

Improving Cross-National Survey Research
by Measuring the Intensity of Response Categories

Tom W. Smith

National Opinion Research Center
University of Chicago

February, 1997

GSS Cross-National Report No. 17

This research was done for the General Social Survey project directed by James A. Davis and Tom W. Smith. The project is supported by the National Science Foundation, Grant No. SES-9122262.

Introduction

Scientific research rests on the reliable and consistent measurement of phenomenon. In cross-national or cross-cultural survey research between countries or social groups that speak different languages the goal of replicative measurement is greatly complicated by the necessity of designing and administering questionnaires in two or more languages. Only by assuring that the items in each language and questionnaire have equivalent meaning and response scales can comparable measurement be obtained and valid inferences be drawn. But the complexity of both survey measurement and of languages makes the goal of equivalency an extremely difficult challenge (Kumata and Schramm, 1956; Scheuch, 1989; Smith, 1988; Van de Vijver and Leung, forthcoming).

Each question has two parts: 1) the point of the inquiry or substance of what is being asked about and 2) the implicit or explicit categories in which the response is requested. When the question is open-ended, the requested response is unstructured (e.g. "What is the most important problem facing the country today?" and "Why did you vote for Bill Clinton for President?"). But most survey questions are closed-ended with an explicit set of response categories or some type of response mechanism described (e.g. "If you were to consider your life in general these days, how happy or unhappy would you say you are on the whole: Completely happy, Very happy, Fairly happy, Not very happy, or Not at all happy?" and "Do you favor or oppose the death penalty for people convicted of murder?").

While there are effectively an unlimited number of subjects that questions ask about (and an wide variety of ways of asking about each subject), survey researchers tend to use a much smaller number of response categories in their questions. As Davis' review (1993) of 301 questions on the 1985-1993 International Social Survey Program (ISSP) modules showed, several response scales were repeatedly used. For example...

<u>Scale</u>	<u># of Items</u>
Agree strongly/Agree/Neither agree nor disagree/ Disagree/Disagree strongly/Can't choose	92
Essential/Very important/Fairly important/ Not very important/Not important at all/Can't choose	26
Definitely allowed/Probably allowed/Probably not allowed/Definitely not allowed/Can't choose	22
Strongly in favor of/In favor of/Neither in favor of nor against/Strongly against	11
Very important/Important/Neither important nor unimportant/Not important/Not important at all/ Can't choose	9

Moreover, not only are the same exact scales utilized again and again, but certain terms tend to be repeated across scales. Note,

for example, the use of "very" and "important" in the second and fifth examples above, of "strongly" in the first and fourth examples, and of "can't choose" in all but the fourth example. Thus, by focusing on the response-scale part of questions, one deals with a set of measurement and translation issues that have widespread application across questions and surveys.

In addition, most survey scales seek to arrange responses along a underlying continuum such as agreement/disagreement, importance, allowance, being in favor of/against, etc.¹ By assessing the position of each response category on the underlying continuum, the intensity of the response is determined. If this is done for items in two languages, it becomes possible to determine the equivalency of the individual response categories and ultimately of the response scale as a whole. The task then becomes to develop a method for assessing where categories fall on a response continuum.

This paper will examine 1) how response categories influence the reported distribution of results, 2) how to measure the intensity of response categories, 3) results from American and German pilot studies of response scales, and 4) the implications of these results for cross-national research.

Response Scales and Reported Distributions

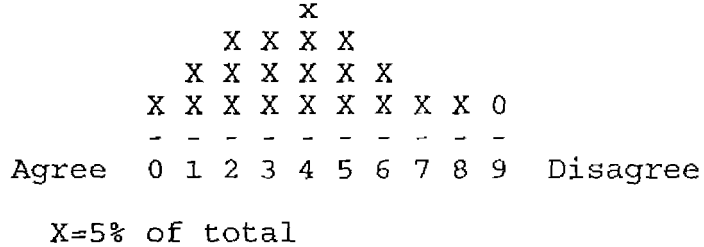
Reported distributions are a function of a) the true distribution of attitudes in the population and b) measurement properties of the response scale. How much of an underlying distribution is captured by a given term/category is a function of a) the underlying distribution and b) the number, intensity, positioning, and intervals between the scale points utilized. In general, a) the more the number of points used, the less of the distribution will be captured by a particular point, b) the closer two points are in intensity, the less of the distribution will be captured by each individual point, c) broader terms may capture more of the distribution than narrower terms (i.e. its not only the mean intensity of a term, but its range that determines how much of the distribution will be covered), and d) adding a new, more intense point to a scale can change how the previous end point was understood and alter (and typically increase) the share of the distribution captured by the displaced endpoint. The effect seems to be that some people avoid "extreme" categories where extremity is based on a category representing the end or extreme position on a scale, not on the extremity of the term actually used to express the scale point.

To illustrate these points, let us start with the simplest case of a dichotomy: agree/disagree. Given the hypothetical distribution of attitudes in Figure 1, the reported distribution would be about 65% agree, 35% disagree. Now suppose a third category

¹Nominal scales do not do this, but these are relatively rare in attitudinal scales.

Figure 1

Hypothetical Distribution



"neither agree nor disagree" was added and that half of the people closest to the mid-point (4.5) were attracted to this mid-category. The revised distribution would be agree 55%, neither agree nor disagree 17.5%, and disagree 27.5%. Next suppose that "agree" was replaced with two categories "completely agree" and "somewhat agree". If "completely agree" was at point 0 and "somewhat agree" at point 3, then the new distribution might be 15% completely agree and 40% somewhat agree. But suppose the two new categories were "completely agree" and "strongly agree" with the former at point 0 and the latter at 2. The distribution would be something like 10% completely agree and 45% strongly agree. But if "strongly agree" was added as a third new category on the agreement side, then the distribution might become completely agree 10%, strongly agree 20%, and somewhat agree 25%. But if "completely agree" was then dropped, then the distribution might become 30% strongly agree and 25% somewhat agree.

Or assume that "Slightly agree" was added and represented at 4. It might not only take over much of the somewhat agree cases, but draw in some of the distribution from neither agree or disagree. Along with a matching "slightly disagree" category these might bring back in say half of the distribution lost to the middle category above leaving 59.375% in the combined agree categories, 8.75% in neither agree nor disagree, and 31.875% in disagree. Thus, assuming a fixed true distribution, seven different response scales ranging from two to seven categories, and the simplest of rules for allocating cases, there is considerable variation in distributions reported. For example, the % agreeing varies from 55% to 65% and strongly agreeing from 20% to 45%.

Now consider what the impact might be of using two scales to measure two different populations with the same true distribution of an attitude as in Figure 1. In population A the completely agree/somewhat agree/neither... scale was used with the resulting distribution of 15%, 40%, 17.5%, 25%, and 2.5%. In population B the strongly agree/somewhat agree/neither... scale was employed and the distribution was 30%, 25%, 17.5%, 17.5%, and 10%. Now assume that population B was interviewed in another language and the researcher was told that the second (non-English) scale was a translation of

and equivalent to the first (English) scale. Comparing these two scales using the typical values of 1-5 for the five categories one would draw the conclusion that there was more agreement in population B than in population A (means respectively of 2.5 vs. 2.6) and that there was much more extremity in population B than in A (1+5 = 40% vs. 17.5%). Neither conclusion would be correct, but merely the artifact of mistranslations and/or misinterpretations of scales in two languages.

Measuring the Intensity of Response Categories

There are several ways to measure the strength of response categories along an underlying response scale. One approach is to have respondents rate the strength of terms defining each point on the scale. There are three standard variants of this approach.

First, one can rank the terms from weaker to stronger (or from less to more or along any similar continuum) (Spector, 1976). This of course only indicates their relative position and not the absolute strength or distance between terms.

Second, one can rate each term on a numerical scale (usually with 10 to 21 points) (Wildt and Mazis, 1978; Worcester and Burns, 1975; Myers and Warner, 1968; Cliff, 1959; Jones and Thurstone, 1955; Mosier, 1941; Vidali, 1975; Mittelstaedt, 1971; Bartram and Yelding, 1973; Traenkle, 1987). This allows the absolute strength or distance between each term to be known and thus facilitates the creation of equal interval scales. It is also possible to use an alphabetical scale or unlabelled spaces, rungs, or boxes as in a semantic differential scale (Osgood, Suci, Tannenbaum, 1957). The letters or spaces are then transformed into their numerical equivalents.

Finally, magnitude measurement techniques can be used to place each term on a ratio scale (Lodge, et al., 1975, 1976, 1979, 1981, 1982; Hougland, Johnson, and Wolf, 1992). The magnitude measure techniques gives an arbitrary value to a reference term and has respondents rate other terms as ratios to the base term. This allows more precision than the numerical scale approach (since the terms are not constrained by the artificial limits of the bounded number scale).

Of these three variants the middle seems most useful. On the one hand, the ranking method fails to provide the numerical precision that is necessary to calibrate terms across languages. On the other hand, the magnitude measurement technique is much more difficult to administer and much harder for respondents to do (about 10-15% seem unable to master the procedure). In addition, the extra precision that the magnitude measurement procedure can provide over that achievable using a 21-point scale approach does not appear to be needed.

