes, nor is it a lame duck that has no effect whatsoever. The linguistic epiphanies described by two subjects, and the significant difference in five items out of thirteen between contact and no-contact subsample show the power of linguistic knowledge to change attitudes. The complex pattern of change, though, reminds us of the limitations of an approach to attitude change that focuses on message content only.

The attitudes toward AAVE observed here are characterized by ambivalence. This fact has both practical as well as theoretical ramifications, which I will discuss in the next chapter.

Chapter 6

Conclusion

6.1 Ambivalence and prior research

Ambivalence can only be discovered if a measuring technique is used that is (a) either designed specifically for detecting ambivalence, such as separate scales for attitude-aspects influenced by different values or for implicit and explicit attitudes,¹ or measurements that leave different amounts of time for the subjects to react, triggering both up-to-date (more processing time) and formerly held (less processing time) attitudes, or (b) if the scale value is high enough (interval or ratio scale level) to make the contradictory nature of item endorsements visible. It is likely that the ambivalence discovered in the sample here is not a new development, but that earlier measurements were not of a type that made its discovery possible.

Actually, stringent methodology can make discovering ambivalence more difficult. Some tests for the validity of the data, e.g. the split halves test, assume that attitudes are necessarily non-ambivalent. If the answer behavior of subjects shows an ambivalent pattern, this is taken as an indicator of faulty questionnaire construction. This also means that a questionnaire that discovers ambivalent attitudes may be faced with charges of not measuring what it claims to measure, without being able to refute this using conventional means. In my opinion, a simultaneous endorsement of “Speaking African American English is an indicator for laziness.” and “It would be a terrible thing to lose African American English.” can be intuitively grasped as expressing ambivalence, but this opinion is obviously liable to be

¹See discussion of Cavazza 2007 below. Explicit attitudes are attitudes a subject is aware of and for which the subject can manipulate the outcome of attempts to measure them. A subject may be unaware of his/her implicit attitudes, and may be unable to manipulate his/her scores in measurements of these implicit attitudes (cf. Ajzen and Fishbein 2005, 205)

challenged.

The linguistic literature reviewed for this thesis contained no references to ambivalence of attitudes toward AAVE. Noteworthy, though, are the observations Taylor 1973 made, that I briefly discussed in Section 3.2.2.1:

[T](t)eachers' attitudes relating to various topics of Nonstandard and Black English vary from topic to topic. Thus, teachers do not appear to have a single, generic attitude toward dialects, but, rather, differing attitudes depending upon the particular aspect of dialect being discussed. (Taylor 1973, 197)

Taylor observed that individuals endorsed some items but rejected others with similar evaluative meaning. Instead of assuming ambivalence of attitudes toward one attitude object, he postulates the existence of several attitude objects towards which people hold consistent attitudes.

The difference between these two interpretations is not as radical as it may seem. The exact distinction between intra-attitudinal (my interpretation) and inter-attitudinal (Taylor's interpretation) is less clear cut than may be assumed:

As noted earlier, an overall attitude toward an object might be influenced by evaluations of many specific attributes of the object or emotions associated with the object. Therefore, one could technically refer to many situations as involving inter-attitudinal structure even when only one object is considered. In our discussions, however, we retain the previous labels of intra-attitudinal when a single object is considered and inter-attitudinal when two or more objects are involved (usually at roughly the same level of abstraction). (Fabrigar et al. 2005, 80)

As I have outlined in 2.2.3, sociolinguistic research on attitudes has tended to be atheoretical in regard to the concept of attitude, attitude measurement, etc. Sociolinguistics has much to gain by looking closer at the research done by social psychologists. Not only are there methods to be discovered that might enrich sociolinguistic research, but there are also theories and concepts that might help interpret the data collected.

6.2 Ambivalence and attitude change

Social psychologists have put forward different theories to explain the origin of ambivalent attitudes. These theories imply different ramifications for attitude change programs.

Katz et al. 1986 argued that contradicting values were the reason for the existence of ambivalent attitudes:

Our earlier review of research on whites' racial attitudes revealed that both pro- and antiblack sentiments were prevalent, suggesting that a large proportion of people were ambivalent. We propose that this duality of attitudes emerges in some degree from a corresponding value duality. This means that in present-day America, to embrace the egalitarian precepts of equality of opportunity, social justice, and the worth of all human beings is to be disposed to identify with the needs and aspirations of minority underdogs, to feel sympathy for them, and to support efforts to improve their lot. Furthermore, there should be a tendency to see the disadvantaged group in a favorable light, perhaps emphasizing stereotypic positive traits like "warmth," "spontaneity," "group pride," and the "will to overcome."

On the other hand, a commitment to the individualistic, Protestant ethic ideas of freedom, self-reliance, devotion to work, and achievement should sensitize the observer to minority behavior patterns that deviate from and thereby threaten these cherished values. (...) Moreover, given the element of Puritanism in the Protestant ethic, its adherents should be inclined to attribute these deviant patterns to personality shortcomings in blacks themselves, rather than in situational factors such as lack of job opportunities. (Katz et al. 1986, 43f)

Some of the freetext comments support such an interpretation. On the one hand, AAVE is seen as a hindrance to economic advancement, indicator of laziness and lack of education, and as a means of 'secluding' oneself from mainstream society, mirroring some aspects of individualism ("emphasis on personal freedom, self-reliance, devotion to work, and achievement" (Katz et al. 1986, 42)). On the other hand, it is seen as part of the speakers' individuality that should be protected against interference from without, or as something 'wrong' which should not lead to its speakers being 'judged', though, which is loosely connected with egalitarianism ("democratic and humanitarian precepts" (Katz et al. 1986, 42)):

I do not frown personally upon those who speak in the Black American Vernacular. However, it is comprised of a good deal of incorrect grammar. Again, this is not a comment I make out of harshness or to judge the intelligence of those who speak in this manner. But factually said, the so-called "Black" American Vernacular is not correct English, and yes, it can affect a person on

the job market and in other professional settings. (...) (Subject 99)

African American English (aka, Ebonics) has a social place. All students of mine know “proper” English, but, I don’t interfere with their social language expression. (Subject 148)

Even though it seems to be sensible to interpret ambivalence as a consequence of conflicting values, the ‘fit’ between the statements made and individualism/equalitarianism is not very close. Different notions may be needed to explain the underlying values.

The notion of situational appropriateness/bidialectalism is something several subjects suggest or applaud. It is a way to harmonize these two conflicting values. You allow each person the freedom to express him/herself as he/she sees fit, as long as it is within the family circle. At the same time, you can enforce adaption to mainstream standards (and, through that, improve chances for education and professional success, as many subjects argue). Individuality/ethnic identity and utilitaristic conformism both ‘have their place’ in that suggestion:

Those who use African American English should be able to speak the language of their workplace or school. However, their use of language at home is their choice. (Subject 17)

I do think it is important to speak “standard” English on the job and at school, but for casual conversation and at home, AAE is fine. (Subject 15)

Code-switching is the key. I see no problem with African American English at home or with friends, but I do not think it has a place in school or higher education. Students must be able to switch back and forth between standard English and and African American English to ensure success in life. (Subject 9)

This type of ambivalence, caused by conflicting values, is also sometimes called *symbolic racism*. Symbolic racism assumes that racist values and non-racist values interact.

