Development of a prototype questionnaire to survey public attitudes toward stuttering: principles and methodologies in the first prototype.(Survey)

St. Louis, Kenneth O., Bobbie Boyd Lubker, J. Scott, Trisha A. Adkins, and Jaan Pill. "Development of a prototype questionnaire to survey public attitudes toward stuttering: principles and methodologies in the first prototype.(Survey). ." The Internet Journal of Epidemiology. (June 13, 2008): NA. General OneFile. Gale. California State Univ, Northridge. 2 Oct. 2009 <http://find.galegroup.com/gtx/infomark.do?&contentSet=IAC-Documents&type=retrieve&tabID=T002&prodId=ITOF&docId=A180592546&source=gale&srcprod=ITOF&userGroupName=csunorthridge&version=1.0>.

Abstract

The Experimental Edition of the Public Opinion Survey of Human Attributes (POSHA-E) was designed to investigate public opinions about stuttering and stuttering compared to other attributes cross-culturally and internationally. Respondents (n = 165) rated items on stuttering in the context of eight other attributes, or "anchors," assumed to range from negative (e.g., "mental illness"), to neutral (e.g., "left-handed"), to positive (e.g., "intelligent"). Only those methodological results that inform subsequent development of this prototype survey instrument or the data collection processes are elaborated. The field test evaluated the efficiency of a quasi-continuous scale, order effects, scoring efficiency, and data reduction in nonprobability samples. Order effects were minimal. Problems in respondent scoring and data reduction were appreciable, but these did not appear to affect the mean scores on the POSHA-E. Results have implications for data collection methods for the population-based international project and illustrate complexities of contemporary survey research methods.

Keywords: Stuttering, Attitudes, Stigma, Stereotypes, Sampling

Project Overview, Rationale, And Measurement Issues

Long-Range Goals

A task force was convened in 1999 to launch an international initiative devoted to exploring public attitudes toward stuttering. Its long-term objectives were to develop a survey instrument that could effectively obtain baseline measures of public attitudes toward stuttering in comparison to various other stigmatizing conditions potentially in any part of the world. Once completed, the empirical data to be gained from the instrument could be used by various stakeholders to foster and to evaluate effectiveness of strategies for mitigating societal stigma to which people who stutter are subjected. St. Louis (2005) briefly reported the rationale and vision for this "International Project on Attitudes Toward Human Attributes" (IPATHA). This paper elaborates the rationale and vision followed by explanations of initial steps in the development of

an instrument designed to obtain such baseline measures. Specifically it describes the rationale, design, and initial field-testing of the first experimental version of the Public Opinion Survey of Human Attributes (POSHA-E).

Rationale

Scope of the Problem of Stigma

Stigma, regarded by Goffman (1963) as the manifestation of a "spoiled identity," is a universal human experience. Individuals who are regarded as being undesirable or potentially dangerous often live with ridicule, bullying, and illegal discrimination. As a result they do not seek or receive the health care or specialized treatments they need and may experience lifelong negative consequences in education, employment, promotion, and social acceptance. Since Goffman's seminal work, stigma has been recognized as an important area of scientific inquiry. Stigma and its behavioral manifestation, discrimination, negatively affects health, both physically and mentally, of more than one billion of the world's population (Wahl, 1999; Weiss, Jadhav, Raguram, Vounatsou, & Littlewood, 2001). Moreover, stigma and discrimination are especially powerful in low- to moderate-income (developing) countries and marginalized groups in high-income (developed) nations (Ustun, Rehm, Chatterji, Saxena, Trotter, Room, Bickenbach, et al., 1999). If stigma could be reduced, the well-being and health of millions could be improved.

Available evidence clearly indicates that negative public attitudes can have dramatic negative impacts on the lives of people with a variety of stigmatizing characteristics. Mental illness is one area that has received perhaps the greatest attention (e.g., Crisp, Gelder, Rix, Meltzer, & Rowlands, 2000; Gelder, 2001, Sartorius, Jablonsky, Korton, Ernberg, Anker, Cooper, & Day, 1986; Thompson, 1999; Thompson, Stuart, Bland, Arbolele-Florez, Warner, & Dickson, 2002), but stigma affects people with numerous other conditions, including communication disorders. As noted above, this field study focuses on the specific communication disorder of stuttering.

In the case of stuttering, stigma is often discussed within the context of a "stuttering stereotype" (Blood, 1999; Cooper & Cooper, 1985; Ham, 1990; Reingold & Krishnan, 2002). Moreover, stigma has been shown to change or vary according to specific variables. For example, when stuttering persons are acquainted personally with respondents, reported stigma seems to disappear (Klassen, 2002). None of the foregoing means that physical aspects of many stigmatized conditions are unimportant in consideration of health outcomes. For stuttering it is well known that physiological and neurological differences exist between stuttering speakers and nonstuttering controls, when groups are compared in physiological domains such as genetics and brain function (e.g., Drayna, Kilshaw, & Kelly, 1999; De Nil, Kroll, Lafaille, & Houle, 2003). Nevertheless, social environments play an important role in the way stuttering is experienced, how it develops, and its effect on persons' lives (e.g., Conture, 2001; Smith & Kelly, 1997; Yaruss & Quesal, 2004).

Measuring Stigma

Increasingly, there are calls for public awareness and education campaigns to diminish stigma associated with stuttering and other conditions (e.g., Blood, 1999; ILAE/IBE/WHO Global

Campaign Against Epilepsy, 2002, Klompas & Ross, 2004; Langevin, 1997; NAAFA, 2002; Wahl, 1999; WHO, 2001). The rationale is that if groups who are stigmatized could, through a more educated public, face positive or even neutral public reactions to their conditions, the impact of their conditions would become less handicapping. If this can be achieved, the benefits would be immediate and major. These campaigns seem to assume that providing the public with accurate information will motivate people to become more understanding and/or empathetic, and ultimately behave in less discriminatory ways toward those who have the undesirable conditions. Historically, attitudes have been shown to improve for some stigmatizing labels of the past, e.g., being labeled as "insane" or "a witch" (Porter, 2001). Herek, Capitanio & Widaman (2002) found that the overt stigma for AIDS declined slightly in the 1990s even though stigma still remained strong. Similarly, although public understanding of mental illness improved from the 1950s to the 1990s, stigma was not "defused" (US Surgeon General, 2004). For some physical handicaps, reduction of stigma over time is more encouraging. Individuals in wheelchairs are less stigmatized than they were in the past (Harris, L., et al., 1991; Nabors, 2002; Smart, 2001). Nevertheless, numerous examples are reported where public education campaigns apparently have not changed attitudes to the extent expected (Gelder, 2001; Harris, Walters, & Waschull, 1991; Lee, 2002).

Campaigns utilize a variety of forums and strategies though which to change attitudes. Some are electronic or print media, famous people with disabilities, school curricula, and professional or self-help group advocacy. To effectively evaluate the success of forums and educational strategies, valid and reliable measures of public attitudes must be developed prior to use of techniques for change. With such measures, relevant attitudes of target populations may be measured before and after campaigns to obtain objective indices of the range of success for each of the various forums and strategies (e.g., Zavecz & Halasz, 2001). This information may be used to improve future public awareness/education or advocacy activities, not only for the conditions targeted in specific campaigns, but also for other stigmatizing conditions.