The direct rating approach has been used to rate words along various dimensions. Of most interest to us are those that either rate terms along a general good/bad or positive/negative dimension or which rate the intensity of modifiers (Wildt and Mazis, 1978; Worcester and Burns, 1975; Myers and Warner, 1968; Cliff, 1959;

Jones and Thurstone, 1955; Mosier, 1941; Vidali, 1975; Mittelstaedt, 1971; Bartram and Yelding, 1973; Lodge, et al, 1975, 1976, 1979, 1981, 1982; Hougland, Johnson, and Wolf, 1992). Similarly, other studies have rated probability statements (Wallsten, Budescu, Rapoport, Zwick, and Forsyth, 1986; Lichtenstein and Newman, 1967); frequency terms (Spector, 1976; Schaeffer, 1991; O'Muir-cheartaigh, Gaskell, and Wright, 1993; Strahan and Gerbasi, 1973, Bradburn and Sudman, 1979; Schriesheim and Schriesheim, 1974; Hakel, 1968; Simpson, 1944); and terms used in reported to describe percentages from public opinion surveys (Crespi, 1981 and "RAC...", 1984).

The studies generally show that a) people (usually college students) can perform the required ratings tasks,² b) ratings and rankings are highly similar across different studies and populations, c) there is high test/retest reliability, and d) several different treatments or variations in rating procedures yield comparable results. Thus, the general technique seems robust and reliable.³

A second approach for assessing the intensity of scale terms and response categories is to measure the distributions generated by using different response scales (Smith, 1979; Laumann, Gagnon, MacKuen and Turner, 1984; Michael, and Michaels, 1994; Hougland, Johnson, and Wolf, 1992; Orren, 1978; Sigelman, 1990). In an experimental, across subjects design, one random group is asked to evaluate an object (e.g. presidential popularity or one's personal happiness) with one set of response categories and a second random group evaluates the same object with another set of response categories. Since the stimulus is constant and the sub-group assignment is random, the number of people attracted to each category will depend on the absolute location of each response category on the underlying continuum and the relative position of each of the scale points adopted. With some modelling around what the two observed distributions suggest are the underlying distribution, it is possible to estimate at what point each term is cutting the underlying scale (Clogg, 1982; 1984).

The alternative version uses a within subjects design in which people are asked the same question (i.e. presented with the same stimulus) two (or more) times with different response categories being used (Orren, 1978). This differs from a test/retest

²While reassuring, other studies show that various measurement artifacts can influence responses to numerical scales (Wilcox, Sigelman, and Cook, 1989; Smith, 1993; Schwarz and Hippler, 1995; Schwarz, Hippler, Deutsch, and Strack, 1985; and Schwarz, Knaeuper, Hippler, Noelle-Neumann, and Clark, 1991). See also, O'Muircheartaigh, Gaskell, and Wright, 1993.

³An exception is that vague frequency terms correspond to different absolute values depending on the commonness or rarity of the specified event or behavior. Thus, people who "usually" vote once a year, but people who "usually" dine out more than once a week (Schaeffer, 1991; Bradburn and Sudman, 1979).

reliability design in that a) the measurement instrument is not constant (since the response categories differ) and b) the two administrations are essentially consecutive without any intervening time and/or buffer tasks. This provides additional information since it allows the direct comparison of responses, but the initial evaluations may artificially influence responses to the later scales (e.g. a person may feel constrained to choose the same response in terms of position or term on a subsequent administration as used on the first administration).

The advantage of the distributional approaches is that they ask respondents only to do what they are normally required to do - answer substantive questions with a simple set of response categories. The disadvantages are that a) it is harder to access a large number of response terms and thus is better suited for assessing a discrete response scale already adopted than for evaluating a large number of terms that might be utilized in possible response scales,⁴ b) results will depend on the precise underlying distribution and the modelling procedures adopted, and c) it creates more work for the analysts since the strength of terms must be indirectly estimated from the distributions rather than directly calculated from respondent ratings.

Because the direct rating approach provides the quantified intensity scores needed in the most straight-forward manner, this was adopted as the main technique. In addition, there may be context effects in the rating of the intensity of terms. For example "very" may be rated more intensely if it was the first strong term presented than if it followed other stronger terms (e.g. completely, extremely). Context effects have generally not been searched for in this line of research, but the randomization of order in several studies has tried to average out any such effects. This latter approach is generally utilized here, but an ordered vs. not ordered experiment is also included.

American and German Pilot Studies

Pilot studies were carried out in the United States and Germany to use the above approach to evaluate the translation and equivalency of response scales. The American pilot study was carried out on a quasi-representative sample of adults living in households. Ten sample points were selected to represent all four Census regions (West, South, Midwest, and Northeast) and three size of place strata (central cities, metropolitan areas outside of central cities, and non-metropolitan areas). Interviewers had

⁴It would be possible to evaluate more terms using more random sub-groups, but in order to maintain the same level of precision this would mean increasing the sample size. Similarly, the same people could be asked many repetitions of a question with different response scales, but this would soon become tedious and later repetitions would probably be distorted by the previous administrations.

quotas to fill based on gender, age, and employment status. They proceeded through neighborhoods in the selected communities until the quotas were completed. The study was designed and carried out by the National Opinion Research Center at the University of Chicago. Besides representing the adult population of United States on the stratification and quota variables (region, size of place, gender, age, and employment status), the sample is also representative on race and marital status. The sample does underrepresent the less educated segment of the population (less than a high school degree: pilot study 6%, General Social Survey 17%). Interviews were conducted in July/August, 1995. A total of 119 interviews were collected, but two were lost in the mail for a final total of 117.

The German pilot study stratified the country by states (Bundeslaender) and city size (cities over 100,000 vs. else). Within these areas interviewers filled quotas based on gender, age, and education. The study was designed and supervised by the Zentrum fuer Umfragen, Methoden, und Analysen, Mannheim, and interviewing was conducted by Infratest - Burke Sozialforschung, Munich. The sample closely matches German Census figures on gender, age, and education. Fieldwork was carried out in September, 1995. A total of 221 interviews were conducted.

American Results

In the pilot study attempts were made both to assess the intensity that people assigned to particular terms and therefore response categories and to evaluate the meaning of the underlying continuum on which intensity was being measured. First, people were asked to rate the intensity of 27 phrases on a 21-point agree/disagree scale (See Qs A3 and B3 in the Appendix). Item order was randomized by sorting cards containing the phrases, except for "basically" which was the first term rated by each respondent. Table 1 gives the means and standard deviations for the terms.⁵ In

⁵On a scale-by-scale basis cases were excluded from the analysis that failed to carry out the ratings adequately. This excluded respondents who refused to do items, those with high item non-response, those who could not consistently associate terms with the proper pole, and those showing peculiar response patterns. People were not excluded for a few unusual responses, but for incomplete and erratic responses to the scale as a whole. There were 14 exclusions for Q.3 (agree/disagree), 5 for Q. 4 (important/unimportant), 10 for Q. 5 (in favor of/against), and 12 for Q. 6 (ranges on agree/disagree). Overall, there were 15 respondents who were excluded for two or more individual scales. Exclusions were significantly associated with interviewer assessments that respondents misunderstood the word rating tasks and that these tasks were difficult. (The interviewer evaluation questions were "How was the respondent's understanding of the word rating tasks? Completely understood/Mostly understood/Mostly

terms of magnitude and relative position the terms array themselves just about exactly as one would expect.⁶ The 11 agree terms run from "agree a little" at 12.1 to "completely agree" at 19.4. The four mid-point or uncertain terms are from 10.1 to 9.6. The 11 disagree terms range from "disagree a little" at 7.1 to "completely disagree" at 0.8. In addition, "not agree" exactly matches "disagree" at 3.5.

Standard deviations follow a wave pattern. They are small near the extremes, increase as intensity moderates, and then decrease to their lowest level for the two mid-point categories (in the middle and neither agree nor disagree). The lower range for the categories near the extremes (strongly agree/disagree) is only partly a function of floor and ceiling effects resulting from respondents rating the terms at or near the endpoints. The unbounded end of the range is usually a little smaller than that for broader and more moderate terms. For example, the average upper range for strongly disagree is + 1.6 compared to + 2.2 for disagree, while the average lower range for strongly agree is - 1.9 compared to - 2.2 for agree. Thus these terms appear to have more precise and limited meanings not only because of floors and ceilings, but also because their greater intensity also narrows people's understanding of their meaning. The standard deviations narrow for the middle categories because people have a clear and consistent understanding of what the mid-point of a scale is. The uncertain terms "can't choose" and "undecided" are also placed near the middle, but the

misunderstood/Completely understood" and "How hard were the word rating tasks for the respondent? Very difficult/Somewhat difficult/Somewhat easy/Very easy.") Exclusions were also higher among the less educated although only the association with Q. 5 was statistically significant.

