Chapter 3

Instrument Development

As alluded to in our introduction, our immediate purpose in Phase I was to find items measuring a prejudicial expectation that might act in place of experimenter-provided expectancies in racial-target person-memory research. At the same time, we hoped to screen-out prejudicial expectation items substantially affected by political attitudes and impression management. Items for future related research were likewise considered. To these ends, we proposed the following:

HYPOTHESIS

Hypothesis 1: Twenty items (from various scales) written specifically to assess prejudicial beliefs about Blacks (as unintelligent, undisciplined, ill-mannered or lazy) will load principally on a factor (referred to here as prejudicial expectation or PE) which is distinguishable – via factor analysis – from other facets of racism.

Hypothesis 2: Thirteen items written specifically to assess each subject's personal history of cultural/environmental or developmental reinforcement of racist judgements or comments (RRJC) will load principally on an RRJC factor that is distinguishable from other causes, correlates or facets or racism.

Hypothesis 3: Forty-eight items written to assess subtle behavioral or attitudinal correlates of racial beliefs will load to varying degrees on various facets of racism (segregationist attitudes, fear/dislike of Blacks, perceptions of historical inequities, anti-equal opportunity, etc.). The items less overtly racist in content will suffer less variance compression (floor effect) from socially-desirable response tendencies; however, theses items likely suffer greater confounding with other constructs (e.g., authoritarianism, neuroticism, political attitudes, self-esteem, etc.).

METHODOLOGY

Subjects: 397 White/Caucasian undergraduates completed the Phase I survey (described below) anonymously for extra credit in various psychology, sociology or human resource management classes.

Measurements: The Phase I survey was constructed by merging items from several popular scales targeting racism or ethnocentrism with items designed to measure neuroticism, need for cognition (nCog), impression management (IM), self-deception, and authoritarianism. Specifically, the scales used were the Need for Cognition Scale (Cacciopo and Petty, 1982) seen in items 4, 24, 35, 47, 55, 65, 84, 112, 131, 145, 185, and 197 of Appendix 1, the Self-Deception/Impression Management Scales (Paulhus, 1991, BIDR Version 6) seen in items 13, 15, 17, 34, 43, 69, 74, 80, 81, 88, 90, 93, 99, 100, 103, 109, 114, 128, 144, 157, 162, 167, 171, 172, 179, 186, 189, 190, 193, 196, and 199 of Appendix 1, the Neuroticism sub-scale of the NEO-PI (Costa and McCrae, 1985) seen in items 3, 6, 21, 23, 38, 46, 48, 49, 51, 53, 54, 57, 59, 61, 63, 64, 66, 70, 71, 82, 86, 91, 95, 97, 117, 118, 120, 121, and 130 of Appendix 1, the Right Wing Authoritarianism Scale (Altemeyer, 1981) seen in items 10, 12, 18, 20, 29, 58, 107, 113, 115, 151, 181, 188, and 195 of Appendix 1, the Modern Racism Scale (McConahay, et al, 1981) seen in items 31, 105, 174, 178, 180, and 184 of Appendix 1,

the Ethnocentrism/Negroes Scale (Adorno, et al, 1950) seen in items 159, 165, 169, 175, 182, and 187 of Appendix 1, the Traditional Racism Scale (McConahay, et al, 1981) seen in items 19, 101, 110, and 124 of Appendix 1, relevant items from the General Social Survey (Kluegel and Bobo, 1993) seen in items 33, 44, 67, 85, 108, 140, and 166 of Appendix 1, and relevant items from the National Election Studies Survey (Kluegel and Bobo, 1993) seen in items 56, 73, 75, 94, 106, 116, 127, 136, and 137 of Appendix 1.

To maximize the participants' perceptions of anonymity, the survey was limited to 200 items (so that one OPSCAN answer sheet would be sufficient — avoiding the need for subject ID numbers). This prevented the use of all of the above scales in their entirety. Roughly two-thirds of each published non-racism scale was applied. Nearly all racism items were applied. Since none of these scales specifically addressed developmental influences in the participants' personal histories, thirteen new items were written to assess cultural/environmental or developmental reinforcement of racist judgments or comments (RRJC). Three new items were also written specifically to assess a prejudicial belief in reliable/enduring differences between the races — especially in terms of intellectual differences. These three new items were combined with seventeen items (taken from the other scales) which also tapped into prejudicial beliefs in lower abilities or work ethics in Blacks (referred to here-in as prejudicial expectation or PE).