Needs and Challenges in Measuring Attitudes Related to Stigma Cross-Culturally

>Global research. Most of the past research on public attitudes has been in developed countries (Crisp, et al., 2000). Increasingly, investigators have recognized a need to measure stigma and its effects in less developed nations as well (Ellsberg, Pena, Herrera, Winkvist, & Kullgren, 1999; Link, 2002; Mohit, 1999). For example, several large studies have documented that attitudes toward mental illness are different in populations around the globe (Thompson, Stuart, Bland, Arboleda-Florez, Warner, Dickson, 2002). Studies documenting stigma related to stuttering (see below) have focused on North America, Western Europe, and Australia. Less well-known reports suggest similar results in other parts of the world, including Asia, Africa, the Middle East, and South America (e.g., studies reviewed by Cooper & Cooper, 1993; Jin, 2001).

Translation to other languages. Cross-cultural research presents special challenges. One of these is to translate survey items from one written language precisely to another language. Rogler (1999) addressed the issue of translating survey questionnaires from English to Spanish and pointed out a number of examples in which there are inherent methodological difficulties in meeting conventional requirements for standardization in paper-and-pencil survey instruments.

Standard instruments. In spite of the difficulties involved, there is a need for standard instruments. Whereas excellent research has been carried out with existing surveys, e.g., in the area of mental illness (Gelder, 2001; Thompson, Howard & Jin, 2001), it is difficult to compare findings across investigations. A necessary step in developing a science of reducing stigma involves the development of quantitative and qualitative methods that can investigate and measure the range of severity of stigma within and between conditions as well as its temporal variability. Such measurement can become the base for estimating the effectiveness of interventions (Weiss & Ramakrishna, 2002). As noted above, even though the literature contains numerous survey reports in specific countries, few standard and widely accepted public opinion instruments have been used to measure public attitudes across a diverse variety of human conditions or attributes in different countries and in different languages. The IPATHA project seeks to take that important step.

In the case of stuttering, no standard measures have been widely used to examine public opinions and attitudes or to establish baseline data against which to measure changes in attitudes, beliefs, and reactions. There are at least two important implications for this lack of baseline data. First, it has been impossible to determine which communities, regions, and societies are more or less knowledgeable or negative in their views about stuttering and therefore, where education efforts might be targeted. Second, without baseline data it is difficult to determine if public education initiatives have achieved their desired effects.

Survey Methods: Stuttering

Most measures of public attitudes or stigma ask people about various aspects of disorders or their own reactions to them. In studies of stuttering, a wide variety of data collection methods have been used: paper-and-pencil questionnaires (Blood, Blood, Tellis, & Gabel, 2003; Gabel, Blood, Tellis, & Althouse, 2004; Hulit & Wertz, 1994; Klein & Hood, 2004), semantic differential scales (Doody, Kalinowsky, Armson & Stuart, 1993), questions to store clerks who had just spoken with a severe stutterer (McDonald & Frick, 1954), face-to-face interviews with people on the street (Van Borsel, Verniers, & Bouvry, 1999), telephone interviews (Craig Hancock, Tran, & Craig, 2001; Ham, 1990), open-ended written statements (Ruscello, Lass, Schmitt, & Pannbacker, 1994), and extended tape-recorded interviews (Corcoran & Stewart, 1998; St. Louis, 2001), among others. A similar range of methods has been used with other conditions, and other innovative methods have been reported as well. For example, health professionals have directly rated the amount of stigma associated with health conditions (Bramlett, Bothe, & Franic, 2003; Ustun, et al., 1999).

Instrument Criteria and Considerations

After reviewing established international opinion surveys and contemporary survey research principles (Babbie, 1990, 2004; Dillman, 1978; Dillman, 2000; Quine, 1985; World Values Study Group, 1990-93) the task force established a number of instrument criteria. One primary requirement is that the survey can be translated to different languages and, thereby, be usable in a wide range of cultural settings. A second requirement is that subsequent translations must meet acceptable standards of reliability and validity. In early discussions, the task force recognized the value of providing respondents with exemplars, such as hearing an audio or video clip of a

person stuttering or of providing a standard definition of what is to be judged. This seemed important since stigma is usually associated with the severity of a condition and since respondents may not know anything about the condition. On the other hand, task force members recognized that standardization of exemplars or definitions would be extremely difficult, especially if they were to be translated to other languages, and that definitions are often ambiguous (e.g., ASHA, 1999). A third requirement is that the methodology for measuring public attitudes be accommodated by available personnel, partners, and financial resources in the US and abroad without major external funding.

Therefore, a written questionnaire format was chosen that neither defines nor provides exemplars of the conditions, but allows respondents to indicate that they do not know about the disorder. Despite some threats to the validity of results, e.g., respondents not realizing that stuttering might include speech with silent blocking, the task force believed this to be the best choice considering reliability, expense, availability of telephones or computers, adaptability, and the need for translation to other languages. A hard-copy survey could meet four major challenges to elicit objective nation-specific data in ways that would be (a) interpretable from probability and non-probability samples (described later), (b) obtainable for reliability and validity measures, (c) familiar to other cultures and countries, and (d) readable and amenable to the translation circle from American English to other languages and to back-translations in English, thereby permitting tracking and evaluation of semantic variability in languages. The text that follows reports development and initial field-testing of the first American English prototype of the survey instrument and on other practical and methodological issues. These were considered important first steps in developing an empirically-based survey instrument that (a) applies theory and methods from population research in epidemiology and other health and social sciences with regard to issues such as respondent selection and sampling (Lubker, 1997; Lubker & Tomblin, 1998), (b) conforms to accepted ethical and methodological standards of survey research, (c) conforms to accepted standards of reliability and validity, (d) allows translation into different languages for multi-national use, and (e) allows quick and efficient analysis by investigators.

Purpose

This study was undertaken to field test the first prototype of the POSHA-E using respondents in a nonprobability (i.e., "convenience") sample. Specific research questions of the field study were:

Are stuttering ratings affected by the order of occurrence of stuttering versus other attributes in the questionnaire?

Can systematic results be achieved through convenience sampling when independent research partners distribute questionnaires?

Is a quasi-continuous rating scale efficient for respondents and data tabulators?

Does convenience sampling yield representative demographic characteristics?

Are respondents' comments suggestive of a user-friendly survey instrument?

Method

Questionnaire

Content and Format

The POSHA-E consisted of five components: First component: Instructions. The survey began with a one-page instruction sheet. To minimize the likelihood that respondents might respond differently knowing that the primary intent was to determine public opinion about stuttering, nowhere in the questionnaires was it stated or implied that the questionnaire was designed to obtain information about stuttering per se. The instructions used the word "opinion" throughout rather than "attitude" since in American English, "opinion" has the more neutral connotation.

[FIGURE 1 OMITTED]

Second component: General section about nine human attributes. (Beginning in this section, various terms are italicized in this somewhat unconventional and necessarily detailed paper to assist the reader in knowing when such terms refer to specific components, sections, and orders.) The one-page general section had questions and prompts eliciting ratings of nine human attributes, i.e., stuttering and eight other attributes or "anchors." Task force members decided that the final POSHA would be most sensitive cross-culturally if stuttering could be placed in context with other stigmatizing and nonstigmatizing attributes. That is, stuttering is stigmatizing as compared to what? They further assumed that these anchors would range from negative to positive on the prototype survey: mental illness, wheelchair use, overweight, old, left-handed, good talker, multilingual, and intelligent. The general section prompts asked respondents to rate: (a) their overall impressions of stuttering and the other eight anchor attributes, (b) the extent to which they wanted to be or have each attribute, (c) their knowledge of each attribute, and (d) individuals they know who manifested each attribute. Based on previous research (St. Louis, 1999), respondents were likely to give lower ratings for negative attributes when asked if they "wanted to be or have" the attribute compared to their "first impression." Figure 1 illustrates a sample of respondents' ratings for the general and other components of the POSHA-E. Additional information about this second component occurs in a later discussion of item order.