Some researchers interpreted the notion of ‘symbolic racism’ in such a way that there is a ‘real’ attitude (e.g. a racist attitude toward African Americans), and that people ‘cover’ it with positive statements, instead of a true conflict between conflicting values. This is not the sense in which the notion

was used by, for example, Kinder 1986 (see discussion Kinder 1986, 154ff). Some freetext comments would allow an interpretation according to that interpretation of the notion of symbolic racism:

I currently work at an elementary school where most students do not speak African American English, but when I come across the few that do, the way they speak is heavy and rough like a completely different accent. I find myself believing that it would be an interesting form in poetry and expression but for everyday language it makes people seem lazy. (Subject 85)

The acknowledgement “that it would be an interesting form in poetry and expression” can be interpreted as a truly positive aspect a person sees concerning a disliked attitude object, or as part of a strategy used to make a very critical statement seem more reasonable, as a way to disguise the actual attitude. Both interpretations are possible.

Other researchers assume that ambivalence is a consequence of attitude change. Petty, for example, attributes ambivalent attitudes to processes of attitude change. His model is called *PAST model*, *PAST* being an acronym for ‘Past attitudes are still there’ (Petty et al. 2006, 23). He summarizes his theory as follows:

In brief, the PAST model holds that when prior attitudes are no longer considered appropriate, the individual encodes an association with the rejected evaluation that essentially tags it as false or wrong. (Petty et al. 2006, 23)

Even if an attitude has been rejected and replaced by a new attitude, the attitude is still ‘there’, albeit tagged as ‘wrong’. In some situations, for example under time pressure, it is possible that the old attitude is retrieved, but the ‘wrong’ tag is not, resulting in dual attitudes, creating the impression of a single, ambivalent attitude, but usually without a feeling of conflict in the subject who experiences this special type of dual attitudes.

Based on which theory you endorse, you may be more or less optimistic about the success chances of attitude change programs. Values are fairly fixed, and if you needed to change a value in order to change an attitude, this would pose a major obstacle to attitude change. Breaking the link between the attitude object and the value (“XYZ is not really relevant for ABC”) might be slightly more promising. For representatives of the second model, attitude change is less of a problem, because, obviously, attitude change has already

taken place. If the PAST model applies to attitudes toward AAVE, all we might need to do is strengthen an existing negation. Interviewing teachers about their feelings of being ‘torn apart’ concerning this issue might reveal which model does **not** apply here, without, unfortunately, being evidence that the other model correctly describes the underlying mechanisms of ambivalence in teachers’ attitudes toward AAVE.

I contacted a number of subjects who had previously participated in this study and had given permission to be contacted individually and had volunteered an e-mail address. The question I asked was whether they perceived a feeling of being ‘torn’ concerning the topic of AAVE:

Sometimes, people are not sure whether they really like or dislike something. They may say “I know that carrots are good for you, but I can’t stand their taste!” or I “like the melody of that song, but the text is rather stupid!”. This is also called *ambivalence*.

Some people in this research project expressed opinions that were characterized by ambivalence.

My question: Do you think that your opinions on African American English are ambivalent, i.e. are there things you like and dislike about African American English? If yes, would you say that you feel ‘torn’ about the question of African American English, or does it give you no special feeling at all?

Six subjects responded to this e-mail, three stating that they were feeling torn about this question, three that they did not feel torn about it in any way.

Subjects expressing a feeling of ambivalence:

Yes, I would have to say there are things I like and dislike about the African American English. While I respect the language as part of the African American cultural, and I like the sing-song quality of the language, I also find myself torn when I hear it in an educational environment. Wrong, right, or indifferent the language of “education” in the U.S. is English. Even in a bilingual or ELL environment, English is either the goal or an equal partner in the educational setting. And while my perception has changed over the last couple of years (new teacher, now teaching in an urban environment), the African American English is not generally respected in the educational community as a language of the educated and that severely puts my students at a disadvantage.

I am troubled by how my students are perceived in consideration of the language that they use in everyday conversation and in the classroom. I am not sure if I have answered your question, but I hope I have offered additional insight. (Response 1)

I guess I would say that some of my opinions are ambivalent, in the sense that I see the use of African American English in a school or professional setting as a negative even though I also see the language as something unique to a culture and a short-hand way for people to communicate and a language that draws people together, and those are positive things.

And yes, I guess I would say I'm a bit torn as to how to resolve these different feelings. (Response 2)

I think it is fair to say that I have some ambivalence. I would say that I am torn between a desire to keep all the cultural richness of the language and the fear that we run the risk of not being able to talk to each other. We are already at the point where there is a tremendous divide, much of it created by language. I want everyone to have an equal opportunity to share in the pie, and I believe that this requires a common language. That said, the English language has always been enriched by other languages, and will continue to be in the future. Further, this question goes well beyond African American English. If we are going to be a melting pot we have to make accommodations. How's that for ambivalence. (Response 3)

Subjects not reporting a feeling of ambivalence:

I guess the best way to answer this is with an explanation. I taught English part-time at various community colleges for many years before teaching high school. The last [x, information deleted to ensure anonymity] years were at [a specific school, information deleted to ensure anonymity]. When I was first there, there were a very few Latino students in my classes. When I quit teaching there a few years ago, about 35 to 40 percent of each class was Latino.

As a white teacher (many of the English teachers were white) in an African-American school, I was the minority. Academic or business English was not the language of choice. I was very used to African American English, and frankly, I didn't think about

it. Once when I was tired, I actually heard myself using it in my reply to a student - who did look at me rather oddly for a second, and then didn't think about it further. The fact that I used it without thinking... well, I guess you could take from this that it doesn't give me any special feeling at all. It is a language like any other language, and in America, we have many, many dialects. My approach to teaching English is to tell my students that spoken English and written English are two very different languages, and that business/academic English is what you use in school and in the business world. That has always worked well, and frankly, it is the truth. I don't believe in teaching African American English, as was once proposed in the Oakland, CA school system, simply because it does not serve our students, who need to be prepared for the business world and for college. But it is a valid language, and should be understood as such by any teacher who wishes to truly help their students prepare for their future. (Response 4)

I don't feel torn at all, but as I recall, your survey didn't allow for my opinion. My students are primarily [ethnic group, information deleted to ensure anonymity] and [ethnic group, information deleted to ensure anonymity], so I don't hear African American English much. I feel that African American English is fine as a dialect. Language is a means of communication and AAE works well for that purpose. However, to succeed in the academic world or in some areas of business, academic "standard" English is necessary. There are advantages to being "bilingual" even when both languages are forms of English. Being able to communicate effectively, one needs to take into consideration one's audience and tailor the language to fit the audience and the situation. It's a practical matter. Absolutely, African Americans have a right to speak in a way that is meaningful to them.