⁶On the agree/disagree rating scale the questionnaire was handed back to respondents after the question was completed and told by the interviewer "Please look over your answers. If you want to change any of your responses, indicate in the right hand column, the one headed "CHANGES," what number you now want to give a phrase." Respondents to later questions were not given a chance to review their responses, but at any point while a question was being administered a respondent could change a response. Changes were fairly rare. 62.4% made no changes, 17.0% 1-2 change, 14.6% 3-6 changes, and 6.0% 7+ changes. On average 1.7 changes were made among the 28 phrases rated.

Two type of changes were common. First, there minor upward or downward adjustments to have responses better fit in with other phrases being rated. Second, there were pole corrections when respondents realized they had oriented their response to the wrong end of the scale. These usually resulted in large changes (e.g. from 2 to 18). In almost all cases, the changes moved answers towards the modal response.

standard deviations are a bit higher because some people wanted to rate them as off-scale and gave some different responses such as 0 to try to convey this idea. (In addition, a few more people than for the other terms did not rate these terms for the same reason.)

Second, a similar exercise was carried out on two important/unimportant scales (Qs. A4 and B4 in Appendix). As in the case of the agree/disagree scale, order of presentation was randomized by sorting. Table 2 reports the means and standard deviations. On one half of the sample people rated terms on a unipolar scale measuring degree of importance and on the other half on a bipolar scale of important/unimportant. The unipolar scale ran from 19.4 for "extremely important" to 1.4 for "not at all important." The bipolar scale extended from 19.4 "extremely important" to 0.8 for "extremely unimportant." On this scale middle terms were placed very near the mid-point ("in between" = 10.0; "neither important nor unimportant" = 9.5). There were 15 "important" terms that were rated on both scales. In 13 of these cases the terms were rated somewhat higher on the bipolar scale than on the unipolar scale. It appears that on the important/not important scale people adjust terms down towards the not important end of the scale. For example, "neither important nor unimportant" is scored at 9.0 instead of the mid-point of 10.0. This indicates that "unimportant" defines a more extreme position than the lack of importance does. The latter is seen by at least some as indicating the absence of importance rather than the presence of unimportance.

Standard deviations are smaller for high terms on both the unipolar and bipolar scales, but the pattern is less clear at the lower end of these scales. "Not important" terms have the largest standard deviations of all terms on the bipolar scale and on the unipolar scale they have among the largest values. Some negative phrases tend to confuse people in general (Smith, 1995) and especially on the bipolar scale people were less sure where to rate these terms visa vie the "unimportant" terms.

Third, Table 3 rates another set of terms, "against/in favor of," and also carries out an order experiment (Qs. A5 & B5). In terms of the means and standard deviations both orders are similar to each other and to the pattern shown with "agree/disagree" in Table 1. In particular, the means have magnitudes and relative positions as one would expect and the standard deviations show the same wave pattern of going from small for extreme terms to larger for more moderate, general terms and then smaller for the middle term.

The order experiment did however reveal a decided difference in terms of the consistency of ratings. On the version that arranged items in ascending order from "strongly against" to "strongly in favor of" as they would be present as part of a response scale, 72% of people presented all seven items in ascending order without any inconsistency. On the version that presented the terms in a fixed, unordered sequence only 37% of respondents rated all seven items in ascending order. This indicates that presenting the items in ascending order, as they are presented as actual response scales, provides people with

additional information and constrains how people perceive and evaluate the terms. When terms are organized as a scale, people are more likely to perceive and treat them as such.

Fourth, Table 4 shows that the values assigned to terms at both ends of the scales for agree/disagree, important/unimportant, and against/in favor of (Table 1-3) are highly symmetrical. The first column gives the mean rating for each term when associated with the positive/top end of the scale. The second column gives the rating when used in conjunction with the lower end of the continuum. The third column reverses the numbers in the second column to show what they equal if rated at the opposite end. Comparing the first and third columns show how similar and symmetrical the ratings are. With one exception, all terms rated at the positive end practically match how they are rated at the negative pole. This indicates that people assign these terms a consistent value regardless of their positive or negative orientation.

Fifth, Table 5 shows that terms are also rated in a highly similar manner when the underlying continuum varies. Part A indicates that terms rated on the agree/disagree and important/unimportant scales have highly consistent values. Part B reveals that terms rated on agree/disagree and in favor of/against are also quite similar. Along with the results from Table 4, this indicates that ratings are robust and that terms probably have similar intensities across various scales.

Sixth, the rating of scales are also quite stable across sub-groups. Sub-group differences were examined for all items rated in the agree/disagree and important/unimportant scales. Differences by gender, age, education, and race were examined. While a few statistically significant results emerged, there were no consistent differences either across samples or demographics. Education showed the most significant differences (6 of 43) Pearson's correlations, but only one significant one-way analysis of variance. The education effects that do appear seem to be related to the greater difficulty of less educated respondents in carrying out the rating task, rather than to systematic differences in the meaning of terms.

Finally, Table 6 examines intra- and inter-respondent variability in the rating of terms. Intra-respondent variability was measured by selecting eight terms rated on the agree/disagree scale and reminding people what score they had assigned to the terms. Next, people were asked what was the lowest value they would accept for the term and what was the highest (Qs. A6 & B6). If they thought no variation from their assignment was acceptable, then that same value was entered as the minimal and maximum score for the term and the acceptable range was 0. The first column shows the mean interval between the top and bottom values. First, these values follow the wave pattern described earlier for the standard deviations (which are presented in the second column for comparison). Acceptable ranges are narrow at the extremes and at the middle and widest between the middle and extremes. Second, the ranges are almost perfectly symmetrical with strongly agree/

strongly disagree and agree/disagree showing the same means. Third, most people see these term as somewhat malleable. They do not believe that the terms have only a precise and invariant value (like agree=16.1), but see terms covering a range of values (e.g. 14-18).

Next, assessments were made of the meaning of the underlying dimension on which the above terms were arrayed. First, the similarity of different pair of words were examined. In Table 7 five pair of words were compared with the pair "agree/disagree" (Qs. A8 & B8). People evaluated how similar the "agree/disagrees pair was to each of the other pairs. "For/against" and "favor/oppose" were considered to be the most similar, "positive/negative" the next closest, and "like/dislike" and "important/unimportant" the least alike. This indicates that "agree/disagree," and "for/against" and "favor/oppose" come closest to tapping a similar underlying dimension, while the other pairs define more distinct continuums.

Then, the similarity of other terms to those used in the "agree/disagree" ("agree," "neither agree nor disagree," and "disagree") and "important/unimportant" ("important" and "unimportant") dimensions are assessed by an open-ended item that asks people to define these terms.⁷ Table 8 lists the terms offered to define "agree." The list basically includes synonyms along with repetitions of "agree" itself. The use of this list will be discussed in the comparative section below.

America and Germany Compared

The preceding analysis indicates how useful the evaluation of the response terms are for understanding response scales in general. Here the use of this information for comparing scales in two countries and languages is considered. Tables 9 and 10 show that overall there is a high correspondence between the agree/disagree and important/unimportant scales in the United States and their counterparts in Germany. Table 9 presents the mean ratings for the agree/disagree and the two German counterparts *stimme zu/lehne ab* and *stimme zu/stimme nicht zu*. The American scores correlate almost as highly with both German scales (respectively $r=0.993$ and 0.986) as the two German scales associate with each other ($r=0.995$) and most means are close and not statistically different from one another (Mohler, Harkness, Smith, and Davis, 1996). Despite this extremely high correlation and the general correspondence in scale scores, there are some important differences in the mean values. First, the base words (e.g. agree, *stimme zu*, disagree, *lehne ab*, etc.) have more extreme meanings in German than in English. For example agree is 16.1 in English and

⁷The definition tasks were found to be fairly hard by many people. A number of interviewers noted in the evaluation section that particular people had problems expressing themselves and often used the word itself as part of the definition.

stimme zu 17.4-17.5 in German. Second, "definitely" is a stronger term in English than "bestimmt" is in German. Third, while "strongly" is a weaker term in English than either "completely" or "definitely," this does not appear to be the case in German where "voll und ganz" shows up as the strongest German term. Fourth, while "a lot" is an intensifier in English, "ziemlich" is a deintensifier in German. In the ZUMA survey 58-85% of respondents rated it less strongly than they rated the base words (stimme zu, lehne ab, stimme nicht zu). It maybe that "ziemlich" is not an appropriate translation of "a lot."

Table 10 shows the mean scores for important/unimportant and wichtig/unwichtig. As before the cross-national scale scores correlate very strongly ($r=0.987$) and most means are quite close. But again there are some notable differences. First, as in previous comparisons, the base words are stronger in German than in English (e.g. unimportant=3.6 and unwichtig=2.2). Second, "definitely" again shows up as stronger than "bestimmt."

The difference in the intensity of base terms may be a general difference between English and German. The pattern appears not only for agree/disagree and important/unimportant (See Tables 9 & 10), but also for in favor of/against.