Table 1. Eigenvalues Associated with First Twenty Factors

Factor:	Eigenvalue:	Factor:	Eigenvalue:
1	10.9	11	1.5
2	5.4	12	1.4
3	4.7	13	1.3
4	3.8	14	1.2
5	3.0	15	1.0
6	2.4	16	1.0
7	2.0	17	.9
8	1.8	18	.9
9	1.7	19	.9
10	1.6	20	.9

In a more exploratory vein (to be pursued further in the post-dissertation efforts), forty-eight additional racism items were written, as well as sixteen more designed specifically to assess the degree to which a participant openly senses that they are afflicted by a racist bias in their own thinking. The forty-eight additional racism items were written with the goal of achieving greater subtlety, and thus, less variance compression in the response frequencies due to social desirability response sets. It seems fair to argue that in the times in which these older racism scales were written (some as old as the 1950's), some participants would be more comfortable identifying with a racism-indicating response. However, as the stigma of political incorrectness has attached itself to more and more attitudinal manifestations,

items on existing racism scales lose their subtlety. In other words, existing racism scales can easily become dated. Since this elusive subtlety was essential when our final short-form racism items were to be embedded in a much larger non-racism pre-lab (Phase II) questionnaire, it was deemed prudent to begin testing some new items in this first phase — if only in an exploratory sense. These new scales are found in Appendix 2.

Procedure: The data set from the initial subject pool of 480 undergraduates was sorted by self-reported race. For the immediate purposes of item selection for the second phase effort, only the responses provided by Caucasians were used — a total of 397 subjects. The factor analysis method used was SPSS's Principal Axis Factoring/Direct Oblimin (allowing non-orthogonal axes) — since correlations were anticipated among the constructs targeted by the survey items. Given the exploratory nature of this initial factor analysis, the degree to which these axes were allowed to be non-orthogonal was not restricted. A simple structure hinting at logical uni-dimensional sub-tests was the goal (especially regarding PE, RRJC, Right Wing Authoritarianism, and nCog – a construct central to the next phase of this research).

ANALYSIS

The first factor analysis run used the SPSS defaults for factor extraction (eigenvalues greater than 1.0) to support a scree analysis (see Table 1 below). The items included in this initial effort were those anticipated for use in Phase II and the initial confirmatory portion of the later phases (i.e., PE items, RRJC items, nCog items, IM items, authoritarian items and anti-racism items manifesting an intent to seek cross-racial relations). Analysis of break points in the scree analysis suggested factor extractions with the number of factors pre-set iteratively in the range of five through seven. Of these, seven yielded the simplest structure — primarily because the thirteen RRJC items always loaded into two separate RRJC factors. Evaluation of these items' response anchors and their associated frequency charts revealed the reason for this: five of these items used dichotomous response alternatives while the other eight used Likert scales. Of the five dichotomous items, only one produced any variance. The responses to the others were both identical (within subjects) and very nearly unanimous (across subjects). As such, they were dropped.

The factor pattern matrix was analyzed with the intent of identifying PE and RRJC items with minimal loadings on authoritarianism and impression management. One additional RRJC item was eliminated by this selection hurdle — as were five PE items (see Table 2). The thirteen remaining PE items were factor analyzed again to check for stability in their factor structure. The NEO-PI items were added into this structural check since the bulk of the deceptive (or cover) items used in the Phase II pre-test came from the NEO-PI. This check yielded a very similar factor structure.

Table 2. Prejudice Item Factor Loadings (suppressed at .19 to show simple structure)

Item	Fac. 1 Prej.	Fac. 2 nCog	Fac. 3 IM	Fac. 4 RR/d	Fac. 5 RRJC	Fac. 6 RWA	Fac. 7 Anti-R
MRS_184	.66						
NES_116	.61					.26	
NES_106	-.57						
GSS_33	.56				.22		
MRS_105	.55				.22		
GSS_67	.53				.23		
GSS_108	.49						-.19
NES_56	.49					.20	
ETH_187	.49	.19			.26		
GSS_140	-.48				.20		
GSS_85	-.47				.24	-.23	
ETH_175	.45	.20			.35		
R_168	-.44				.21	-.30	
GSS_44	.43				.29		
NES_73	.43				.32		
TRS_124	.41				.40		
R_139	.41					.19	
ETH_165	.37	.22			.21		
NES_94	.34				.32		
R_36	.21						

The items required for near-term applications in upcoming phases (PE and RRJC) were then prepared for analysis via Item Response Theory (IRT). First, the PE items were factor analyzed separately, as were the RRJC items. An unrotated Principal Axis Factoring solution was used to confirm that each set of items ostensibly measures a unitary dimension. The scree analysis confirmed the existence of a dominant first factor for each of these two item sets — especially PE (see Table 3).

BILOG was used for the IRT analysis. Since it is intended for use with dichotomously scored items, the item response data was converted via Excel into dichotomous data (zeros and ones where "one" represents an affirming response for the trait). Negatively-keyed items were flipped in valence as appropriate (these items had negative loadings in the factor analysis as expected). The item analysis results for the thirteen PE items are shown in Table 4 (the dichotomous cut-off used to generate the IRT parameters follows the item number).