In a sense, your topic is comparable to the kids in Aberdeen who were told they couldn't use Doric in school. Same idea. You need to fit into the majority society with more standard English when the occasion demands it. The kids who spoke Doric often were thought of as uneducated, as African American English speakers might be as well. I think we need to value a child's home language, whatever it is, but teach the language that will help students succeed in the larger society. It's their choice to learn standard academic English or not, and though not learning aca-

demic English doesn't totally limit their options, some roads will be closed to them - or at least more difficult.

Does this clarify my opinion for you? Essentially, I'm neutral, not torn, just trying to do my job. (Response 5)

The question of African American English gives me no special feelings at all. I would say the only thing that comes to mind when I think of AAE is "there is a time and a place". (Response 6)

This result seems to favor the value-induced ambivalence model discussed above, even though only half of the subjects who responded reported a feeling of being torn. Interestingly, all subjects who did not report feeling torn on this issue favor some form of bidialectalism. This might be interpreted in such a way that the idea of situational appropriateness serves to reduce the conflict between both values, reducing and/or eliminating the level of perceived ambivalence.

Independent from the question of the origin of ambivalence, different ideas have been brought forward concerning the effect of ambivalence on attitude stability, i.e. the influence of ambivalence on the ease with which attitudes can be changed.

Cavazza and Butera 2008 state that "Having a wider repertoire of responses is indeed a strength and not a weakness. (Cavazza and Butera 2008, 13). How do they come to this conclusion? They noticed that while people high on ambivalence reacted to social pressure by changing their explicit statements about their attitudes, if you tabbed their attitudes using more indirect measures,² you would find very little change, less change that is, than with individuals low on ambivalence. It could be argued, therefore, that in such situations individuals high in ambivalence seem to change a lot, while their attitudes remain comparably stable:

(...) it is possible that ambivalent individuals use a sort of compliance in order to manifest their agreement and at the same time to avoid a more structural change. (...) we hypothesize here that ambivalence might be used to adapt to the social environment, in a way that allows respecting the normative context while avoiding a deep change, one that involves both direct and indirect attitudes. (Cavazza and Butera 2008, 3)

²Cavazza and Butera distinguish between direct and indirect means of measurement. Direct measurements can be manipulated by the subject, indirect measurements hardly so (cf. Cavazza and Butera 2008, 3f).

Basically, an ambivalent attitude would protect individuals under social pressure against unwanted attitude change.

If these observations can be generalized, then studying attitude change in populations with ambivalent attitudes would be more difficult than so far assumed. Direct and indirect measures would have to be used in combination to yield more precise results that also include the more stable indirect attitudes. Furthermore, it would make subjects less prone to undergo ‘real’ attitude change. Even if the measure of direct attitude showed a change, more measures would be needed to be certain that change has actually taken place.

Clark et al. 2008, on the other hand, argue that ambivalence supports attitude change: “(...) if the processing of proattitudinal information decreases ambivalence, this might result in especially potent (strong) attitudes. In contrast, equally high processing of counterattitudinal information might not result in attitudes that are as strong because ambivalence has not been reduced (and might even increase).” (Clark et al. 2008, 575). Unlike Cavazza, Clark assumes that ambivalence is a state of internal conflict, a “subjectively uncomfortable” (Clark et al. 2008, 566) condition. Such a conflict may sharpen the attention to everything that may alleviate it, i.e. move a person to a positive or to a negative attitude, away from the ambivalent state he or she is in.³

More research is needed to find out how ambivalence acts in this specific case: Does it block change, or does it support change? In any case, attitude change programs should take the intra-attitudinal structure of the attitude they wish to change into consideration. The ambivalence of subjects may be utilized in an attitude change program.

6.3 A final thought

If contact with scientific data creates a desired and beneficial attitude change, this can only be applauded.

If this is not sufficient to create the kind of attitude change that is desired, or if attitude change might not be viewed as desired and beneficial by all parties involved, the question gets more complicated. What can we ethically do in order to create attitude change? Are there techniques which work, but from which it would be better to refrain? Is attitude change a ‘greater good’ in itself?

³This explanation, of course, is not possible if you assume the PAST-model, according to which subjects should not be experiencing a state of internal conflict.

We may choose not to use all methods that may potentially evoke attitude change for that purpose. Demonstrating facts, arguing different positions, outlining consequences of certain behaviors are all generally accepted forms of teaching. They address attitudes using a cognitive path at the same time as they teach generally useful knowledge and ideas. Acting as a role model yourself, or using ‘paid’ role-models as often done in advertising, would use the affective path. And you can even use a ‘conative path’ by encouraging or enforcing behavior change and thereby evoking cognitive inconsistency, making change in the underlying attitude likely. These non-cognitive approaches attempt to change attitudes without directly *persuading* the individual. They use (and abuse?) psychological strategies in similar ways as advertising and propaganda do. While I have little doubt that using a cognitive approach to attitude change is ethically acceptable as long as the data used has a factual basis and has not been invented in order to evoke attitude change, some other approaches, even if they have the potential to be more effective, do raise moral questions.

Bibliography

- Abdul-Hakim, I. (2002). *Florida preservice teachers' attitudes toward African American Vernacular English*. Ph. D. thesis, Florida State University. etd.lib.fsu.edu/theses/available/etd-06232003-112932/.
- Abrams, D. and M. A. Hogg (1987). Language attitudes, frames of reference, and social identity: a Scottish dimension. *Journal of language and social psychology* 6, 201–213.
- Ajzen, I. and M. Fishbein (2005). The influence of attitudes on behavior. In D. Albarracín, B. T. Johnson, and M. P. Zanna (Eds.), *The handbook of attitudes*, pp. 173–221. Mahwah: Lawrence Erlbaum Associates.
- Barnes, S. L. (1998). Ebonics and public awareness: who knows? who cares? *Journal of Black studies* 29(1), 17–33.
- Baugh, J. (2001). Applying linguistic knowledge of African American English to help students learn and teachers teach. In S. L. Lanehart (Ed.), *Sociocultural and historical contexts of African American English*, pp. 319–330. Amsterdam: John Benjamins.
- Blake, R. and C. Cutler (2003, 2). AAE and variation in teachers' attitudes: a question of school philosophy? *Linguistics and education* 14, 163–194.
- Bogardus, E. S. (1925). Measuring social distance. *Journal of applied sociology*, 299–308.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* (6th ed.). Heidelberg: Springer.
- Bortz, J. and G. A. Lienert (2003 [1998]). *Kurzgefaßte Statistik für die klinische Forschung: Leitfaden für die verteilungsfreie Analyse kleiner Stichproben* (2nd ed.). Heidelberg: Springer.
- Bowie, R. L. and C. L. Bond (1994). Influencing future teachers' attitudes toward Black English: are we making a difference? *Journal of teacher education* 42(2), 112–118.