Third component: Detailed sections containing three of the nine human attributes or triads. The third component was lengthy, comprised of three detailed sections (or, more precisely, three detailed section triads [explained below]) of two or three pages each. See samples in Figure 1. The detailed stuttering section prompts elicited ratings on: (a) sources of knowledge, (b) causes, (c) sources for help, (d) reactions and feelings in the presence of people who stutter, (e) concern if selected people stutter, and (f) beliefs about stuttering and people who stutter. All of these areas are included in other surveys of attitudes toward stuttering (c.f. citations in section on Survey Methods: Stuttering). The detailed stuttering section served as a template for the other attributes (i.e., anchors), so that attribute-specific detailed sections were similar to the detailed stuttering section and to each other, but some prompts were deleted for specific anchors. For example, groups of items about where to get help were eliminated for the attributes of left-handed, multilingual, good talker, and intelligent since individuals with these attributes would not logically be assumed to need help.

Because a prohibitively large number of potential items would be required if each prototype questionnaire contained all nine attribute-specific detailed sections, a strategy was adopted wherein each respondent would rate only three detailed sections, i.e., a triad. On each questionnaire one of the three detailed sections was always a stuttering section so that we would have complete stuttering data from all respondents. The other two detailed sections of each questionnaire elicited ratings for two of the other eight anchors. The triads varied in order and content across groups of questionnaires (explained below). Other details on construction of this third component appear in the discussion of triad order.

Fourth component: Demographics. All questionnaires ended with a three-page demographic section comprised primarily of nominal variables. The demographic section contained the following prompts for each respondent: (a) date of birth, (b) current and birth residence, (c) citizenship, (d) living arrangement, (e) sex, (f) languages spoken, (g) education, (h) career or vocation, (i) religion, (j) ratings and comments about the respondent's health and other abilities, and (k) how and where the questionnaire was obtained.

Fifth component: Comments. The final page provided space for respondents' optional comments.

Quasi-Continuous Rating Scale

The first POSHA-E prototype used a quasi-continuous response scale for ratings. The instructions showed respondents how to draw a vertical line on continuous dotted line scales with printed markers at each end and in the exact middle. (See Figure 1.) The scale lines were printed to the right of each questionnaire item. The ends of the dotted line scales were designated by descriptors such as "very negative" and "very positive" or "definitely 'no'" and "definitely 'yes.'" The middle of the scales had descriptors such as "neutral" or "not sure." A quasi-continuous scale was chosen for the first prototype in order to sample potentially small differences in attitudes and to employ data analyses not possible with interval data. For example, scale responses could be analyzed with ratio data rather than with interval data (e.g., Siegel, 1956).

Respondents' ratings were converted to numeric data by student and paid data entry assistants. In a procedure described by Breitweiser and Lubker (1991), assistants placed a transparent ruler, on which were printed numeric values, from 0 to 100, over each item on the survey scales and noted the locations of respondents' vertical marks. They then wrote the ruler's number, from 0 to 100, closest to the respondent's mark in the margin to the right of the scale. Respondents' marks at the middle marker (i.e., "neutral") were recorded and tallied as 50. After ratings were converted to numeric values, assistants entered values for each section of each questionnaire into Microsoft Excel worksheets for analysis.

Other, non-scaled items also appeared on the survey. For example, in the last set of prompts in the general section, respondents checked responses about several categories of individuals who might manifest each attribute. The choices included "nobody," "acquaintance," "close friend," "distant relative," "close relative," "me," or "other." The questionnaires also elicited additional information under "other" categories in selected detailed sections, and the demographic section sought detailed information about a wide range of variables.

Length of the Prototype Instrument

The printed length of the POSHA-E ranged from 12 to 14 pages. The reason for the differences was that the detailed sections for the anchor attributes of left-handed, good talker, multilingual, and intelligent were one page shorter than the detailed sections for stuttering and the other attributes. As noted, each POSHA-E contained triads of a detailed stuttering section and two of the other attributes. The possible number of ratings and other responses on the questionnaires ranged from 257 responses to 376 responses, depending on differences in the number of items in triads and on demographic and experiential characteristics of respondents.

Counterbalancing and Randomization Scheme

The first purpose of this field study was to determine the extent to which the order of stuttering ratings was affected by ratings for other attributes. Stuttering ratings occurred both as items along with all the other anchors in the general section and as detailed sections in triads of detailed sections that followed. Since the detailed sections had 85 items, it was far beyond the scope of this investigation to systematize all possible item orders within the detailed stuttering sections. This field study analyzed various possible order effects for the general section and the order of detailed stuttering and other attribute sections in the triads. For example, would respondents be more or less likely to "want to have" stuttering if it were the first item rated, the middle item rated, or the last item rated? Similarly, would opinions about the causes of stuttering be affected if respondents had already speculated about the causes of mental illness versus left-handedness? To reduce item order influence, any potential respondent--or groups of surveys with respondents--should not have an appreciably greater chance of receiving one item order over another item order. That is, distribution of varying item order survey should be randomized.

[FIGURE 2 OMITTED]

Accordingly, three levels of randomization were used. First, a systematic counterbalancing and item randomization scheme controlled for effects of item order of individual items among the four prompts in the general section and for the triad order of entire detailed sections devoted to stuttering or other anchors. The scheme is summarized in Figure 2. Four prompts appeared in the general section: (a) "My overall impression of a person who ...", (b) "I would want to be a person who ...", (c) "The amount I know about a person who ...", and (d) "People I have known who ..." Each prompt was followed by nine line items eliciting ratings of the nine attributes, one of which was stuttering. Stuttering and all attributes held the same order at all prompts on a given survey. However, across survey versions, the place of stuttering in this item sequence was rotated so that stuttering might be the first item for all prompts, or the fifth, or the ninth item at all prompts in the general section.

Within each of these three item orders in the general section, a triad order of detailed sections was varied systematically so that the detailed stuttering section appeared as the first, second, or third section rated, and so that each of the remaining eight detailed anchor sections occurred with all the others in all possible combinations and occurred first or second. The latter pairs of eight anchors yields 56 possible combinations (8 attributes x 7 combinations of two attributes) as shown in Figure 2. One-third of the questionnaires represented each of the three item orders

(first, middle, and last position), each of three triad orders of the subsequent detailed section on stuttering (first, second, or third), and the various combinations of the other attributes. Thus, 504 possible orders existed for a complete set of parallel POSHA-E questionnaires (3 general item orders x 3 detailed triad orders x 56 anchor combinations within the triads).

Second, content of packets with different survey versions was systematically varied before packets were given to partners/distributors for distribution. Research assistants prepared and collated packets in which successive questionnaires followed the various counterbalanced orders according to the samples shown in the bottom of Figure 2. In each packet of 14 or more questionnaires, adjacent POSHA-Es systematically varied in terms of general item order and at least two iterations of all anchors in the triad order. Third, the collated packets then were numbered and selected randomly to be given to distributors (explained below) who were "blind" to packet contents and questionnaire sequences (i.e., they had no knowledge of which respondent might receive which version of the questionnaire). Distributors were instructed to give out questionnaires sequentially, without mixing up the packets' order. All respondents received packets to which variously ordered surveys were randomly allocated.