Next, Table 11 shows the frequency of English terms used to define "agree" and German terms used to define "stimme zu." The next step in the analysis is to take each English term and translate it into German and each German term and translate it into English. "Agree" and "stimme zu" will be judged to mean the same thing to the extent that a) the German terms offered as meaning "stimme zu" match the German terms translated from English terms used to define "agree" and b) the relative frequency of these terms are similar. Perfect correspondence would involve only matched terms appearing and in the same proportions. While the detailed analysis has not been carried out, the terms in Table 11 clearly show both much overlap and some distinctions. For example, "bejahe," etc. means "accept," "give consent" both of which appear among the English terms and translations of "approve," etc. include "zustimmung" which is among the German terms offered. But "positiv" in German is equivalent to "positive" in English and this term is not mentioned in the American survey.

Implications and Future Research

First, in general the comparison of American and German results on the agree/disagree and important/unimportant scales indicate a close, but not perfect correspondence between the scale terms in general and in particular for terms used in prior ISSP scales (e.g. the five point agree/disagree scale). Some scale disparities do exist and the above rating scores could be used to suggest the use of alternative terms in future response scale or the adjustment of past scales according to their position on the underlying continuum.

Regarding the latter, attitudinal scales are often used in analysis as if they were interval scale with equal distances

between each response. For example, a five-point agree/disagree scale will be used in analysis with the response points assigned values of one to five. But the above analysis indicates that the response points do not have equal intervals between them. For example, the scores on the American five-point agree/disagree scale are 18.8, 16.0, 9.9, 3.5, and 1.5 and the intervals are 2.8, 6.2, 6.4, and 2.0. To estimate the impact of these miscalibrations, 16 agree/disagree items on the 1991 ISSP religion module and 18 agree/disagree items on the 1993 ISSP environment module were inter-correlated with themselves and five demographics (gender, age, years of education, highest educational degree, and frequency of church attendance) using both the raw 1-5 scale and the 18.8-1.5 scale.

Overall, there was little difference in the raw or adjusted correlations (Table 12). What impact there is is for the adjusted correlations to decrease. This may mean that the raw scale scores apply more regularity to attitudes than really prevails so that the adjusted figures show the marginally lower and truer associations. Alternatively, when presented with the terms as a scale, people may assign them equal distances and shift from scale independent evaluations of the response terms to more ordered, scale dependent assessments. This would mean that the scale adjustment would be less than optimal since respondents had already self-adjusted their responses.

Second, these results offer some tentative ideas about what kinds of scales might produce more equivalent, cross-national comparisons. Symmetrical, bipolar scales with an explicit middle point are probably best for cross-national scales. First, people have a very clear understanding of what is the mid-point. It provides people with a third anchor point (in addition to the end points). Second, the division into two sides means that even if sub-categories within the two sides do not match that summing the categories within each side should produce comparable recoded categories. Third, modifiers generally appear to be balanced. For example, strongly agree and strongly disagree have reciprocal values. Of course it is important that bipolar pairs exist in each language.

Unipolar scales without an explicit mid-point that ask about the amount of some quantity are likely to be more problematic. First, setting aside the translation of specific terms, it would be harder to match categories across languages since on these scales the mid-point is either not clearly defined or subsumed into some broad middle category. Second, the terms used tend to be asymmetrical which makes the matching across languages harder to achieve. Third, research indicates that on unipolar scales people confound terms and position (Klockars and Yamagishi, 1988). Without the mid-point clearly defined people will often assume that the middle category represents the middle even when the term used (e.g. good or bad) is clearly towards the positive or negative end. Fourth, people may not consistently understand what the low end of a pure unipolar scale mean. For example, if people are rating values as from high to low on "conservativism," does a low

conservatism score mean the value is very liberal or merely that it is not conservative and perhaps moderate. Fourth, without a clear mid-point it is possible for unipolar scales to "slide over," so that categories are unintentionally tilted towards the upper or lower end.⁸

Finally, more research is needed. It would be extremely useful both to cover more languages and have larger samples. Specific issues that need further study are:

1. How common is it that German terms are stronger (more near the extremes) than corresponding English and what can be done to compensate for this?

2. Would numerical scales with only the endpoint labelled be more equivalent across languages than labelled scales? What about a scale with only the ends and mid-point labeled?

3. To what extent does presenting terms in scales change the intensity that people associate with them (and the distance between categories)?

4. When people hold a attitude between two scale points, how do they decided to chose the category that is higher or lower than their precise position. Do they select the nearest category or are other decision rules used?

5. Are the positions of terms/categories independent of the substance of the scale (i.e. its subject matter and the nature of public attitudes toward it)? Would people rate "completely agree" about a moderate statement about the economy the same as they would in reference to an extreme statement about religion?

⁸On unipolar and bipolar scale in general see Ostrom, 1987. A suspect scale is the Eurobarometer life satisfaction scale ("On the whole are you very satisfied, fairly satisfied, not very satisfied, or not at all satisfied with the life you lead?) (Euporean Commission, 1996). Year-to-year changes across the European Union are minor and inter-country difference are large and pretty stable. It is suspected that the large inter-country differences are in part due to differences on the intensity of terms used in the scale, variations in translation of the underlying dimension itself, or both.

References

- Bartram, Peter and Yelding, David, "The Development of an Empirical Method of Selecting Phrases Used in Verbal Rating Scales: A Report on a Recent Experiment," Journal of the Market Research Society, 15 (July, 1973), 151-156.
- Bradburn, Norman M. and Sudman, Seymour, Improving Interview Method and Questionnaire Design. San Francisco: Jossey-Bass, 1979.
- Cliff, Norman, "Adverbs as Multipliers," Psychological Review, 66 (January, 1959), 27-44.
- Clogg, Clifford C., "Using Association Models in Sociological Research: Some Examples," American Journal of Sociology, 88 (1982), 114-134.
- Clogg, Clifford C., "Some Statistical Models for Analyzing Why Surveys Disagree," in Surveying Subjective Phenomena, edited by Charles F. Turner and Elizabeth Martin. Volume 2. New York: Russell Sage, 1984.
- Crespi, Leo P., "Semantic Guidelines to Better Survey Reportage," Office of Research, International Communication Agency, Memorandum, August 11, 1981.
- Davis, James A., "[Memorandum to] ISSP Methodology Group," September, 1993.
- European Commission, Eurobarometer: Public Opinion in the European Union. No 45. Brussels: European Commission, 1996.
- Jones, Lyle V. and Thurstone, L.L., "The Psychophysics of Semantics: An Experimental Investigation," Journal of Applied Psychology, 39 (February, 1955), 31-36.
- Hakel, Milton D., "How Often is Often?" American Psychologist, 23 (July, 1968), 533-534.
- Houglund, James G.; Johnson, Timothy P.; and Wolf, James G., "A Fairly Common Ambiguity: Comparing Rating and Approval Measures of Public Opinion," Sociological Focus, 25 (August, 1992), 257-271.
- Klockars, Alan J. and Yamagishi, Midori, "The Influence of Labels and Positions in Rating Scales," Journal of Educational Measurement, 25 (Summer, 1988), 85-96.
- Kumata, Hideya and Schramm, Wilbur, "A Pilot Study of Cross-Cultural Meaning," Public Opinion Quarterly, 20 (1956), 574-584.

- Laumann, Edward O.; Gagnon, John H.; Michael, Robert T.; and Michaels, Stuart, The Social Organization of Sexuality: Sexual Practices in the United States. Chicago: University of Chicago Press, 1994.
- Lichtenstein, Sarah and Newman, J. Robert, "Empirical Scaling of Common Verbal Phrases Associated with Numerical Probabilities," Psychon. Sci., 9 (1967), 563-564.
- Lodge, Milton, Magnitude Scaling: Quantitative Measurement of Opinions. Beverly Hills: Sage Publications, 1981.
- Lodge, Milton; Cross, David; Tursky; Tanenhaus, Joseph; and Reeder, Richard, "The Psychophysical Scaling of Political Support in the 'Real World'," Political Methodology, 3 (1976), 159-182.
- Lodge, Milton; Cross, David V.; Tursky, Bernard; and Tanenhaus, Joseph, "The Psychological Scaling and Validation of a Political Support Scale," American Journal of Political Science, 19 (November, 1975), 611-649.
- Lodge, Milton; Tanenhaus, Joseph; Cross, David; Tursky, Bernard; Foley, Mary Ann; and Foley, Hugh, "The Calibration and Cross-Modal Validation of Ratio Scales of Political Opinion in Survey Research," Social Science Research, 5 (1976), 325-347.
- Lodge, Milton and Tursky, Bernard, "Comparisons between Category and Magnitude Scaling of Political Opinion Employing SRC/CPS Items," American Political Science Review, 73 (1979), 50-66.
- Lodge, Milton and Tursky, Bernard, "On the Magnitude Scaling of Political Opinion in Survey Research," American Journal of Political Science, 25 (May, 1981), 376-419.
- Lodge, Milton and Tursky, Bernard, "The Social-Psychological Scaling of Political Opinion," in Social Attitudes and Psychophysical Measurement, edited by Bernd Wegener. Hillsdale, NJ: Lawrence Erlbaum Associates, 1982.
- MacKuen, Michael B. and Turner, Charles F., "The Popularity of Presidents, 1963-1980," in Surveying Subjective Phenomena, edited by Charles F. Turner and Elizabeth Martin. Volume 2. New York: Russell Sage, 1984.
- Mittelstaedt, Robert A., "Semantic Properties of Selected Evaluative Adjectives: Other Evidence," Journal of Marketing Research, 8 (May, 1971), 236-237.
- Mohler, Peter Ph.; Harkness, Janet; Smith Tom W., and Davis, James A., "Calibrating Response Scales Across Two Languages and Cultures," Paper presented to the International Sociological Association Conference, Colchester, July, 1996.