- Bradac, J. L., A. C. Cargile, and J. S. Hallett (2001). Language attitudes: retrospect, conspect, and prospect. In P. W. Robinson and H. Giles (Eds.), *The new handbook of language and social psychology*, pp. 137–155. Chichester: John Wiley & Sons.
- Braddock, R., R. Lloyd-Jones, and Lowell-Schoer (1963). *Research in written composition*. Champagne: National Council of teachers of English.
- Buendgens, J. (2002). In how far is AAVE intelligible for speakers of SAE? a study on the remote present perfect. Unpublished seminar paper.
- Buendgens-Kosten, J. (2006). The teacher as language expert. Poster presentation at the *International Symposium on African & Diasporic Languages and Education*, New York: Columbia University, Teachers' College, October 5-7, 2006.
- Buendgens-Kosten, J. (2007). Definitions of AAVE in lay discourse. Presentation at the *49th Annual M/MLA Convention*, 11/8/2007-11/11/2007, Cleveland.
- Cargile, A. C., J. Takai, and J. I. Rodriguez (2006). Attitudes toward African-American Vernacular English: A US export to Japan? *Journal of multilingual and multicultural development* 27(6), 443–456.
- Casper, K. (2002). *Spracheinstellungen: Theorie und Messung*. [n.p.]: Books on Demand.
- Cavazza, N. and F. Butera (2008). Bending without breaking: Examining the role of attitudinal ambivalence in resisting persuasive communication. *European journal of social psychology* 38, 1–15.
- Cecil, N. C. (1988). Black Dialect and academic success: a study of teacher expectations. *Reading improvement* 25, 34–38.
- Clark, J. K., D. T. Wegener, and L. R. Fabrigar (2008). Attitudinal ambivalence and message-based persuasion: motivated processing of proattitudinal information and avoidance of counterattitudinal information. *Personality and social psychology bulletin* 34, 565–577.
- Cooper, R. L. and J. A. Fishman (1974). The study of language attitudes. *International journal of the sociology of language* 3, 5–19.
- Covington, A. J. (1975). Teacher's attitudes toward Black English: effects on student achievement. In R. L. Williams (Ed.), *Ebonics: the true language of Black folks*, pp. 40–54. St. Louis: Institute for Black studies.
- Danziger, K. (1997). *Naming the mind: how psychology found its language*. London: Sage Publ.

- Di Giulio, R. C. (1973). Measuring teacher attitudes toward Black English: a pilot project. *The Florida FL reporter Spring/Fall*, 25–26; 49.
- Dillard, J. L. (1972). *Black English: its history and usage in the United States*. New York: Random House.
- Eagly, A. H. and S. Chaiken (1993). *The psychology of attitudes*. Fort Worth: Harcourt Brace Hovanovich College.
- Edwards, A. L. (1957). *Techniques of attitude scale construction*. New York: Appleton-Century-Crofts.
- Edwards, A. L. and F. P. Kilpatrick (1948). Scale analysis and the measurement of social attitudes. *Psychometrika* 13(2), 99–114.
- Edwards, J. R. (1982). Language attitudes and their implications among English speakers. In E. B. Ryan and H. Giles (Eds.), *Attitudes towards language variation: social and applied contexts*, pp. 20–33. London: Edward Arnold.
- Eller, R. G. (1989). Johnny can't talk, either: the perpetuation of the deficit theory in classrooms. *The reading teacher* 42, 670–674.
- Fabrigar, L. R., T. K. MacDonald, and D. T. Wegener (2005). The structure of attitudes. In D. Albarracín, B. T. Johnson, and M. P. Zanna (Eds.), *The handbook of attitudes*, pp. 79–124. Mahwah: Lawrence Erlbaum Associates.
- Fasold, J. L. (1972). *Black English: Its history and usage in the United States*. New York: Random House.
- Fasold, R. W. (1981). The relation between Black and White speech in the South. *American speech* 56, 163–189.
- Fishbein, M. (1963). An investigation of the relationships between beliefs about an object and the attitude toward that object. *Human relations* 16, 233–239.
- Ford, J. F. (1978). The prospective foreign language teacher and the culturally and linguistically different learner. *Foreign language annals* 11, 381–390.
- Garrett, P., N. Coupland, and A. Williams (2003). *Investigating language attitudes: social meanings of dialect, ethnicity and performance*. Cardiff: University of Wales Press.
- Giles, H. and E. B. Ryan (1982). Prolegomena for developing a social psychological theory of language attitudes. In E. B. Ryan and H. Giles (Eds.), *Attitudes toward language variation: social and applied contexts*, pp. 208–223. London: Edward Arnold.

- Goldwater, B. C. (1972). Psychological significance of pupillary movements. *Psychological bulletin* 77(5), 340–355.
- Green, L. J. (2002). *African American English: a linguistic introduction*. Cambridge: Cambridge University Press.
- Greenwald, A. G. (1968). Cognitive learning, cognitive response to persuasion, and attitude change. In A. G. Greenwald, T. C. Brock, and T. M. Ostrom (Eds.), *Psychological foundations of attitudes*, pp. 147–170. New York: Academic Press.
- Guttman, L. (1967 [1944]). A basis for scaling qualitative data. In M. Fishbein (Ed.), *Attitude theory and measurement*, pp. 96–107. New York: John Wiley & Sons.
- Haas, G. R. and et al. (1991). Cross-racial appraisal as related to attitude, ambivalence and cognitive complexity. *Personality and social psychology bulletin* 17, 83–92.
- Hoffman, M. J. (1998). Ebonics: the third incarnation of a thirty-three year old controversy about Black English in the United States. *Links & letters* 5, 75–87.
- Hoover, M. R., F. McNair-Knox, S. A. R. Lewis, and R. L. Politzer (1996). African American English attitude measures for teachers. In R. L. Jones (Ed.), *Handbook of tests and measurements for Black populations*, Volume I, pp. 383–259. Hampton, VA: Cobb & Henry Publishers.
- Joiner (1980). Memorandum opinion and order. In G. Smitherman (Ed.), *Black English and the education of Black children and youth: proceedings of the National Invitational Symposium of the King Decision*, pp. 336–358. Detroit: Center for Black Studies.
- Jonas, K. and R. Ziegler (2007). Attitudinal ambivalence. In M. H. et al. (Ed.), *The scope of social psychology: theory and applications: essays in honour of Wolfgang Stroebe*, pp. 29–42. Hove: Psychology Press.
- Katz, I. and D. C. Glass (1973). An ambivalence-amplification theory of behavior toward the stigmatized. In W. G. Austin and S. Worchel (Eds.), *The social psychology of intergroup relations*, pp. 55–70. Monterey, CA: Brooks/Cole Publishing Company.
- Katz, I., J. Wackenhut, and R. G. Hass (1986). Racial ambivalence, value, duality, and behavior. In J. F. Dovidio and S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism*, pp. 35–59. Academic Press.
- Kinder, D. R. (1986). The continuing American dilemma: White resistance to racial change 40 years after Myrdal. *Journal of social issues* 42(2),

151–171.