Distribution of Questionnaires

Nonprobability Respondent Sampling Scheme

Respondent sampling procedures are emphasized in survey research and in the health sciences, particularly in epidemiology (Greenberg, Daniels, Flanders, Eley, & Boring, 1993; Gordis, 1996). Simple random sampling to select survey respondents seems to be the sine qua non of probability sampling in survey research, unrealistic though it sometimes may be; certainly probability sampling of several kinds (e.g., systematic sampling, stratified sampling, or block sampling) has many advantages. The task force decided to field test the POSHA-E using nonprobability (i.e., "convenience") samples even though the long-range goals of the IPATHA initiative include development of satisfactory methods for probability sampling of public opinion as well (St. Louis, 2005). As noted earlier, this field study was a vehicle to establish consistent distribution procedures for international partners and to set a model for determining return rates for surveys. We hoped to identify problems inherent in nonprobability sampling that we would encounter and early on to make some decisions about controlling through analyses. This was relevant since, at the time of this field test, partners in Bulgaria, Turkey, and Brazil had already agreed to help with translations and to accept surveys for distribution.

Distribution Procedures

A central distributor gave questionnaire packets to other distributors who in turn gave individual questionnaires to potential respondents who either completed them or, in a few cases, recruited others to complete the survey. This is described by Dillman (1978) as the "person in charge" method. Specifically, the first author (KSL) gave packets of questionnaires to seven university "partners." Four students were enrolled in an honors class, and two were engaged in independent research experiences. The seventh student was visiting from another institution. The students assumed the roles of distributing and collecting that international partners are expected subsequently to perform. The student partners distributed the questionnaires to family members,

friends, acquaintances, and small groups. The only inclusion criteria were that survey recipients must be able to read English, be at least 18 years of age to comply with approved human subject protection policies, and be willing to consider filling out the POSHA-E. Four student partners received packets for distribution of 56 collated POSHA-Es each. They suggested this to be "... as many collated POSHA-Es as they believed they might be able to hand out to individuals likely to complete and return them." Three student partners accepted 20, 22, and 37 questionnaires, respectively).

Results

As noted, the purposes of this field study were to field test the first prototype of a survey instrument to measure public attitudes toward stuttering with special attention to (a) potential order effects inherent in the questionnaire; (b) a partner-recruitment questionnaire distribution pattern; (c) problems in scoring, tallying, and accuracy of responses on a quasi-continuous scale; (d) readability; and (e) comments and suggestions from respondents.

Respondent Characteristics

Return rate. Of 303 questionnaires distributed, 165 (55%) were returned. Distributor-specific return rates ranged from 35% to 89%. This response rate was not as uniformly high as some survey experts suggest (e.g., Dillman, 1978), but was high enough to obtain sufficient analyzable data when returns from all partners were combined into a single sample. This was encouraging especially since the survey was long. Subsequent inspection of demographic data suggested that the nonprobability sampling scheme resulted in respondents representing a wide variety of occupations and ages, the majority of whom were not members of distributors' extended families. That is, all partners were able to recruit respondents not only among family but also among friends, acquaintances, and friends of friends or relatives who were willing to complete the survey.

Demographic characteristics. Table 1, column 2, summarizes frequency distributions for demographic variables. The majority of respondents were women (71%) and had a mean age of 31.5 years. They were well educated, reporting a mean number of 15.3 years of education. One half were single and one-third, married. Thirty-six percent were students with the rest working for a salary. Virtually all (99%) reported English as their native language, and relatively few spoke other languages. Ninety-three percent reported their race as Caucasian, and 85% indicated their religious affiliation as Christian. From the fourth item of the general section where respondents were asked who they knew with each of the attributes, very few respondents in indicated that they were stutterers, mentally ill, wheelchair users, or old (2-5%). Somewhat more respondents indicated that they were overweight, left-handed, or multilingual (10-26%), and even more respondents marked that they were good talkers (30%) and intelligent (48%). Overall, as seen in such areas as language spoken, religious affiliation, and race, this was a relatively homogeneous group.

As noted, the intent of the nonprobability sampling procedure was not to achieve a representative (i.e., probability) sample but to explore the feasibility of methods that can be used in many cultures and in developing countries. Nevertheless, information about the extent to which the

respondent sampling procedure produced a respondent sample deviating from census data could be used to inform analysis and utility of such sampling in interpretation and comparisons with public opinion data. Table 2 was prepared to compare the respondents in this field study with data from the US Census (US Census Bureau, 2000). It is clear that the collection procedure produced a sample of respondents, 67% (111/165) of whom were from West Virginia (WV), that was quite different from census characteristics of (a) the host county, (b) other counties representing at least 3% of the respondents, or (c) the state. US Census data indicate that only 15% of WV adults over 25 years report a bachelor's degree or higher. The table shows that the percent of ?25-year-old respondents whose demographic information reported at least 16 years of education (assumed to be generally equivalent to a bachelor's degree) in the sample groups is generally higher than would be expected in a probability sample of adults from counties and the state. The only notable exception occurred for the host county (Monongalia)--where 27% of the respondents resided--in which US Census data indicated that 32% of adults over 25 possessed bachelor's degrees or higher. More important, Table 2 reveals that county and state populations have an approximately equal sex distribution, different from the unbalanced distribution of the respondent sample, which contained about 2 1/2 times more females than males.

POSHA-E Order Effects

Distribution of orders in returned questionnaires. Of the 165 returned questionnaires, 46.7% were characterized by item orders in general sections that contained stuttering in the middle position, 20.0% had stuttering in the first position, and 33.3% in the last position. Obviously, the position distributions in returned questionnaires did not achieve even distributions across the three positions, even though respective percentages of distributed questionnaires were 32.3%, 37.0% and 30.7%. Nevertheless, the returned questionnaires had a more evenly proportionate distribution of positions in triad orders. The stuttering section occurred first among the triads in 30% of returned questionnaires, second in 36%, and third in 33%. Moreover, all 56 possible orders of the two detailed attribute (anchor) sections occurred in the 165 returned questionnaires. The frequency of occurrence for each possible order averaged 2.9 occurrences across all orders, which is the expected occurrence (165 / 56 = 2.9) for an equal distribution of orders. These results are not surprising given that individual distributors received packets for only one item order but up to all of the various triad orders.

General section order effects: Scaled ratings. In order to generate one comparable score, overall means of ratings for all nine attributes were combined from general section scale (i.e., 0-100) responses for comparisons among each of the three item orders, which are shown in row 1 of Table 3. The similarity among mean ratings of items appearing in each of the three orders, ranging from 54.1 to 56.6, indicate that the item order of attributes had little effect on the ratings.

These means reflect the combined ratings of all nine attributes across all 27 scale items from the first three prompts in the general section. To investigate potential order effects further, all pairwise t tests for independent observations compared the mean ratings of items eliciting opinions on overall impression, wanting to be, and knowledge about each attribute. For example, were there differences in ratings of overall impression for mental illness when stuttering occurred first versus last, or first versus middle, or middle versus last position? Of the 81 t tests, 74 (91%) were nonsignificant or had probability levels that were equal or greater than .05, with a

mean probability level for the t tests of .48. The seven order pairs that were significantly different were not distributed in any recognizable pattern: stuttering, 1 of 9 pair-wise tests; overweight, 1 of 9 tests; multilingual, 3 of 9 tests; and mental illness, 2 of 9 tests. Nonsignificant comparisons in 91% of the tests support a conclusion that there were no systematic order effects for the order of stuttering in the general section.

It should be noted that a correction for multiple t tests, such as the Bonferroni correction (Maxwell & Satake, 1997) to minimize the probability for making a Type I error was not applied in these and in subsequent parallel comparisons. In these cases not correcting the alpha value is the more conservative alternative since our emphasis was on not making a Type II error (i.e., accepting--not rejecting--the null hypothesis that there is no difference between pairs of means) or concluding that a difference for first, middle, or last position for the stuttering attribute existed when in fact there was none.