- Mosier, Charles, "A Psychometric Study of Meaning," Journal of Social Psychology, 13 (February, 1941), 123-140.
- Myers, James H. and Warner, W. Gregory, "Semantic Properties of Selected Evaluation Adjectives," Journal of Marketing Research, 5 (November, 1968), 409-412.
- O'Muircheartaigh, Colm A.; Gaskell, George D., and Wright, Daniel B., "The Impact of Intensifiers," Public Opinion Quarterly, 57 (Winter, 1993), 552-565.
- Orren, Gary R., "Presidential Popularity Ratings: Another View," Public Opinion, 1 (May/June, 1978), 35.
- Osgood, Charles E.; Suci, George J.; and Tannenbaum, Percy H., The Measurement of Meaning. Urbana, IL: University of Illinois Press, 1957.
- Ostrom, Thomas M., "Bipolar Survey Items: An Information Processing Perspective." in Social Information Processing and Survey Methodology, edited by Hans-J. Hippler, Norbert Schwarz, and Seymour Sudman. New York: Springer-Verlag, 1987.
- "RAC Quantifies the Vast Majority," The Sampler from Response Analyses, 31 (March, 1984), 2.
- Schaeffer, Nora Cate, "Hardly Ever or Constantly? Group Comparisons Using Vague Quantifiers," Public Opinion Quarterly, 55 (Fall, 1991), 395-423.
- Schriesheim, Chester and Schriesheim, Janet, "Development and Empirical Verification of New Response Categories to Increase the Validity of Multiple Response Alternatives Questionnaires," Educational and Psychological Measurement, 34 (Winter, 1974), 877-884.
- Scheuch, Erwin, "Theoretical Implications of Comparative Survey Research: Why the Wheel of Cross-Cultural Methodology Keeps on Being Reinvented," International Journal of Sociology, 4 (June, 1989), 147-167.
- Schwarz, Norbert and Hippler, Hans-J., "The Numeric Values of Rating Scales: A Comparison of their Impact in Mail Surveys and Telephone Interviews," International Journal of Public Opinion Research, 7 (Spring, 1995), 72-74.
- Schwarz, Norbert and Hippler, Hans-J., "What Response Scales May Tell Your Respondents: Informative Functions of Response Alternatives," in Social Information Processing and Survey Methodology, edited by Hans-J. Hippler, Norbert Schwarz, and Seymour Sudman. New York: Springer-Verlag, 1987.

- Schwarz, Norbert; Hippler, Hans-J.; Deutsch, Brigitte; and Strack, Fritz, "Response Scales: Effects of Category Range on Reported Behavior and Comparative Judgments," Public Opinion Quarterly, 49 (Fall, 1985), 388-395.
- Schwarz, Norbert; Knaeuper, Baerbel; Hippler, Hans-J.; Noelle-Neumann, Elisabeth; Clark, Leslie, "Rating Scales: Numeric Values May Change the Meaning of Scale Labels," Public Opinion Quarterly, 55 (1991), 570-582.
- Sigelman, Lee, "Answering the 1,000,000-Person Question: The Measurement and Meaning of Presidential Popularity," Research in Mircropolitics, 3 (1990), 209-226.
- Simpson, Ray H., "The Specific Meanings of Certain Terms Indicating Differing Degrees of Frequency," Quarterly Journal of Speech, 30 (October, 1944), 328-330.
- Smith, Tom W., "An Analysis of Response Patterns to the Ten-Point Scalometer," American Statistical Association 1993 Proceedings. Alexandria, VA: ASA, 1993.
- Smith, Tom W., "Happiness: Time Trends, Seasonal Variations, Intersurvey Differences, and Other Mysteries," Social Psychology Quarterly, 42 (1979), 18-30.
- Smith, Tom W., "The Holocaust Denial Controversy," Public Opinion Quarterly, 59 (Summer, 1995), 269-295.
- Spector, Paul E., "Choosing Response Categories for Summated Rating Scales," Journal of Applied Psychology, 61 (June, 1976), 374-375.
- Strahan, Robert and Gerbasi, Kathleen Carrese, "Semantic Style Variance in Personality Questionnaires," Journal of Psychology, 85 (September, 1973), 109-118.
- Traenkle, Ulrich, "Auswirkungen der Gestaltung der Antworskala suf Quantitative Urteile," Zeitschrift fuer Sozial Psychologie, 18 (1987), 88-99.
- Van de Vijver, Fons and Leung, Kwok, "Methods and Data Analysis of Comparative Research," in Handbook of Cross-Cultural Personality, forthcoming.
- Vidali, Joseph J., "Context Effects on Scales Evaluatory Adjective Meaning," Journal of the Market Research Society, 17 (January, 1975), 21-25.
- Wallsten, Thomas S.; Budescu, David V.; Rapoport, Amnon; Zwick, Rami; and Forsyth, Barbara, "Measuring the Vague Meanings of Probability Terms," Journal of Experimental Psychology, 115

(December, 1986), 348-365.

Wilcox, Clyde; Sigelman, Lee; and Cook, Elizabeth, "Some Like it Hot: Individual Differences in Responses to Group Feeling Thermometers," Public Opinion Quarterly, 53 (Summer, 1989), 246-257.

Wildt, Albert R. and Mazis, Michael B., "Determinants of Scale Response: Label Versus Position," Journal of Marketing Research, 15 (May, 1978), 261-267.

Worcester, Robert M. and Burns, Timothy R., "A Statistical Examination of the Relative Precision of Verbal Scales," Journal of the Market Research Society, 17 (July, 1975), 181-197.

Table 1

Mean Scores on Agree/Disagree Terms

Term	Mean	Standard Deviation
Completely agree	19.4	1.6
Definitely agree	19.0	1.5
Strongly agree	18.8	1.3
Very much agree	18.5	2.2
Agree a lot	17.2	2.8
Agree	16.1	2.9
Basically agree	13.8	3.1
Probably agree	13.6	2.9
Tend to agree	13.5	2.8
Moderately agree	13.3	2.3
Somewhat agree	12.9	2.4
Agree a little	12.1	2.6
In the middle	10.1	0.7
Neither agree nor disagree	9.9	1.3
Can't choose	9.8	2.7
Undecided	9.6	1.8
Disagree a little	7.1	2.2
Somewhat disagree	6.6	2.1
Moderately disagree	6.4	2.3
Tend to disagree	6.4	2.7
Probably disagree	6.2	3.1
Disagree	3.5	2.9
Not agree	3.5	3.1
Disagree a lot	3.0	3.6
Strongly disagree	1.5	2.2
Very much disagree	1.4	1.6
Definitely disagree	1.0	1.3
Completely disagree	0.8	2.3

N=97-101

Table 2

Mean Scores on Important/Unimportant Terms

Term	Important Only List		Important/Unimportant List		Combined List	
	Mean	SD	Mean	SD	Mean	SD
Important:						
Extremely	19.4	1.2	19.4	0.9	19.4	1.0
Very, very	19.0	2.4	--	--	--	--
Exceptionally	18.9	2.4	--	--	--	--
Completely	18.6	2.9	19.1	1.8	18.8	2.5
Definitely	18.5	2.0	18.4	2.0	18.4	2.0
Highly	18.2	1.9	--	--	--	--
Very	18.2	1.5	18.3	2.6	18.2	2.1
Quite	16.8	2.8	--	--	--	--
IMPORTANT	15.1	3.6	16.3	2.8	15.5	3.5
Pretty	15.0	3.1	15.6	2.6	15.3	2.9
Probably	13.0	3.4	14.0	3.1	13.5	3.3
Fairly	13.4	3.5	13.9	2.2	13.6	3.0
Somewhat	12.2	3.5	13.2	2.5	12.7	3.1
Slightly	10.8	3.4	12.0	2.7	11.3	3.2
A little bit	10.1	4.2	12.2	2.4	11.1	3.7
Neither imp. nor unimp.	9.0	3.6	9.5	2.4	9.3	3.1
Not too	6.8	3.6	--	--	--	--
Not very	4.7	3.4	5.5	4.2	5.1	3.8
Not	2.4	3.4	4.1	4.4	3.2	4.0
Not at all	1.4	3.3	3.0	4.2	2.2	3.8
Unimportant:						
In between	--	--	10.0	2.2	--	--
Slightly	--	--	8.0	2.7	--	--
A little bit	--	--	7.9	3.0	--	--
Somewhat	--	--	6.6	2.7	--	--
Probably	--	--	6.1	3.8	--	--
Fairly	--	--	5.8	3.2	--	--
Very	--	--	5.1	6.7 ^a	--	--
Pretty	--	--	4.7	3.1	--	--
UNIMPORTANT	--	--	3.6	3.9	--	--
Definitely	--	--	1.8	3.3	--	--
Completely	--	--	1.3	3.6	--	--
Extremely	--	--	0.8	1.8	--	--
N	56-58		51-54		109-112	

Table 2 (continued)

^aThis item has a small number of cases coded near the high end of the scale (16-20). These cases create the large standard deviation (6.7) and also make the mean (5.1) much higher than the median (2). In all other cases the mean and median are very close (almost always within +/- 1). Inspection of the cases to see why a high number of pole reversals (i.e. errors of reference) occurred on this item did not reveal any special cause.