- Kretzschmar, Jr., W. A. (2008). Public and academic understandings about language: The intellectual history of Ebonics. *English world-wide* 29(1), 70–95.
- Kruglanski, A. W. and W. Stroebe (2005). The influence of beliefs and goals on attitudes: issues of structure, function, and dynamics. In D. Albarracín, B. T. Johnson, and M. P. Zanna (Eds.), *The handbook of attitudes*, pp. 323–368. Mahwah: Lawrence Erlbaum Associates.
- Labov, W. (1975 [1972]). *Language in the inner city: studies in the Black English Vernacular*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (2008). The cognitive capacities of the sociolinguistic monitor. <http://www.ling.upenn.edu/~wlabov/PowerPoints/SLM.SS17.ppt>. Keynote presentation at Sociolinguistics Symposium 17 in Amsterdam on March 5, 2008. [Online; accessed 09/09/2008].
- Lambert, W. E. (1967). A social psychology of bilingualism. *Journal of social issues* 23(2), 91–109.
- Likert, R. (1932). *A technique for the measurement of attitudes*. Archives of psychology. 140. New York: [n.p.].
- Likert, R. (1967). The method of constructing an attitude scale. In M. Fishbein (Ed.), *Attitude theory and measurement*, Chapter 11, pp. 90–95. New York: John Wiley & Sons.
- Lippi-Green, R. (1997). *English with an accent: Language, ideology and discrimination in the United States*. London: Routledge.
- Lippi-Green, R. (2000). That's not my language: the struggle to (re)define African American English. In R. D. Gonzles and I. Melis (Eds.), *Language ideologies: critical perspectives of the Official English Movement: volume 1: education and the social implications of official language*, pp. 230–247. Mahwah: Lawrence Erlbaum Associates.
- Löffler, H. (1980). Dialektfehler: Ansätze zu einer deutschen "Fehlergeographie". In D. Cherubim (Ed.), *Fehlerlinguistik: Beiträge zum Problem der sprachlichen Abweichung*, pp. 94–105. Tübingen: Max Niemeyer.
- McGuire, W. J. (1986). The vicissitude of attitudes and similar representational constructs in twentieth century psychology. *European journal of social psychology* 16, 89–130.

- McKenzie, R. M. (2008). Social factors and non-native attitudes towards varieties of spoken English: a Japanese case study. *International Journal of Applied Linguistics* 18(1), 63–88.
- Niedzielski, N. A. and D. R. Preston (2003). *Folk linguistics*. Berlin: Mouton de Gruyter.
- Okawa, G. Y. (2000). From “bad attitudes” to(ward) linguistic pluralism: developing reflective language policy among preservice teachers. In R. D. Gonzáles and I. Melis (Eds.), *Language ideologies: critical perspectives of the Official English Movement: volume 1: education and the social implications of official language*, pp. 276–296. Mahwah, NJ: Lawrence Erlbaum Associates.
- Osgood, C. E. (1967 [1965]). Cross-cultural comparability in attitude measurement via multilingual semantic differentials. In M. Fishbein (Ed.), *Attitude theory and measurement*, pp. 108–116. New York: John Wiley & Sons.
- Osgood, C. E., W. H. May, and M. S. Miron (1975). *Cross-cultural universals of affective meaning*. Urbana: University of Illinois Press.
- Osgood, C. E., G. J. Suci, and P. H. Tannenbaum (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Pandey, A. (2000). Linguistic power in virtual communities: the Ebonics debate on the internet. *World Englishes* 19(1), 21–38.
- Petty, R. E., Z. L. Tormala, P. Briñol, and W. B. G. Jarvis (2006). Implicit ambivalence from attitude change: an exploration of the PAST model. *Journal of personality and social psychology* 90(1), 21–41.
- Priester, J. R. and R. E. Petty (1996). The gradual threshold model of ambivalence: relating the positive and negative bases of attitudes to subjective ambivalence. *Journal of personality and social psychology* 71(3), 431–449.
- Rickford, J. R. (1999). *African American Vernacular English: features, evolution, educational implication*. Language in society. 26. Malden, MA: Blackwell.
- Rickford, J. R. and R. J. Rickford (2000). *Spoken Soul: the story of Black English*. New York: John Wiley & Sons.
- Robinson, J. A. (1996). The relationship between personal characteristics and attitudes toward Black and White speakers of informal nonstandard English. *The western journal of Black studies* 20(4), 211–220.

- Rosenthal, M. S. (1974). The magic boxes: pre-school children's attitudes toward Black and Standard English. *The Florida FL Reporter Spring/Fall*, 55–62; 92–93.
- Ryan, E. B., H. Giles, and R. J. Sebastian (1982). An integrative perspective for the study of attitudes toward language variation. In E. B. Ryan and H. Giles (Eds.), *Attitudes towards language variation: social and applied contexts*, pp. 1–19. London: Edward Arnold.
- Schenk, J. (1972). Stereotype. In H. J. Eysenck, W. Arnold, and R. Meili (Eds.), *Encyclopedia of psychology*, pp. 273–275. New York: Herder and Herder.
- Sherif, M. and C. I. Hovland (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change*. New Haven: Yale University Press.
- Shores, D. L. (1977). Black English and Black attitudes. In D. L. Shores and C. P. Hines (Eds.), *Papers in language variation: SAMLS-ADS collection*, pp. 177–187. [n.p.]: University of Alabama Press.
- Shuy, R. W. (1970). Teacher training and urban language problems. In R. W. Fasold and R. W. Shuy (Eds.), *Teaching Standard English in the inner city*, pp. 120–141. Washington, D.C.: Center for applied linguistics.
- Smitherman, G. (1973). “God don’t never change”: Black English from a Black perspective. *College English* 34,5, 827–833.
- Smitherman, G. (Ed.) (2006). *Words from the mother: language and African Americans*. New York: Routledge.
- Smitherman-Donaldson, G. (1988). Discriminatory discourse on Afro-American speech. In G. Smitherman-Donaldson and T. A. van Dijk (Eds.), *Discourse and discrimination*, pp. 145–175. Detroit: Wayne University Press.
- Speicher, B. L. and S. M. McMahon (1992). Some African American perspectives on Black English Vernacular. *Language in society* 21, 383–407.
- St. Clair, R. N. (1982). From social history to language attitude. In E. B. Ryan and H. Giles (Eds.), *Attitudes towards language variation: social and applied contexts*, pp. 164–174. London: Edward Arnold.
- Taiffel, H. (1972). Prejudice. In H. J. Eysenck, W. Arnold, and R. Meili (Eds.), *Encyclopedia of psychology*, Volume 3, pp. 37–39. New York: Herder and Herder.