Detail section order effects: Scaled ratings. In spite of the foregoing, it is evident that the order of stuttering items in the general section could potentially affect responses to stuttering items in subsequent detailed sections for stuttering. Again, to generate one comparable score, overall mean ratings of all judgments in the detailed sections of stuttering are shown in row 2 of Table 3 and were very similar, ranging from 42.9 to 43.7. Again, t tests compared all pairs of orders for the 85 scale items in the detailed stuttering section. Of the 255 pairs (85 items x 3 comparisons = 255), 247 (97%) were not statistically significant at the .05 level, with probability levels averaging .52.

Finally, and perhaps of greater interest, the overall mean ratings for detailed stuttering items were analyzed separately according to its occurrence in first, second, or third positions in the triads (row 3 of Table 2). Again, very few differences were found; 235 (92%) of the t test comparisons were nonsignificant. The mean probability level was .45.

General Section order effects: Categorical ratings. We also compared the percentages of respondents who knew someone with each attribute in the general section. (i.e., "nobody," "acquaintance," "close friend," "distant relative," "close relative," "me," or "other"). We ran Chi Square tests of the independence of expected versus observed frequencies for these experience categories. The categories were corrected for actual proportions of occurrence by adjusting the actual frequencies according to the number of returned questionnaires for each item order so that each set would represent one third of the respondents. The relevant data for item order are shown in Table 4.

These results were highly significant when stuttering appeared as the first or middle attribute in the general section item lists of nine attributes ($p < .00001$) but were not significant when stuttering appeared at the end of the list ($p = .98$). Table 4 indicates that respondents were most likely to categorize individuals (or themselves) with the various attributes (including "nobody") when stuttering was the middle attribute. They were less likely to do so if stuttering occurred last but least likely when stuttering occurred first. Surprisingly, these differences in personal experience with attributes within the three orders occurred even though the attitudes regarding them did not (see Table 3). We cannot explain this difference but speculate that those distributors having POSHA-Es containing stuttering as the middle attribute of general items were more likely

to distribute questionnaires to college students in speech-language pathology or other similar helping professions who seem to be more predisposed to carefully identify various attributes. Unpublished data from a preliminary report of a subset of the data reported here indicates that students identified many more people they knew with the various attributes than did adult nonstudents (St. Louis, Schiffbauer, Phillips, Sedlock, Hriblan, & Dayton, 2000). Order effects: Summary. Overall, the counterbalancing appears to have provided a reasonably consistent sampling of all possible orders. More important, it is quite clear that such counterbalancing had little if any systematic effect on the results, except perhaps for the unexpected findings of differences in categorization frequencies of various attributes in the previous section. We concluded that counterbalancing presentation orders was essentially unnecessary for future field studies and that respondents rated items similarly regardless of the order in which they were presented.

Accuracy and Care in Responding

Marking problem on the quasi-continuous scales. All correctly marked items were retained in quantitative analyses. However, some respondents seemed not to mark the quasi-continuous scale for POSHA-E items with specific attention to each item. We inspected all questionnaires for evidence of respondent markings that might bias results. Of the 165 returned questionnaires, 52 (31.5%) contained instances in which respondents had drawn an unbroken line through three or more questionnaire items. The affected questionnaires had a mean of 5.5 instances of lines drawn through multiple items. These instances were tallied as legitimate responses unless marks appeared outside the printed response range or in other uninterpretable forms. Sometimes it appeared that respondents had intended to mark the extreme left, middle, or extreme right of the scale but drew a non-vertical line that, when measured and scored, was likely a higher or lower rating than what had been intended.

The average number of scale responses was 245 per questionnaire calculated from 14 randomly selected questionnaires that were counted individually. Of the total responses, fewer than 1% of the ratings could be regarded as suspect for the above reasons [287 errors / (245 items x 165 respondents) x 100% = 0.7%].

Respondents occasionally rated two items on the same scale line. In these instances, no responses were tallied for either item. Nine questionnaires (5.5%) contained at least one of these kinds of errors, with a total of 10 errors that represented only 0.02% of all scale ratings.

Based on the very small percentages of presumed errors, we concluded that respondents generally marked the quasi-continuous scale with sufficient care that the tallied responses could be considered valid measures of their opinions. Nevertheless, the number of questionnaires that contained at least some problems (nearly one-third) ultimately led the task force to consider a different response mode.

Missing data. In another problematic response pattern, several items or, more typically, entire prompts with all following items were left unrated. Incidents where respondents appeared to have omitted one or two specific items but answered others were not counted. Also not counted were those associated with turning two pages at a time thereby omitting an entire page. Eight

(4.8%) of the 165 questionnaires contained 11 separate omissions of four or more consecutive items following a prompt or question. There were approximately 10 questionnaire items per prompt; therefore, a total of 11 blank sections could be only as high as 0.3% of the total number of potential responses [(11 sections x 10 items) / (245 items x 165 respondents) x 100% = 0.3%]. In fact, the true percentage was somewhat less than this because, in some cases, a few of the items in a group containing consecutive omissions were marked and tallied. We considered this percentage of missing data negligible.

Tallying consistency. To measure the consistency with which data entry assistants carried out measuring and scoring continuous scale values, a trained research assistant rescored five randomly selected returned questionnaires. Using a transparent ruler identical to that used originally, she re-measured each scale response and either agreed with the original score or recorded the value and the magnitude of difference from the original value. If the original value was greater, it was recorded as a plus value; if less, as a minus value. She then counted the number of original ratings and other responses (e.g., other comments and biographical data). On the five rescored questionnaires, there was a mean of 242.8 scaled ratings (range = 218 - 278) on the quasi-continuous scales and a mean of 38.8 other categorical or written responses (range = 26 - 56), for a total of 281.6 average responses for the five questionnaires (range = 244 - 317). For the scaled responses, a mean of 6.8 tallying disagreements (range = 0 - 30) was found in the five questionnaires with a mean combined (plus and minus) scale value of +21.8 (range = -12 to +144). The average magnitude of error was +0.96, or about 1 out of 100, with a range of -6 to +3.8. If we assume that each potential rating difference could be as high as 100, the contribution of the disagreements to the overall data set is virtually zero (.96 / (242.8 responses X 100)) = .00004). Based on these analyses, we believe that the detailed task of measuring and tallying of the questionnaires was carried out with sufficient consistency to have no adverse effect on the results.

Readability

Several authorities on survey research have addressed the importance of readability, wording and question clarity in questionnaires (e.g., Babbie, 1990; Dillman, 1978, 2000; Fowler, 1995). A general principle among experts is "keep it simple." However, Dillman (1978; 2000) also discussed the pitfalls of over-simplification in wording and item construction.

Components of the POSHA-E were analyzed according to the computerized Flesch Reading Ease Index that is based on the average number of syllables per word and the average number of words per sentence (Evaluating Readability Statistics, 1993-94). Index scores range from 0 to 100, wherein standard writing (e.g., the target level for most newspapers) averages 60 to 70. The higher the score, the greater the proportion of the population who can readily understand the document. A 60-70 Flesch readability index corresponds to writing at the seventh to eighth grade level.

The readability rating of the POSHA-E instructions component was 58.2 or at the 9.4 grade of school reading levels. This score means that populations who read at a mid-ninth grade level or higher would understand the instructions. Flesch readability estimates of the general and detailed sections ranged from 53.4 to 88.3 (mean = 70.6). In other words, text grade-level ranged from

2.8, or early second grade, to 7.8 or nearly eighth grade with a mean of 5.3 or early fifth grade level.