Table 3

Ratings of In Favor of and Against

	Ascending Order		Mixed Order	
	Mean	StdDev	Mean	StdDev
Strongly against	1.6	2.0	1.3	1.9
Against	3.9	2.2	3.5	3.4
Slightly against	6.9	2.4	7.2	1.9
Neither against nor in favor of	9.6	1.7	9.8	1.4
Slightly in favor of	12.2	2.3	12.1	1.9
In favor of	15.9	2.4	15.6	2.8
Strongly in favor of	18.9	1.3	18.6	1.6
% with all items rated in ascending order	71.7	--	37.0	--
N	53		52-54	

Table 4
Symmetry in Ratings

A. Ratings of Agree/Disagree (Samples A and B)

	Agree	Disagree	20 - Disagree
Completely	19.4	0.8	19.2
Definitely	19.0	1.0	19.0
Strongly	18.8	1.5	18.5
Very much	18.5	1.4	18.6
A lot	17.2	3.0	17.0
AGREE/DISAGREE	16.1	3.5	16.5
Not agree	--	3.5	16.5
Probably	13.6	6.2	13.8
Tend to	13.5	6.4	13.6
Moderately	13.3	6.4	13.6
Somewhat	12.9	6.6	13.4
A little	12.1	7.1	12.9

B. Ratings of Importance/Unimportance (Sample B)

	Important	Unimportant	20 - Unimp.
Extremely	19.4	0.8	19.2
Completely	19.1	1.3	18.7
Definitely	18.4	1.8	18.2
Very	18.3	5.1	14.9 ^a
IMPORTANT/ UNIMPORTANT	16.3	3.6	16.4
Pretty	15.6	4.7	15.3
Probably	14.0	6.1	13.9
Fairly	13.9	5.8	14.2
Somewhat	13.2	6.6	13.4
A little bit	12.2	7.9	12.1
Slightly	12.0	8.0	12.0

C. Ratings of In Favor of/Against (Samples A and B)

	In Favor of	Against	20 - Against
Strongly	18.7	1.5	18.5
In Favor of/ Against	15.8	3.7	16.3
Slightly	12.2	7.1	12.9

Table 5

Comparisons Across Rating Scales

A. Agree/Disagree and Important/Unimportant (Sample B)

	Agree	Important	Disagree	Unimportant
Completely	19.4	19.1	0.8	1.3
Definitely	18.9	18.4	1.2	1.8
Very much/very	18.2	18.3	1.4	5.1 ^a
BASE WORD	15.6	16.6	3.7	3.6
Probably	13.5	14.0	6.5	6.1
Somewhat	12.5	13.2	6.6	6.6
A little/ a little bit	12.1	12.2	7.1	7.9

B. Agree/Disagree and In Favor of/Against (Samples A and B)

	Agree	In Favor of	Disagree	Against
Strongly	18.8	18.7	1.5	1.5
BASE WORD	16.1	15.8	3.5	3.7
Neither Agree/In Favor or...	9.9	9.7	--	--

Table 6
Range of Acceptable Values

	Mean Range ^a	StdDev ^b
Strongly agree	2.6	1.3
Basically agree	4.0	3.1
Agree	4.1	2.9
Neither agree nor disagree	1.6	1.3
Can't choose	1.6	2.7
Disagree	4.1	3.5
Not agree	3.9	3.1
Strongly disagree	2.6	2.1

^aDifference between high and low limits in Q. 6

^bStandard deviation of items in Q. 3.

Table 7

Closeness of the Meaning of Various Pairs of Words
to Agree/Disagree

	Mean ^a	% Very much the same
For/against	2.4	34.7
Important/unimportant	3.2	8.9
Like/dislike	2.9	8.8
Favor/Oppose	2.3	29.7
Positive/negative	2.9	14.9

n=101

^aResponse scale ran from 1=Very much the same to 4=Very much different. Lower number indicates pairs are closer.

Table 8

Terms Used in the Definition of Agree

Accept, acceptance	6
Accord, accordance	2
Against (not)	1
Agree, agreeable, agreeing, agreement	13
Alike	1
Approve	3
Congenial	1
Consensus, consent	4
Disagree (not)	1
Favor	6
For	6
Harmony	2
In line with	2
Like, liking	3
Mutual	1
OK	2
Same	16
Similar	1
Support	2
True	2
Valid	1

Table 9

American/German Scores on
Agree/Disagree, Stimme zu/Lehne ab, and Stimme zu/Stimme Nicht zu

English/German	America		Germany			
	Means	SD	Stimme zu/ Lehne ab		Stimme zu/ Nicht	
	Means	SD	Means	SD	Means	SD
Agree/Stimme zu:						
Completely/Voellig	19.4	1.2	19.3	1.8	19.3	1.4
Definitely/Bestimmt	19.0	1.5	17.9	2.5	17.6	3.1
Strongly/Voll und ganz	18.8	1.3	19.7	1.4	19.8	0.7
Very much/Sehr	18.5	2.2	17.6	2.8	18.3	2.1
A Lot/Ziemlich	17.2	2.8	16.0	2.2	16.4	2.6
AGREE/STIMME ZU	16.1	2.9	17.5	2.5	17.4	2.8
Basically/Im grunde	13.8	3.1	14.4	2.9	14.6	3.5
Probably/Wahrscheinlich	13.6	2.9	13.8	2.8	14.0	3.1
Tend to/Eher	13.5	2.8	13.8	2.5	13.8	3.1
Moderately/Maessig	13.3	2.3	12.3	2.4	10.4	3.9
Somewhat/Teilweise	12.9	2.4	13.3	2.5	12.8	2.5
A little/Ein bisschen	12.1	2.6	12.5	2.5	11.7	3.5
Middle/Mitte:						
In the middle/ In der mitte	10.1	0.7	10.0	1.3	9.9	1.3
Neither agree nor disagree/Stimme weder zu noch lehne ab/Stimme weder zu noch nicht zu	9.9	1.3	9.7	1.5	9.6	2.4
Can't Choose/Kann ich nicht sagen	9.8	2.7	9.5	2.5	8.5	3.7
Undecided/Unentschieden	9.6	1.8	10.0	1.1	10.0	0.5
Diasagree/Lehne ab/Stimme nicht zu						
A little/Ein bisschen	7.1	2.2	6.7	2.4	---	---
Somewhat/Teilweise/ Zum teil nicht	6.6	2.1	6.8	2.5	7.6	3.1
Moderately/Maessig	6.4	2.3	6.6	3.1	---	---
Tend to/Eher	6.4	2.7	5.9	2.4	6.0	3.1
Probably/Wahrscheinlich	6.2	3.1	6.1	2.6	4.9	2.5
DISAGREE/LEHNE AB	3.5	2.9	2.9	2.6	1.2	2.4
Not agree/STIMME NICHT ZU	3.5	3.1	3.5	3.6	1.8	2.2
A lot/Ziemlich/ Ueberwiegend nicht	3.0	3.6	4.1	2.6	4.4	3.1

Table 9 (continued)

English/German	America		Germany			
	Means	SD	Stimme zu/ Lehne ab		Stimme zu/ Nicht	
			Means	SD	Means	SD
Strongly/Stark/ Ueberhaupt nicht	1.5	2.2	1.7	2.7	0.4	0.9
Very much/Sehr/ Entschieden nicht	1.4	1.6	2.0	2.6	1.2	2.9
Definitely/Bestimmt	1.0	1.3	2.7	3.0	1.7	2.8
Completely/Voellig/ Ganz und gar nicht	0.8	2.3	1.1	2.6	0.6	1.6