Table 3. Eigenvalues for First Eight Factors for Examining Uni-dimensionality

PE		RRJC	
Factor:	Eigenvalue:	Factor:	Eigenvalue:
1	5.6	1	3.1
2	1.2	2	1.2
3	1.1	3	1.0
4	.9	4	.8
5	.8	5	.7
6	.7	6	.6
7	.6	7	.5
8	.6	8	.5

DISCUSSION

The existing racism scales included (but did not differentiate) items assessing fear of Blacks, segregationist attitudes (regarding dating, marriage, schools, neighborhoods, etc.), perceptions of unfairness in public policy, perceptions of minority responsibility, perceptions of historical inequities, as well as measures of prejudicial beliefs (including beliefs in inferior intellect, inferior work ethics, inferior preparation for mainstream culture, etc.). Though the factor structure or dimensionality of these existing scales may have changed over the years or decades of their existence, such a possibility did not motivate our item analysis efforts. Our intent was to identify a sub-set of items targeting beliefs of race-based inferiority and distinguish

Table 4. Item Analysis Results for 13 Surviving PE Items

Item/ Dichot. Cut Pt.	Item-Total Corr.	Discrimination Slope (a)	Difficulty Threshold (b)	Pseudo-Random Guessing (c)	Mean Response	Std. Dev.
GSS_33/1	.76	1.37	.54	.0013	1.2	1.48
R_36/3	.43	.44	-1.26	.0048	3.0	1.49
GSS_44/1	.67	1.60	.67	.0035	.5	1.07
GSS_67/2	.71	.98	.22	.0033	1.6	1.58
NES_73/1	.67	1.6	.73	.0045	.6	1.09
NES_94/2	.62	.95	.24	.0033	1.6	1.49
MRS_105/2	.72	.99	.63	.0025	1.2	1.33
TRS_124/1	.70	1.4	.29	.0033	.8	1.27
GSS_140/2	.34	.39	.39	.0048	2.5	1.48
ETH_165/1	.61	.94	.38	.0030	.8	1.22
ETH_175/1	.71	1.4	.81	.0046	1.0	1.23
MRS_184/2	.70	.92	.40	.0027	1.4	1.34
ETH_187/1	.70	1.2	.64	.0016	1.1	1.26

them from items targeting reinforcement for racist judgments or comments. It was hypothesized that logically-related items would load primarily on the two factors PE and RRJC, and that these two factors could then be distinguished from other facets of racism. Given the exploratory nature of the effort, the outcomes were reasonably consistent with our expectations.

To meet our immediate needs (i.e., support of Phase 2), we focused further efforts only on the PE items. Although we were able to identify PE items with minimal loadings on impression management and right-wing authoritarianism, the surviving items provide information in roughly the same levels of the PE trait dimension (see the difficulty thresholds in Table 4 above). Unfortunately, the only exceptions to this effect came from items offering very little discriminating power (R_36 and GSS_140 – with the latter spreading its information across the range of prejudice). As a result, the standard error function is not as "reliability-friendly" as hoped for across the common range of prejudice.

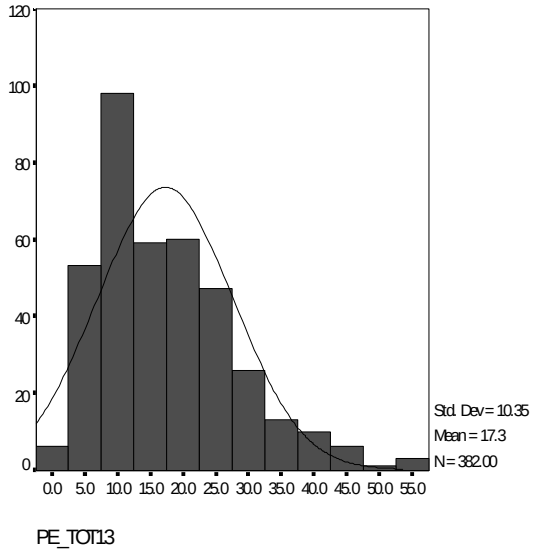

Figure 3. Prejudicial Expectation Scores (Likert scores summed on 13 items for 397 Caucasians)

A quick look at Table 4 reveals further justification for eliminating these two items: they have — far and away — the lowest item-total correlations. This is probably related to the arguments behind our second and third instrument-related hypotheses. We're not presenting extensive analysis results (within this dissertation) relative to these hypotheses — since they deal with issues related largely to our post-dissertation effort. However, some comments are in order to the degree they relate to PE item selection. Recall we predicted that "items less overtly racist in content [would] suffer less variance compression (i.e., floor effect) from socially-desirable response tendencies" (p. 30). The two items just mentioned appear to support this prediction. Unfortunately, these items instead suffer a much lower item-total correlation. As we also predicted, these "less overtly racist" items likely suffer from multi-dimensionality. In other words, as items become less overtly racist in content, they're more likely to tap into other attitude-related constructs. One consequence of this problem would be lower item-total correlations.