- Taylor, O. (1973). Teachers' attitudes toward Black and nonstandard English as measured by the language attitude scale. In R. W. Shuy and R. W. Fasold (Eds.), *Language attitudes: current trends and prospects*, pp. 174–201. Washington, D.C.: Georgetown University Press.
- Taylor, O. L. (1975). Black language and what to do about it: some Black community perspectives. In R. L. Williams (Ed.), *Ebonics: the true language of Black folks*, pp. 28–39. St. Louis: The institute for Black studies.
- Thomas, W. I. and F. Znaniecki (1918). *The Polish peasant in Europe and America: monograph of an immigrant group*, Volume 1. Boston: The Gorham Press.
- Thurstone, L. (1967). Attitudes can be measured. In M. Fishbein (Ed.), *Attitude theory and measurement*, pp. 77–89. New York: John Wiley & Sons. Reprint from 1928.
- Thurstone, L. L. and E. J. Chave (1966 [1929]). *The measurement of attitude: a psychophysical method and some experiments with a scale for measuring attitude toward the church*. Chicago: University of Chicago Press.
- Tittle, C. R. and R. J. Hill (1967). Attitude measurement and prediction of behavior: an evaluation of conditions and measurement techniques. *Sociometry* 30, 199–213.
- U.S. Census Bureau (2001). Questions and answers for census 2000 data on race, march 14, 2001. <http://www.census.gov/Press-Release/www/2001/raceqandas.html>. [Online; accessed 12/24/2008].
- U.S. Census Bureau ([n.d.]a). Census 2000 EEO data tool. <http://www.census.gov/eo2000/index.html>. [Online; accessed 12/19/2008].
- U.S. Census Bureau ([n.d.]b). Facts for features: Special edition: Teacher appreciation week (may 2-8). http://www.census.gov/Press-Release/www/releases/archives/facts_for_features_special_editions/001737.html. [Online; accessed 12/24/2008].
- U.S. Department of Commerce Economics and Statistics Administration U.S. Census Bureau, Geography Division ([n.d.]). Census regions and divisions of the United States. http://commons.wikimedia.org/wiki/File:Census_Regions_and_Divisions.PNG. [Online; accessed 12/27/2008].
- White, M. J., B. J. Vandicer, M. L. Becker, B. G. Overstreet, L. E. Temple, K. L. Hagan, and E. P. Mandelbaum (1998). African American

- evaluations of Black English and Standard American English. *Journal of Black psychology* 24, 60–75.
- Williams, F. (1973). Some research notes on dialect attitudes and stereotypes. In R. W. Shuy and R. Fasold (Eds.), *Language attitudes: current trends and prospects*, pp. 113–128. Washington, D.C.: Georgetown University Press.
- Williams, F., N. Hewett, R. Hopper, L. M. Miller, R. C. Naremore, and J. L. Whitehead (1976). *Explorations of the linguistic attitudes of teachers*. Rowley, MA: Newbury House Publishers.
- Wolfram, W. (1991). *Dialects and American English*. Englewood Cliffs: Prentice Hall.
- Wolfram, W. (1999). Repercussions from the Oakland Ebonics controversy: the critical role of dialect awareness programs. In C. T. Adger and et al. (Eds.), *Making the connection: language and academic achievement among African American students*, Chapter 61-80. [n.p.]: Center for applied linguistics.
- Wolfram, W. and E. R. Thomas (2002). *The development of African American English*. Language in society. 31. [n.p]: Blackwell Publishing.
- Woodworth, W. D. and R. T. Salzer (1971). Black children's speech and teachers' evaluations. *Urban education* 6, 167–173.
- Zimbardo, P. G. (1992). *Psychologie* (5th ed.). Berlin: Springer Verlag.

Index

- AAVE
 - awareness of, 7, 129
 - history of research, 16
 - lay perspective, 7
 - linguistic perspective, 4
- ambivalence, 22, 150, 158, 160, 161
- anchor effect, 35, 55
- attitude
 - conscious attitudes, 11
 - definition, 10, 18
 - direct, 168
 - history of research, 10, 13
 - in biology, 11
 - in sociolinguistics, 23
 - in the arts, 10
 - indirect, 168
 - internal structure, 22
 - interrelationship with linguistics, 111, 124, 152
 - motor attitude, 11
 - origin, 19
 - purpose, 19
- attitude accessibility, 21
- attitude change, 109, 169
 - and ambivalence, 161
 - perception, 155
- attitude measurement, 12, 26, 34
 - linked to specific indicator classes, 48
 - problems, 26, 27
- attitude object, 3
 - associated with ambivalence, 151
- attitude stability, 22, 168
- attitude strength, 21, 169
- Bewußtseinslage, 11
- bidialectalism, 163
- Bogardus, 12
- character trait, 20
- classes of evaluative responding, 48
- conscious attitudes, 11
- contact with linguistics, 158
- content analysis, 53
- Edwards-Kilpatrick scale, 38
- expectations (of teachers), 1, 65, 69
- F-matrix, 135, 182
- facial electromyographic activity, 50
- galvanic skin response, 50
- German dialects, 111
- Guttman, 13, 37
- history
 - of AAVE research, 16
 - of attitude research, 10, 13
- hypothetical construct, 21, 26
- ideology, 20
- incarnations of AAVE debate, 15
- indicators, 48
 - affective, 50
 - cognitive, 48
 - behavioral, 51
- internal consistency, 136, 184
- item
 - scaling, 132

- selection, 113, 129
- language attitudes
 - history of research, 13
- latent process, 21
- latitude of acceptance, 148, 158
- law of comparative judgment, 132
- Likert, 39, 54, 72, 76, 82, 85, 88, 90, 97
- linguistic epiphany, 157
- linguistics
 - and attitudes, 1, 111, 124, 152, 158
 - contact with, 152
- magnitude estimation, 36
- matched guise, 28, 47
- method of equal-appearing intervals, 34, 35
- method of paired comparisons, 34, 35, 55, 132
- method of successive intervals, 34, 35
- mood, 20
- motor attitude, 11
- opinion polling, 13
- Osgood, 45, 54, 58
- P-matrix, 135, 182
- PAST model, 158, 164
- person scaling, 39
- prejudice, 19
- pupillary response, 50
- Pygmalion effect, 1, 65, 69, 169
- questionnaire, 126
- Rosenthal effect, 1, 65, 69, 169
- scale construction, 136
- scale level, 32
 - ordinal, 114
- schema, 20
- semantic differential, 45
- situational appropriateness, 163
- stereotype, 19
- Stevens, 36
- stimulus, 27, 127
- stimulus-then-person-scaling, 34
- symbolic racism, 163
- terminology, xi, 30, 127
- Thurstone, 34, 54, 132
- value, 20, 162
- Z-matrix, 135, 182