Comments from Respondents

Thirty of the 165 respondents (18.2%) wrote comments about the POSHA-E at the end of their questionnaires. Seven (4.2%) commented on the limited response options, such as the lack of a "doesn't matter" or that marking a 1-5 scale would be better. Six respondents (3.6%) highlighted examples of unclear questions; another six stated that the questionnaire was too long. Five respondents (3.0%) described the questionnaire as boring, uninteresting, making no sense, or dumb. Another five criticized the requirement to generalize to all persons with a given attribute. Finally, eight (4.8%) suggested other topics for inclusion in future surveys.

Summary

This field study generated valuable data for proceeding with the IPATHA initiative. Whereas surveys are often generated and tested with limited attention to methodological issues presumed not to affect results, the task force believed that the POSHA-E must undergo rigorous evaluation of such concerns. In general, order of stuttering items did not affect the results in any large or systematic ways. With a carefully constructed counterbalancing and distribution scheme, convenience sampling could achieve unbiased distribution of different versions of the POSHA-E. A quasi-continuous scale can be used to measure opinions of stuttering and other attributes; however, respondent errors and omissions must be noted, even though they likely do not affect overall response results. Errors in measuring scaled responses with a special ruler and errors in tallying do not apparently affect results. Nevertheless, we concluded that the inconvenience, time, and difficulty of measuring and scoring a quasi-continuous scale is not consistent with the IPATHA goal of simplicity and user-friendliness. Furthermore, problems associated with computer formatting adaptations, spacing, varying sentence length, and pagination in various iterations of the survey in multiple languages become evident for ruler calibrations.

Not surprisingly, the model of convenience sampling employing partner distributors did not achieve a respondent sample that was representative of the geographic population of adults; however, it did achieve a sample that reflected a considerable range of ages, occupations, religious affiliations, marital status, and other demographic variables. Finally, although most respondents filled out the questionnaire without comment, those who did comment were likely to point out problems of length, confusion, and difficulty.

Discussion

Overview

A field study was conducted to field-test the first experimental prototypes for the Public Opinion Survey of Human Attributes (POSHA-E), a paper-and-pencil instrument designed to identify public attitudes toward stuttering and other attributes among multi-national populations. This report does not focus on attitudes but on characteristics of the questionnaire and very importantly on identification of methodological problems that we expected in cross-cultural research. Field-

testing the POSHA-E helped to identify its strengths and weaknesses with regard to procedural efficiency, sample representativeness, order effects, scoring ease and accuracy, data reduction, and response reliability and validity. A number of companion field studies informed by these efforts are already under way.

Respondents

The nonprobability sampling scheme was carried out by seven different student partners, each distributing questionnaires to acquaintances, friends, family members, and others known by these individuals. The most important caution of these data collection procedures is that they produced a respondent sample biased by selection. The selection resulted from cultural, social, and educational characteristics of the selecting partner. A major consequence is that results from analyses of this respondent sample are not generalizable to public opinion and attitudes. This sample helps to inform the need for control strategies when data are analyzed. The return rates from 35% to 89% per partner suggests that this strategy could effectively be replicated with other national and international partners.

Questionnaire

Length

"There is a widespread view that long questionnaires ... should be avoided" (de Vaus, 2002, p. 112). A review of the evidence, however, indicates that there is little research to support this commonsense assumption. Seventy-five percent response rates for mail questionnaires with as many as 24 pages are reported (Dillman, 1978). The POSHA-E in this and on-going studies was quite long (i.e., up to 14 pages with 376 items) in order to obtain extensive field test data on many items and to inform decisions about reducing the questionnaire to a shorter and user-friendly version. None of the student partners reported that potential respondents refused to complete the survey after seeing its length, although refusal may well have occurred by not turning in the survey. Future field tests will include a query about completion time. Furthermore, respondent comments about the tedious nature of the instrument dictate strong consideration of shortening the questionnaire.

Order Effects

The distributed questionnaires in this field study contained uniform representations of three triad orders for detailed stuttering section and various permutations of the two detailed sections of the other eight anchors. The exception is that an approximate equal representation was not achieved for the three item orders of stuttering in the general section.

The field study indicated quite clearly that the order of presentation achieved by counterbalancing strategies had little effect on respondents' ratings on any measures or any sections. Results suggested that in subsequent field tests it is appropriate to utilize only one item order in the general section. Moreover, there was no concern about potential order effects as a function of the detailed stuttering section's place in the questionnaire.

Readability

The principles of readability have important implications for those who seek to improve knowledge, to change attitudes and to measure these changes. The questionnaire readability rankings seem not to be a problem for this sample of respondents who, as a group, are more highly educated than the population in the surrounding community. The levels of readability may pose significant problems for less educated respondent samples. The inconsistency between tenth grade level of the instructions, mean fifth grade level of the survey items, and readability inconsistency among items will be addressed in subsequent field tests. An essential consideration is whether instructions at a high school level are useful among less well-educated populations whose knowledge and opinions about stuttering and other characteristics may be underrepresented in results of this and other surveys. Their knowledge, beliefs, and susceptibility to change agents may be significantly different from those in more highly educated strata.

Rating and Tallying Accuracy

The quasi-continuous scale occasioned a small but appreciable number and variety of inaccurately marked responses. Thus, methodological concerns for response attrition and for unusable responses were raised. While respondents' recording errors had inconsequential effects on the results, it became increasingly evident that the quasi-continuous rating scale too often produced invalid, unusable responses. The data reduction procedure was cumbersome and time-consuming for the quasi-continuous scale, requiring 20-40 minutes for each questionnaire. For this reason, if for no other, an interval scale will be considered for the final POSHA instrument.

A sample of re-measured and retallied questionnaires in the study indicated that inconsistencies between data entry assistants and a validator were negligible. Furthermore, comparisons of responses to the quasi-continuous scale with the 9-point EAI scale found similar attitude ratings.

Future Research

Other field tests of POSHA-E are under way around the US and in other countries (St. Louis, 2005). Investigations are under way to estimate reliability and validity of the instrument. POSHA-E translations to other languages have also begun. We will compare survey results from translated versions to results from American English versions and identify problems inherent in translations, which despite our best efforts to eliminate them may be influenced by American English idiomatic expressions.

When additional field test data are available, item analyses will determine which items are most discriminating and which are redundant or not useful. The next activity is to develop a shortened, revised version of the POSHA-E and to field-test it with convenience and probability samples here and abroad. Inclusion of salary strata for the demographic section is under consideration to have direct SES assignment rather than having SES assigned from proxy variables.

With acquired sampling information, we are alerted to specific sampling questions for the larger international project. As noted above and shown in Table 2, respondents did not represent the area population. Educated persons and white females were over-represented. Future field studies

will be designed to investigate these distributions in other non-representative samples and compare these distributions with results of probability sampling methods in some settings.

Considering that many international samples will not be generated by probability sampling methods, the possibilities for control at analysis rather in design, weighting procedures for sample representation, and other sampling manipulations can be explored for data generated in other cultures and communities.

This initial field study represents progress in moving toward the long-range goal stated in the introduction, that is to explore specific methodological issues in the ongoing process of developing an instrument to measure public attitudes toward stuttering. In this process, we are amassing a vast amount of data from on-going field studies that can be analyzed to identify country- and culture-specific attitudes and biases toward stuttering, other conditions, and differences. Inefficiency with the quasi-continuous scale and the use of nonprobability samples notwithstanding, this and other field studies under way allow us to carry out analyses of comparative data on attitudes toward stuttering within the context of other human attributes. Such contextual information can provide preliminary baseline data to inform strategies to modify negative attitudes. Accordingly, we intend to generate baselines indicating where attitudes are more or less positive and how attitudes are associated with one other.