Table 10

American/German Scores on
Important/Unimportant and Wichtig/Unwichtig

English/German	America		Germany	
	Means	SD	Means	SD
Important/Wichtig:				
Extremely/Aeusserst	19.4	1.2	18.6	3.0
Completely/Voellig	18.8	2.5	18.8	3.0
Definitely/Bestimmt	18.4	2.0	16.5	2.6
Very/Ganz	18.2	2.1	18.4	2.9
IMPORTANT/WICHTIG	15.5	3.5	16.4	2.8
Pretty/Schon ziemlich	15.3	2.9	15.7	2.7
Probably/Wahrscheinlich	13.5	3.3	13.0	3.5
Fairly/Einigermassen	13.6	3.0	11.9	3.9
Somewhat/Teilweise	12.7	3.1	12.9	2.8
A little bit/Ein wenig	11.1	3.7	11.0	3.5
Not very/Nicht sehr	5.1	3.8	5.8	3.7
Not/Nicht	3.2	4.0	2.6	3.4
Not at all/Ueberhaupt nicht	2.2	3.8	1.5	2.8
Middle:				
Neither important nor unimportant/Weder wichtig noch unwichtig	9.3	3.1	9.7	2.3
In between/Dazwischen	10.0	2.2	10.1	1.9
Unimportant/Unwichtig:				
A little bit/Ein wenig	7.9	3.0	7.0	3.2
Somewhat/Teilweise	6.6	2.7	6.6	2.6
Probably/Wahrscheinlich	6.1	3.8	5.3	3.1
Fairly/Einigermassen	5.8	3.2	5.2	3.1
Very/Sehr	5.1	6.7	1.5	3.7
Pretty/Schon ziemlich	4.7	3.1	4.8	4.6
UNIMPORTANT/UNWICHTIG	3.6	3.9	2.2	3.4
Definitely/Bestimmt	1.8	3.3	2.6	3.1
Completely/Voellig	1.3	3.6	1.1	3.0
Extremely/Aeusserst	0.8	1.8	0.9	2.8

Table 11

Terms Used in the Definition of Agree and Stimme zu

A. Agree

Accept, acceptance	6
Accord, accordance	2
Against (not)	1
Agree, agreeable, agreeing, agreement	13
Alike	1
Approve	3
Congenial	1
Consensus, consent	4
Disagree (not)	1
Favor	6
For	6
Harmony	2
In line with	2
Like, liking	3
Mutual	1
OK	2
Same	16
Similar	1
Support	2
True	2
Valid	1

B. Stimme zu

Akzeptieren, akzeptabel, akzeptiere	3
Anerkennen	1
Befueworte, befueworten, befuewortung	4
Bejahe, bejahen, bejahung	8
Dafuer	42
Einverstanden, Einverständnis, einverstandniserkaerung	55
Gleiche, gleichen, gleicher	23
Grosse	1
Grund, grunde	4
Gut	8
Identisch	1
Positiv, positive	7
Richtig, richtige, richtigkeit	9
Selbe, selben, selber	6
Voll, volle, volles	11
Uebereinstimmen, uebereinstimmung	7
Ueberzeugt	9
Ueberzeugung	3
Zustimme, zustimmen, zustimmung	14
Zutreffend	1
Zuveriaessig	1

Table 12

Mean Correlations Using Raw and Adjusted Response Values
for Agree/Disagree ISSP Questions

(Pearson's r)

A. Religion

	Raw	Adjusted
Mean Inter-item Correlations	.169	.160
Mean Correlation with		
Gender	.069	.071
Age	.127	.123
Years of Education	.153	.147
Highest Educational Degree	.141	.136
Church Attendance	.219	.215

B. Environment

	Raw	Adjusted
Mean Inter-item Correlations	.148	.130
Mean Correlation with		
Gender	.088	.072
Age	.094	.089
Years of Education	.181	.162
Highest Educational Degree	.177	.156
Church Attendance	.076	.073

Source: ISSP, 1991 and 1993

Appendix: American Questionnaires

Sample A:

Time: _____

Section A: Translation

1. If you were to consider your life in general these days, how happy or unhappy would you say you are on the whole...

Very happy	1
Fairly happy	2
Not very happy	3
Not at all happy	4

2. What is your opinion of the following statement?

It is the responsibility of the government to reduce the differences in income between people with high incomes and those with low incomes.

Do you...

Agree strongly	1
Agree	2
Neither agree nor disagree	3
Disagree	4
Disagree strongly	5
CAN'T CHOOSE	8

3. In order to help us write better and more understandable questions, we need to know how people like you use certain words. Here is a scale that goes from 0 to 20 . The zero (0) point means you totally and completely **disagree** with an idea and 20 means you totally and completely **agree** with an idea. I'm going to read you some terms and I'd like you to tell me what number best represents how much agreement or disagreement the word or phrase means.

A. What score between 0 and 20 would you give to ...

HAND CARD Q.3	First Response	Changes
a. basically agree	_____	_____
SHUFFLE CARDS AND ASK REST OF PHRASES		
b. agree	_____	_____
c. agree a little	_____	_____
d. agree a lot	_____	_____
e. can't choose	_____	_____
f. completely agree	_____	_____
g. completely disagree	_____	_____
h. definitely agree	_____	_____
i. definitely disagree	_____	_____
j. disagree	_____	_____
k. disagree a little	_____	_____
l. disagree a lot	_____	_____
m. in the middle	_____	_____
n. moderately agree	_____	_____
o. moderately disagree	_____	_____
p. neither agree nor disagree	_____	_____
q. not agree	_____	_____
r. probably agree	_____	_____
s. probably disagree	_____	_____
t. somewhat agree	_____	_____
u. somewhat disagree	_____	_____
v. strongly agree	_____	_____
w. strongly disagree	_____	_____
x. tend to agree	_____	_____
y. tend to disagree	_____	_____
z. undecided	_____	_____
aa. very much agree	_____	_____
bb. very much disagree	_____	_____

Code 95=verbatim; 96=can't rate term; 98=Don't know term; 99=missing/no answer/unreadable. Code 0.5 for value between the 21 numbered scale points. If range given, code 95 and indicate range specified.

B. HAND RESPONSE SHEET WITH ANSWERS RECODED TO RESPONDENT AND SAY: Please look over your answers. If you want to change any of your responses, indicate in the right hand column, the one headed "CHANGES," what number you now want to give a phrase.

4. Now we're going to use a similar scale that goes from 0 to 20 to rate some additional phrases. On this scale 0 indicates something of the lowest importance possible, something last and least in importance, and 20 indicates the highest importance possible, something that is first and foremost in importance. As I read you each phrase, please tell me what number best represents how much importance the phrase indicates.

HAND CARD Q. A4 SHUFFLE CARDS AND ASK IN THAT ORDER

- a. Pretty important _____
- b. Definitely important _____
- c. Not too important _____
- d. Extremely important _____
- e. Not very important _____
- f. Fairly important _____
- g. Highly important _____
- h. Probably important _____
- i. Not at all important _____
- j. Exceptionally important _____
- k. Not important _____
- l. Very important _____
- m. Somewhat important _____
- n. Important _____
- o. Neither important nor unimportant _____
- p. Quite important _____
- q. Very, very important _____
- r. A little bit important _____
- s. Slightly important _____
- t. Completely important _____

5. And now consider a similar scale going from 0 to 20. Point 0 indicates that someone is totally and completely **against** an idea and point 20 means that someone is totally and completely **in favor** of the idea. I'm going to read you some terms and I'd like you to tell me what number best represents how much someone is either **against** or **in favor of** an idea.

HAND CARD Q. A5

- a. Slightly against _____
- b. Strongly in favor of _____
- c. Against _____
- d. Strongly against _____
- e. Neither against nor in favor of _____
- f. Slightly in favor of _____
- g. In favor of _____

GO BACK TO Q. 3, LOOK UP THE RATINGS GIVEN TO TERMS USED IN 6A-H, AND ENTER IN MIDDLE COLUMN, "PREVIOUSLY GIVEN," BELOW:

6. Now let's consider again a few of the terms you rated about disagreement and agreement. Here again is the scale that goes from 0 to 20. The zero (0) point means you totally and completely disagree with an idea and 20 means you totally and completely agree with an idea. You gave "basically agree" a score of [mention number given to Q. 3]. Now I want you to think about what is the lowest score that you feel would still represent the phrase "basically agree" and what would be the highest score that would still mean "basically agree," that is what numbers would represent the range from high to low that would describe where "basically agree" fits on our scale from 0 to 20. First, what would be the lowest number for "basically agree?" And what would be the highest number?

REPEAT FOR 6B-H.

HAND CARD Q. 6

INTERVIEWER:
FILL-IN FROM
Q. 3

	Lowest	Previously Given	Highest
a. Basically agree	_____	_____	_____
b. Strongly agree	_____	_____	_____
c. Neither agree nor disagree	_____	_____	_____
d. Disagree	_____	_____	_____
e. Can't Choose	_____	_____	_____
f. Not agree	_____	_____	_____
g. Strongly Agree	_____	_____	_____
h. Agree	_____	_____	_____

NOTE: IF ANY OF THE TERMS WERE NOT RATED 0-20 IN Q. 3, THEN SKIP AND DO NOT ASK IN Q. 6.

7. Now, I'm going to ask you about some of words we've just been discussing. What does the word "agree" mean? What does it involve? How about "disagree"? What does it mean or involve? And WHAT does the phrase "neither agree nor disagree" mean? What does it involve? And what about "important"? And how about "unimportant"?

A. Agree

B. Disagree

C. Neither agree nor disagree

D. Important

E. Unimportant

8. I'm going to read several pairs of words and I would like you to compare and contrast these with the pair "agree/disagree". I want you to tell me whether you think they mean very much the same as "agree/disagree", somewhat the same as "agree/disagree", somewhat different from "agree/disagree" or very much different from "agree/disagree."