Calculations belonging to the scale construction

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	na	19	11	15	5	5	0	0	4	1	0	1	1
2	1	na	3	3	1	1	1	1	1	1	1	1	1
3	9	17	na	14	8	5	1	1	1	0	0	1	1
4	5	17	6	na	5	6	1	2	1	1	1	1	1
5	15	19	14	15	na	8	1	1	1	1	0	3	1
6	15	19	15	14	12	na	1	1	2	1	1	4	1
7	20	19	19	19	19	19	na	17	13	9	1	10	1
8	20	19	19	18	19	19	3	na	8	4	2	7	1
9	16	19	19	19	19	18	7	12	na	6	3	10	1
10	19	19	20	19	19	19	11	16	14	na	2	8	2
11	20	19	20	19	20	19	19	18	17	18	na	15	9
12	19	19	19	19	17	16	10	13	10	12	5	na	3
13	19	19	19	19	19	19	19	19	19	18	11	17	na

Table 1: F-matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	na	0.95	0.55	0.75	0.25	0.25	0	0	0.2	0.05	0	0.05	0.05
2	0.05	na	0.15	0.15	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
3	0.45	0.85	na	0.7	0.4	0.25	0.05	0.05	0.05	0	0	0.05	0.05
4	0.25	0.85	0.3	na	0.25	0.3	0.05	0.1	0.05	0.05	0.05	0.05	0.05
5	0.75	0.95	0.7	0.75	na	0.4	0.05	0.05	0.05	0.05	0	0.15	0.05
6	0.75	0.95	0.75	0.7	0.6	na	0.05	0.05	0.1	0.05	0.05	0.2	0.05
7	1	0.95	0.95	0.95	0.95	0.95	na	0.85	0.65	0.45	0.05	0.5	0.05
8	1	0.95	0.95	0.9	0.95	0.95	0.15	na	0.4	0.2	0.1	0.35	0.05
9	0.8	0.95	0.95	0.95	0.95	0.9	0.35	0.6	na	0.3	0.15	0.5	0.05
10	0.95	0.95	1	0.95	0.95	0.95	0.55	0.8	0.7	na	0.1	0.4	0.1
11	1	0.95	1	0.95	1	0.95	0.95	0.9	0.85	0.9	na	0.75	4.5
12	0.95	0.95	0.95	0.95	0.85	0.8	0.5	0.65	0.5	0.6	0.25	na	0.15
13	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.9	0.55	0.85	na

Table 2: P-matrix

CALCULATIONS BELONGING TO THE SCALE CONSTRUCTION 182

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.000	-	0.126	0.674	-0.674	-0.674	-	-	-0.842	-	-	-	-
2	-	0.000	-1.036	-1.036	-	-	-	-	-	-	-	-	-
3	-0.126	1.036	0.000	0.524	-0.253	-0.674	-	-	-	-	-	-	-
4	-0.674	1.036	-0.524	0.000	-0.674	-0.524	-	-1.282	-	-	-	-	-
5	0.674	-	0.524	0.674	0.000	-0.253	-	-	-	-	-	-1.036	-
6	0.674	-	0.674	0.524	0.253	0.000	-	-	-1.282	-	-	-0.842	-
7	-	-	-	-	-	-	0.000	1.036	0.385	-0.126	-	0.000	-
8	-	-	-	1.282	-	-	-1.036	0.000	-0.253	-0.842	-1.282	-0.385	-
9	0.842	-	-	-	-	1.282	-0.385	0.253	0.000	-0.524	-1.036	0.00	-
10	-	-	-	-	-	-	0.126	0.842	0.524	0.000	-1.282	-0.253	-1.282
11	-	-	-	-	-	-	-	1.282	1.036	1.282	0.000	0.674	-0.126
12	-	-	-	-	1.036	0.842	0.00	0.385	0.00	0.253	-0.674	0.000	-1.036
13	-	-	-	-	-	-	-	-	-	1.282	0.126	1.036	0.000

Table 3: Z-matrix

	2-1	3-2	4-3	5-4	6-5	7-6	8-7	9-8	10-9	11-10	12-11	13-12
1	-	-	0.548	-1.348	0	-	-	-	-	-	-	-
2	-	-1.036	0	-	-	-	-	-	-	-	-	-
3	1.162	-1.036	0.524	-0.777	-0.421	-	-	-	-	-	-	-
4	1.71	-1.56	0.524	-0.674	0.15	-	-	-	-	-	-	-
5	-	-	0.15	-0.674	-0.253	-	-	-	-	-	-	-
6	-	-	-0.15	-0.271	-0.253	-	-	-	-	-	-	-
7	-	-	-	-	-	-	1.036	-0.651	-0.511	-	-	-
8	-	-	-	-	-	-	1.036	-0.253	-0.589	-0.44	0.897	-
9	-	-	-	-	-	-1.667	0.638	-0.253	-0.524	-0.512	1.036	-
10	-	-	-	-	-	-	0.716	-0.318	-0.524	-1.282	1.029	1.029
11	-	-	-	-	-	-	-	-0.246	0.246	-1.282	0.674	-0.8
12	-	-	-	-	-0.194	-0.842	0.385	-0.385	0.253	0.927	0.674	-1.036
13	-	-	-	-	-	-	-	-	-	-1.156	0.91	-1.036
Sums	2.872	-3.632	1.596	-3.744	-0.971	-2.509	3.811	-2.106	-1.649	-3.745	5.22	-1.843
n	2	3	6	5	6	2	5	6	6	6	6	4
Means	1.436	-	0.266	-	-	-	0.7622	-0.351	-	-	0.87	-
Mean*100	143.63	-	26.6	0.7488	0.1618p3	1.2545	76.22	-35.1	0.2748p3	0.6241p6	87.0	0.46075
(rounded)		121.07		-74.88	-16.18	-	125.45		-27.48	-62.42		-46.08