The data from these ongoing and subsequent studies will allow analyses of demographic variables to determine differences in attitudes in controlled, stratified and appropriately weighted analyses of respondent populations' characteristics. Such knowledge about factors associated with people's attitudes is postulated as an essential foundation for those attempting to change attitudes.

Acknowledgements

This paper is dedicated to the late Charles C. Diggs, who was on the original task force. He would have been a coauthor of this paper were it not for his untimely death in 2006. We wish to acknowledge the valuable assistance of Jami Ebbert, Sharon Helman, James D. Schiffbauer, Carolyn I. Phillips, Andrea B. Sedlock, Lisa J. (Hriblan) Gibson, Rebecca M. Dayton, Amanda K. Cassidy, Rachel Fender, Amanda K. Efaw, Maren Greig, and Brenda Davis in collecting, scoring, tallying, and analyzing data.

This research project was carried out at the Department of Speech Pathology and Audiology at West Virginia University using resources of the Department.

References

American Speech-Language-Hearing Association (ASHA). (1999). Terminology pertaining to fluency and fluency disorders: Guidelines. Asha, Supplement No. 19, 41-2, 29-36.

Babbie, E. (1990). Survey research methods. Belmont, CA: Wadsworth.

Babbie, E. (2004). The practice of social research. Tenth edition. Belmont, CA: Wadsworth.

Blood, G. W. (1999). The stigma of stuttering: Centuries of negative perceptions and stereotypes. Paper presented at the annual convention of the American Speech-Language-Hearing Association. San Francisco, CA.

Blood, G. W., Blood, I. M., Tellis, G. M., & Gabel, R. M. (2003). A preliminary study of self-esteem, stigma, and disclosure in adolescents who stutter. Journal of Fluency Disorders, 28, 143-159.

Bramlett, R. E., Bothe, A. K., & Franic, D. M. (2003). Measuring the impact of stuttering on quality of life. Paper presented at the Annual Convention of the American Speech-Language-Hearing Association. Chicago, IL.

Breitweiser, S.S. & Lubker, B.B. (1991). Hospital schools: A primer on management models and policy implications. Health Care Management Review, 16, 27-37.

Conture, E. G. (2001). Stuttering: Its nature, diagnosis, and treatment. Boston: Allyn & Bacon.

Cooper, E. B., & Cooper, C. S. (1985). Clinician attitudes toward stuttering: A decade of change. Journal of Fluency Disorders, 10, 19-33.

Cooper, E. B., & Cooper, C. S. (1993). Fluency disorders. In D. E. Battle (Ed.), Communication disorders in multicultural populations. Boston: Andover Medical Publishers.

Corcoran, J. & Stewart, M. (1998). Stories of stuttering: A qualitative analysis of interview narratives. Journal of Fluency Disorders, 23, 247-264.

Craig, A., Hancock, K, Tran, Y, & Craig, M. (2001). Prevalence and stereotypes of stuttering in the community. In H-G. Bosshardt, J. S. Yaruss, & H. F. M. Peters (Eds.), Fluency disorders: Theory, research, treatment, and self-help, Proceedings of the Third World Congress on Fluency Disorders in Nyborg, Denmark (pp. 588-592). Nijmegen, The Netherlands, Nijmegen University Press.

Crisp, A. H., Gelder, M. G., Rix, S., Meltzer, H. I, & Rowlands, O. J. (2000). Stigmatisation of people with mental illnesses. British Journal of Psychiatry, 177, 4-7.

De Nil, L. F., Kroll, R. M., Lafaille, S. J., & Houle, S. (2003). A positron emission tomography study of short- and long-term treatment effects on functional brain activation in adults who stutter. Journal of Fluency Disorders, 28, 357-380.

Doody, I., Kalinowski, J., Armson, J., & Stuart, A. (1993). Stereotypes of stutterers and non-stutterers in three rural communities in Newfoundland. Journal of Fluency Disorders, 18, 363-373.

de Vaus, D. (2002) Surveys in social research. Fifth Edition. London: Routledge of Taylor and Francis Group (58-69).

Dillman, D. A. (1978). Mail and telephone surveys. New York: John Wiley & Sons.

Dillman, D. A. (2000). Mail and telephone surveys: The tailored design method. New York: John Wiley & Sons.

Drayna, D., Kilshaw, J., & Kelly, J. (1999). The sex ratio in familial persistent stuttering. American Journal of Human Genetics, 65, 1473-1475.

Ellsberg, M. C., Pena, R., Herrera, A., Winkvist, A., & Kullgren, G. (1999). Domestic violence and emotional distress among Nicaraguan women: Results from a population-based study. American Psychologist, 54, 30-36.

Evaluating Readability Statistics (1993-94). User's Guide, Microsoft Word. Microsoft Corporation, p. 101.

Fowler, F. J., Jr. (1995). Improving survey questions: design and valuation. Applied Social Research Methods Series. Vol.38. Thousand Oaks CA: SAGE Publications, Inc.

Gabel, R. M., Blood, G. W., Tellis, G. M., & Althouse, M. T. (2004). Measuring role entrapment of people who stutter. Journal of Fluency Disorders, 29, 27-49.

Gelder, M. (2001). The nature of such stigmatization: The Royal College of Psychiatrists' survey of public opinions about mentally ill people. In A. H. Crisp (Ed.). Every family in the land: Understanding prejudice and discrimination against people with mental illness. Robert Mond Memorial Trust. Retrieved September 1, 2004 from http://www.stigma.org/everyfamily/mgelder.html.

Goffman, E. (1963). Stigma: Notes on the management of spoiled identity. Englewood Cliffs, NJ: Prentice-Hall.

Gordis, L. (1996). Epidemiology. Philadelphia, PA: W. B. Saunders Company

Greenberg, R.S., Daniels, S.R., Flanders, W.D., Eley, J.W., & Boring, J.R. (1993). Medical Epidemiology. East Norwalk, CT: Appleton Lange.

Ham, R. E. (1990). What is stuttering: Variations and stereotypes. Journal of Fluency Disorders, 15, 259-273.

Harris, L., & Associates. (1991). Public attitudes toward people with disabilities. New York: Author.

Harris, M. B., Walters, L. C., & Waschull, S. (1991). Altering attitudes and knowledge about obesity. Journal of Social Psychology, 131, 881-884.

Herek, G. M., Capitanio, J. P., & Widaman, K. F. (2002). HIV-related stigma and knowledge in the United States: Prevalence and trends, 1991-1999. American Journal of Public Health. 92, 371-377.

Hulit, L. M., & Wertz, L. (1994). The association of attitudes towards stuttering with selected variables. Journal of Fluency Disorders, 19, 247-267.

International League Against Epilepsy/International Bureau for Epilepsy/World Health Organization (ILAE/IBE/WHO). (2002). Bringing epilepsy "out of the shadows." Annual Report 2001 for the ILAE/IBE/WHO Global Campaign against Epilepsy. Retrieved September 1, 2004 from http://www.ibe-epilepsy.org/uploadEpyARbro.pdf.

Jin, X. (2001). Public awareness of stuttering in China. Paper presented at the 6th World Congress for People Who Stutter / 3rd International Stuttering Association Conference. Ghent, Belgium

Klassen, T. R. (2002). Social distance and the negative stereotype of people. Journal of Speech-Language Pathology and Audiology, 26, 90-99.