First, does the phrase "for/against" mean very much the same as "agree/disagree", somewhat the same as "agree/disagree", somewhat different from "agree/disagree" or very much different from "agree/disagree"?

REPEAT FOR 8B-E.

	Very Much the Same	Somewhat the Same	Somewhat Different	Very Much Different	DK
a. for/against	1	2	3	4	8
b. important/ unimportant	1	2	3	4	8
c. like/ dislike	1	2	3	4	8
d. favor/ oppose	1	2	3	4	8
e. positive/ negative	1	2	3	4	8

9. What language did you mainly speak at home when you were a child?

10. What languages do you now speak at home?

First Mentioned: _____

Second Mentioned: _____

Q. 3

0 - Totally and Completely Disagree
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20 - Totally and Completely Agree

Q.A4

0 - Lowest Possible Importance/Last and Least in Importance
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20 - Highest Possible Importance/First and Foremost in Importance

Q. A5

0 - Totally and Completely Against
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20 - Totally and Completely in Favor of

Q. 6

0 - Totally and Completely Disagree
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20 - Totally and Completely Agree

Q. 8

Agree/Disagree

Very Much the Same	1
Somewhat the Same	2
Somewhat Different	3
Very Much Different	4

Sample B:

Time: _____

Section A: Translation (B)

1. If you were to consider your life in general these days, how happy or unhappy would you say you are on the whole...

Completely happy	1
Very happy	2
Fairly happy	3
Not very happy	4
Not at all happy	5

2. What is your opinion of the following statement?

It is the responsibility of the government to reduce the differences in income between people with high incomes and those with low incomes.

Do you...

Completely agree	1
Somewhat agree	2
Neutral	3
Somewhat disagree	4
Completely disagree	5
CAN'T CHOOSE	8

3. In order to help us write better and more understandable questions, we need to know how people like you use certain words. Here is a scale that goes from 0 to 20 . The zero (0) point means you totally and completely **disagree** with an idea and 20 means you totally and completely **agree** with an idea. I'm going to read you some terms and I'd like you to tell me what number best represents how much agreement or disagreement the word or phrase means.

A. What score between 0 and 20 would you give to ...

HAND CARD Q. 3	First Response	Changes
a. basically agree	_____	_____
SHUFFLE CARDS AND ASK REST OF PHRASES		
b. agree	_____	_____
c. agree a little	_____	_____
d. agree a lot	_____	_____
e. can't choose	_____	_____
f. completely agree	_____	_____
g. completely disagree	_____	_____
h. definitely agree	_____	_____
i. definitely disagree	_____	_____
j. disagree	_____	_____
k. disagree a little	_____	_____
l. disagree a lot	_____	_____
m. in the middle	_____	_____
n. moderately agree	_____	_____
o. moderately disagree	_____	_____
p. neither agree nor disagree	_____	_____
q. not agree	_____	_____
r. probably agree	_____	_____
s. probably disagree	_____	_____
t. somewhat agree	_____	_____
u. somewhat disagree	_____	_____
v. strongly agree	_____	_____
w. strongly disagree	_____	_____
x. tend to agree	_____	_____
y. tend to disagree	_____	_____
z. undecided	_____	_____
aa. very much agree	_____	_____
bb. very much disagree	_____	_____

Code 95=verbatim; 96=can't rate term; 98=Don't know term; 99=missing/no answer/unreadable. Code 0.5 for value between the 21 numbered scale points. If range given, code 95 and indicate range specified.

B. HAND RESPONSE SHEET WITH ANSWERS RECODED TO RESPONDENT AND SAY: Please look over your answers. If you want to change any of your responses, indicate in the right hand column, the one headed "CHANGES," what number you now want to give a phrase.

4. Now we're going to use a similar scale that goes from 0 to 20 to rate some additional words. On this scale 0 indicates something that is completely and totally unimportant and 20 indicates something that is completely and totally important. As I read you various terms, please tell me what number best represents how much importance the word or phrase means.

HAND CARD Q. B4

SHUFFLE CARDS AND ASK IN THAT ORDER

- a. Pretty important _____
- b. Definitely important _____
- c. Probably unimportant _____
- d. Extremely important _____
- e. Not very important _____
- f. Fairly important _____
- g. Very unimportant _____
- h. Probably important _____
- i. Not at all important _____
- j. Fairly unimportant _____
- k. Not important _____
- l. Very important _____
- m. Somewhat important _____
- n. Unimportant _____
- o. Completely important _____
- p. Neither important nor unimportant _____
- q. Somewhat unimportant _____
- r. Important _____
- s. Pretty unimportant _____
- t. A little bit unimportant _____
- u. Slightly unimportant _____
- v. A little bit important _____
- w. Definitely unimportant _____
- x. Extremely unimportant _____
- y. Slightly important _____
- z. Completely unimportant _____
- aa. In between _____

5. And now consider a similar scale going from 0 to 20. Point 0 indicates that someone is totally and completely **against** an idea and point 20 means that someone is totally and completely **in favor** of the idea. I'm going to read you some terms and I'd like you to tell me what number best represents how much someone is either **against** or **in favor** of an idea.

HAND CARD Q. B5

- a. Strongly against _____
- b. Against _____
- c. Slightly against _____
- d. Neither against nor in favor of _____
- e. Slightly in favor of _____
- f. In favor of _____
- g. Strongly in favor of _____

GO BACK TO Q. 3, LOOK UP THE RATINGS GIVEN TO TERMS USED IN 6A-H, AND ENTER IN MIDDLE COLUMN, "PREVIOUSLY GIVEN," BELOW:

6. Now let's consider again a few of the terms you rated about disagreement and agreement. Here again is the scale that goes from 0 to 20. The zero (0) point means you totally and completely disagree with an idea and 20 means you totally and completely agree with an idea. You gave "basically agree" a score of [mention number given to Q. 3]. Now I want you to think about what is the lowest score that you feel would still represent the phrase "basically agree" and what would be the highest score that would still mean "basically agree," that is what numbers would represent the range from high to low that would describe where "basically agree" fits on our scale from 0 to 20. First, what would be the lowest number for "basically agree?" And what would be the highest number?

REPEAT FOR 6B-H.

HAND CARD Q. 6

INTERVIEWER:
FILL-IN FROM
Q. 3

	Lowest	Previously Given	Highest
a. Basically agree	_____	_____	_____
b. Strongly agree	_____	_____	_____
c. Neither agree nor disagree	_____	_____	_____
d. Disagree	_____	_____	_____
e. Can't Choose	_____	_____	_____
f. Not agree	_____	_____	_____
g. Strongly Agree	_____	_____	_____
h. Agree	_____	_____	_____

NOTE: IF ANY OF THE TERMS WERE NOT RATED 0-20 IN Q. 3, THEN SKIP AND DO NOT ASK IN Q. 6.

7. Now, I'm going to ask you about some of words we've just been discussing. What does the word "agree" mean? What does it involve? How about "disagree"? What does it mean or involve? And WHAT does the phrase "neither agree nor disagree" mean? What does it involve? And what about "important"? And how about "unimportant"?

A. Agree

B. Disagree

C. Neither agree nor disagree

D. Important

E. Unimportant

8. I'm going to read several pairs of words and I would like you to compare and contrast these with the pair "agree/disagree". I want you to tell me whether you think they mean very much the same as "agree/disagree", somewhat the same as "agree/disagree", somewhat different from "agree/disagree" or very much different from "agree/disagree."

First, does the phrase "for/against" mean very much the same as "agree/disagree", somewhat the same as "agree/disagree", somewhat different from "agree/disagree" or very much different from "agree/disagree"?

REPEAT FOR 8B-E.

	Very Much the Same	Somewhat the Same	Somewhat Different	Very Much Different	DK
a. for/against	1	2	3	4	8
b. important/ unimportant	1	2	3	4	8
c. like/ dislike	1	2	3	4	8
d. favor/ oppose	1	2	3	4	8
e. positive/ negative	1	2	3	4	8

9. What language did you **mainly** speak at home when you were a child?

10. What languages do you now speak at home?

First Mentioned: _____

Second Mentioned: _____

Q. 3

- 0 - Totally and Completely Disagree
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20 - Totally and Completely Agree

Q. B4

- 0 - Completely and Totally Unimportant
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20 - Completely and Totally Important

Q. B5

Strongly against

Against

Slightly against

Neither against nor in favor of

Slightly in favor of

In favor of

Strongly in favor of

-	0 - Totally and Completely Against
-	1
-	2
-	3
-	4
-	5
-	6
-	7
-	8
-	9
-	10
-	11
-	12
-	13
-	14
-	15
-	16
-	17
-	18
-	19
-	20 - Totally and Completely in Favor of

Q. 6

0 - Totally and Completely Disagree
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20 - Totally and Completely Agree

0.8

Agree/Disagree

Very Much the Same	1
Somewhat the Same	2
Somewhat Different	3
Very Much Different	4