Table 4: Column differences

Calculations belonging to the internal consistency check

		1	2	3	4	5	6	7	8	9	10	11	12	13
		0	1.44	-	0.27	-	-	-	0.77	-	-	-	0.87	-
				1.21		0.75	0.16	1.25		0.35	0.27	0.62		0.46
1	0													
2	1.44	-												
		1.44												
3	-	1.21	2.65											
4	0.27	-	1.17	-										
		0.27		1.5										
5	-	0.75	2.18	-	1.01									
		0.75		0.46										
6	-	0.16	1.6	-	0.43	-								
		0.16		1.05		0.59								
7	-	1.26	2.7	0.04	1.52	0.51	1.1							
		1.26												
8	0.76	-	0.67	-	-	-	-	-						
		0.76		2.0	0.5	1.51	0.92	2.02						
9	-	0.35	1.79	-	0.62	-	0.19	-	1.11					
		0.35		0.86		0.4		0.90						
10	-	0.28	1.71	-	0.54	-	0.11	-	1.04	-				
		0.28		0.94		0.47		0.98		0.08				
11	-	0.62	2.06	-	0.89	-	0.46	-	1.39	0.27	0.35			
		0.62		0.59		0.13		0.63						
12	0.87	-	0.57	-	-	-	-	-	-	-	-	-		
		0.87		2.08	0.60	1.62	1.03	2.13	0.11	1.22	1.15	1.49		
13	-	0.46	1.9	-	0.73	-	0.3	-	1.223	0.11	0.19	-	1.30	
		0.46		0.75		0.29		0.79			0.16			

Table 5: Z' matrix (rounded)

		0	1.44	-	0.27	-	-	-	0.76	-	-	-	0.87	-
1	0			1.21		0.75	0.16	1.26		0.35	0.28	0.62		0.46
2	1.44	0.08												
3	-	0.89	0.60											
	1.21													
4	0.27	0.40	0.88	0.07										
5	-	0.77	0.99	0.32	0.85									
	0.75													
6	-	0.56	0.95	0.20	0.67	0.28								
	0.16													
7	-	0.90	1.00	0.52	0.94	0.70	0.86							
	1.26													
8	0.76	0.22	0.75	0.0	0.31	0.07	0.18	0.00						
9	-	0.64	0.96	0.20	0.73	0.35	0.51	0.18	0.87					
	0.35													
10	-	0.61	0.96	0.18	0.71	0.32	0.55	0.16	0.85	0.47				
	0.28													
11	-	0.73	0.98	0.28	0.81	0.45	0.67	0.26	0.92	0.59	0.64			
	0.62													
12	0.87	0.19	0.71	0.02	0.27	0.05	0.15	0.02	0.46	0.11	0.15	0.07		
13	-	0.68	0.97	0.23	0.77	0.49	0.62	0.21	0.89	0.54	0.57	0.44	0.90	
	0.46													

Table 6: P' matrix (rounded)

	1	2	3	4	5	6	7	8	9	10	11	12
1												
2	0.05											
3	-0.42	0.25										
4	-0.15	-0.03	0.23									
5	0.02	0.95	0.38	-0.1								
6	0.19	0.95	0.6	0.03	0.32							
7	1	0.95	0.95	0.95	0.95	0.95						
8	1	0.95	0.95	0.59	0.95	0.95	0.13					
9	0.16	0.95	0.95	0.95	0.95	0.39	0.17	-0.27				
10	0.95	0.95	1	0.95	0.95	0.95	0.39	-0.05	0.23			
11	1	0.95	1	0.95	1	0.95	0.95	-0.02	0.26	0.26		
12	0.95	0.95	0.95	0.95	0.8	0.65	0.48	0.19	0.39	0.45	0.18	
13	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.33	0.12	-0.05
Sum	0.94	0.22	1.21	0.72	1.12	1.04	1.16	0.53	0.88	1.04	0.3	0.05
N	5	2	3	3	2	2	4	4	3	3	2	1
Av	0.19	0.11	0.40	0.24	0.56	0.52	0.29	0.13	0.29	0.35	0.15	0.05

Table 7: Difference between P and P' matrix (rounded)

Text of the main questionnaire

Page 1: Questionnaire

Thank you for participating in this study. If you have any questions, feel free to contact me at [contact e-mail address].

Demographics

Age: [freetext field]

Teaching experience (in years): [freetext field]

Sex: [dropdown menu]

Ethnicity: [freetext field]

State: [dropdown menu]

School School type

Kindergarten

Elementary

Middle School

Junior High

High School

College

Other type of school (please give details)

Details: [freetext field]

I teach special education.

Employment

I currently work as a teacher.

I am not currently employed as a teacher, but have been so in the past (e.g. retired, between jobs)

I am a teaching assistant or similar.

I am a school administrator or similar; I do not teach students myself.

I am a school administrator or similar; I either teach students myself or have done so in the past.

I am currently undergoing teacher training.

Contact with linguistics

I have never had⁴ contact with linguistics in any form.

I have had a little contact with linguistics (e.g. read an article in a journal, watched a documentary)

I have had some contact with linguistics (e.g. attended at least one seminar on linguistics at university or attended a teacher training workshop, read one or more books on linguistics)

I have had a lot contact with linguistics (e.g. attended several seminars on linguistics at university, etc.)

I am a linguist (e.g. degree in linguistics)

Overall, I have had contact with linguistics

in my free time (hobbies and entertainment)

through my job (e.g. on-the-job training, books and journals for teachers) during university training

Questionnaire

It is important that you know what this questionnaire is actually about. All sentences will center around a form of speech called African American English or African American Vernacular English by linguists. It is spoken by many African Americans, especially in informal situations (for example at home or when together with friends). A typical sentence would be: He be workin hard. ('He is always working hard') Sometimes, this form of speech is also called Black English or Ebonics.

Is is not important that you personally use the same name for it as is used here, as long as you basically understand what it is about.

I know what is meant by African American English.

I do not know what is meant by African American English.

Do you agree with the following statements:

I agree I disagree Statement

⁴In an early version of this questionnaire, this and the following items had been phrased in the simple past instead of the present perfect simple. This was changed only after a few subjects had already responded to the questionnaire.

African American English has a place at the home of its speakers.
Speaking African American English can cause difficulties on the job market.
African American English can add flavour and soul to a poem, song or novel.
It would be a terrible thing to lose African American English.
African American English is a perfectly normal thing in American media.
African American English is used to strengthen a common identity between its speakers.
Speakers of African American English do not express complete thoughts.
Speaking African American English is an indicator for laziness.
African American English is very structured.
African American English is broken English.
African American English has very little grammar.
African American English is not cool.
African American English differs from other forms of speech.

Comments Any comments you would like to make? [freetext field]

Do you believe that what you learned in linguistics changed your opinions regarding African American English?

yes

no

I don't know

not applicable

To confirm you are a human being, please solve this equation: $5+3$ [freetext field]

Page 2: Questionnaire

Thank you very much!

If you want to keep informed about the progress of this study, please enter your e-mail address here.

Email: [freetext field]

May I contact you with further questions?

yes

no

Page 3, version 1: Questionnaire
Thank you very much!

You will receive information on the progress of this study at your e-mail address "[]".

Page 3, version 2: Questionnaire
Thank you very much!

You will **not** receive information on the progress of this study.