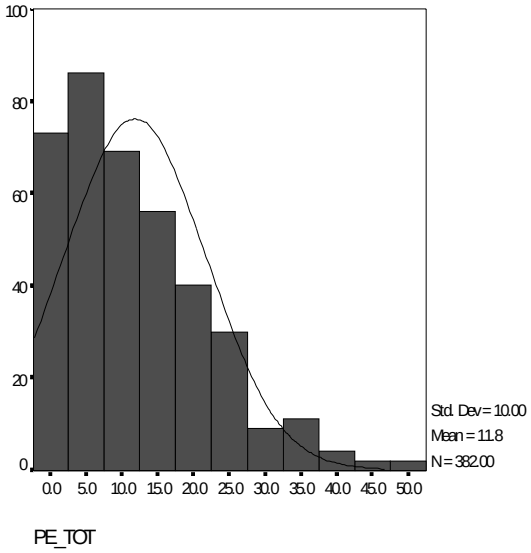

Figure 4. Prejudicial Expectation Scores (Likert scores summed on 11 items for 397 Caucasians)

Less floor effect allows an item to tap racial prejudice at subtler (i.e., lower) levels. Unfortunately, our most sensitive items do so with diminished precision. In other words, although these items provide information at lower trait levels, they provide very little discriminatory power at *any* level. We believe this is related to the trade-offs alluded to in Hypothesis Three. Here we suggested that impression management (IM) factor loadings will be negatively correlated with mean responses on our PE items. In other words, applying our subtle tangential approach, what we profit in avoiding IM we may lose in racism specificity. Though this will be more appropriately done via a post-dissertation confirmatory factor/LISREL analysis, we did find this kind of relationship using exploratory factor loadings (but using these two factors alone). Specifically, the correlation between PE item mean responses and IM factor loadings for these items was -.743 ($p < .01$, though with one very influential outlier removed, this dropped to $r = -.401$, $p = .11$). This outcome foretold a likely range restriction problem at the bottom end of our principal individual difference measure: prejudicial expectation (see Chapters 5 and 6 for the implications of this floor effect). This range restriction becomes more apparent when the two items least afflicted with IM are removed to create our final scale (compare Figures 3 and 4). We re-checked the factor structure after removing these items. The ratio of the first two eigenvalues (5.4 to .9) remained supportive for our assertion of apparent uni-dimensionality.

The range restriction was less of a problem when our instrument development subjects (N=397) completed the items with total anonymity assured (typically in large groups with no ID numbers of any kind). The subjects we later used in our 90 to 120 minute laboratory phase (N=79 survivors) turned out to be another matter. IM was likely to be more of a problem because we had to code random ID numbers on the response sheets (to match subjects' responses with their other laboratory-based behaviors). Their anonymity was truly protected, but they were probably a little suspicious of this procedure. In any case, the smaller groups in the computer lab setting and the intensity of the exercise may have increased self-consciousness. McCrae (1997) has demonstrated that this laboratory effect can temporarily encourage subjects to minimize their negative opinions about negatively-stereotyped groups. This would probably move people toward lower apparent scores on PE.

Regardless of cause, we suffered such consequences (compare Figures 4 and 5). This was likely to be behind some of the apparent chaos in our ultimate scatterplots (again see Chapter 5).

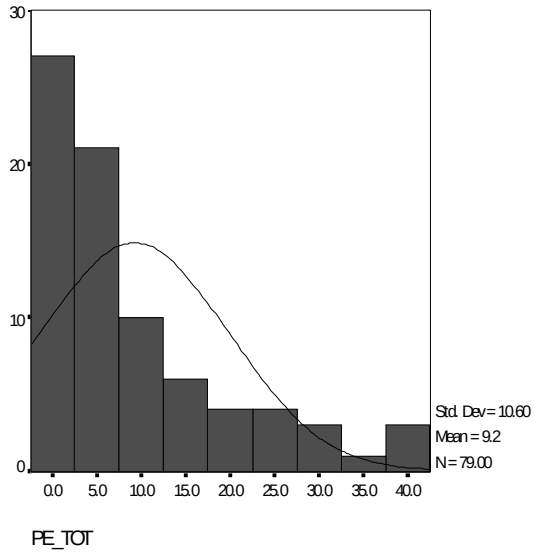

Figure 5. Prejudicial Expectation Scores (Likert scores summed on 11 items for 79 survivors)