Klein, J. F, & Hood, S. B. (2004). The impact of stuttering on employment opportunities and job performance. Journal of Fluency Disorders, 29, 255-273.

Klompas, M., & Ross, E. (2004). Life experiences of people who stutter, and the perceive d impact of stuttering on quality of life: Personal accounts of South African individuals. Journal of Fluency Disorders, 29, 275-305.

Langevin, M. (1997). Peer teasing project. In E. C. Healey & H. F. M. Peters (Eds.), 2nd World Congress on Fluency Disorders Proceedings. Nijmegen, The Netherlands: The International Fluency Association.

Lee, S. (2002). The stigma of schizophrenia. Paper presented at the International Conference on Stigma and Global Health: Developing a Research Agenda. Bethesda, MD. Retrieved September 1, 2004 from http://www.stigmaconference.nih.gov/abstracts/LeeAbstract.htm

Link, B. G. (2002). On stigma and its public health implications. Paper presented at the International Conference on Stigma and Global Health: Developing a Research Agenda. Bethesda, MD. Retrieved January 8, 2008 from http://www.stigmaconference.nih.gov/abstracts/LinkAbstract.htm

Lubker, B. B. (1997). Epidemiology: An essential science for speech-language pathology and audiology. Journal of Communication Disorders. 30, 251-268.

Lubker, B. B. & Tomblin, J. B. (1998). Epidemiology principles and perspectives: Informing clinical practice and research on language disorders of children. Topics in Communication Disorders. 19, 1-27.

Maxwell, D. L., & Satake, E. (1997). Research and statistical methods in communication disorders. Baltimore: Williams & Wilkins.

McDonald, E. T., and Frick, J. V. (1954). Store clerks' reaction to stuttering. Journal of Speech and Hearing Disorders, 19, 306-311.

Mohit, A. (1999). Mental health in the eastern Mediterranean region of the World Health Organization with a view of the future trends. Eastern Mediterranean Health Journal, 5, 231-240.

National Association to Advance Fat Acceptance (NAAFA). (2002). Obesity research. Retrieved January 8, 2008 from http://www.naafa.org/documents/policies/obesity_research.html.

Nabors, L. A. (2002). The effects of brief interventions on children's playmate preferences for a child sitting in a wheelchair. Journal of Developmental and Physical Disabilities, 14, 403-413.

Porter, R. (2001). The history of stigmatization of the mentally ill: Is mental illness inevitably stigmatizing? In A. H. Crisp (Ed.). Every family in the land: Understanding prejudice and discrimination against people with mental illness. Robert Mond Memorial Trust. Retrieved January 8, 2008 from http://www.stigma.org/everyfamily/royporter.html.

Quine, S. (1985). Does the mode matter? A comparison of three modes of questionnaire completion. Community Health Studies, 9, 151-156.

Reginold, A. L., & Krishnan, S. (2002). The study of potentially stigmatizing conditions: an epidemiologic perspective. Paper presented at the International Conference on Stigma and Global Health: Developing a Research Agenda. Bethesda, MD. Retrieved January 8, 2008 from http://www.stigmaconference.nih.gov/ReingoldPaper.htm.

Rogler, L. H. (1999). Methodological sources of cultural insensitivity in mental health research. American Psychologist, 54, 424, 433.

Ruscello, D. M., Lass, N. J., Schmitt, J. F. & Pannbacker, M. D. (1994). Special educators' perceptions of stutterers. Journal of Fluency Disorders, 19, 125-132.

Sartorius, N., Jablensky, A., Korten, A., Ernberg, G., Anker, M., Cooper, J. E., & Day, R. (1986). Early manifestations and first-contact incidence of schizophrenia in different cultures: A preliminary report on the initial evaluation phase of the WHO collaborative study on determinants of outcome of severe mental disorders. Psychological Medicine, 16, 909-928.

Siegel, S. (1956). Nonparametric statistics for the behavioral sciences. NY: McGraw-Hill.

Smart, J. (2001). Disability, society, and the individual. Gaithersburg, MD: Aspen.

Smith, A. & Kelly, E. (1997). Stuttering: A dynamic, multifactorial model. In R. Curlee & G. Siegel (Eds.). Nature and Treatment of stuttering: New directions, 2nd ed. (pp. 204-217). Boston: Allyn & Bacon.

St. Louis, K. O. (2005). A global project to measure public attitudes of stuttering. The Asha Leader, 10, 12-13; 22.

St. Louis, K. O. (2001). Living with stuttering: Stories, basics, resources, and hope. Morgantown, WV: Populore.

St. Louis, K. O. (1999). Person-first labeling and stuttering. Journal of Fluency Disorders, 24, 1-24.

St. Louis, K. O., Schiffbauer, J. D., Phillips, C. I., Sedlock, A. B., Hriblan, l. J., & Dayton, R. M. (2000). The public environment where attitudes develop: Stuttering versus mental illness and intelligence. Paper presented at the 3rd International Stuttering Awareness Day On-Line Conference.

Thompson, G. (1999). The burden of mental illness in Alberta: An overview of research and the views of Albertans. Commissioned by the Alberta Provincial Mental Health Advisory Board for "Health Summit '99".

Thompson, A. H., Howard, A. W., & Jin, Y. (2001). A social problem index for Canada. Canadian Journal of Psychiatry, 46, 45-51.

Thompson, A. H., Stuart, H., Bland, R., Arboleda-Florez, J., Warner, R., & Dickson, R. A. (2002). Attitudes about schizophrenia from the pilot site of the WPA worldwide campaign against the stigma of schizophrenia. Social Psychiatry and Psychiatric Epidemiology, 37, 475-482.

US Census Bureau (2000). United States Department of Commerce. Retrieved January 8, 2008 from http://quickfacts.census.gov/qfd/states/54/54061.html.

US Surgeon General (2004). Mental health: A report of the Surgeon General. Retrieved September 1, 2004 from http://www.surgeongeneral.gov/library/mentalhealth/toc.html.

Ustun, T. B., Rehm, J., Chatterji, S., Saxena, S., Trotter, R., Room, R., Bickenbach, J., & the WHO/NIH Joint Project CAR Study Group. (1999). Multiple-informant ranking of the disabling effects of different health conditions in 14 countries. The Lancet, 354, 111-115.

Van Borsel, J., Verniers, I., & Bouvry, S. (1999). Public awareness of stuttering. Folia Phoniatrica and Logopaedica, 51, 124-132.

Wahl, O. F. (1999). Mental health consumers' experience of stigma. Schizophrenia Bulletin, 25, 467-478.

Weiss, M. G., & Ramakrishna, J. (2002). Intervention: Research on reducing stigma. Paper presented at the International Conference on Stigma and Global Health: Developing a Research Agenda. Bethesda, MD. Retrieved January 8, 2008 from http://www.stigmaconference.nih.gov/abstracts/WeissAbstract.htm.

Weiss, M. G., Jadhav, S., Raguram, R., Vounatsou, P., & Littlewood, R. (2001). Psychiatric stigma across cultures: Local validation in Bangalore and London. Anthropology and Medicine 8, 71-87.

World Health Organization (WHO). (2001). Epilepsy: Social consequences and economic aspects. Retrieved January 8, 2008 from http://www.who.int/mediacentre/factsheets/fs166/en/

World Values Study Group (1990-93). World Values Survey. Ann Arbor, MI: Inter-University Consortium for Political and Social Research (ICPSR).

Yaruss, J. S., & Quesal, R. W. (2004). Stuttering and the International Classification of Functioning, Disability, and Health (ICF): An update. Journal of Communication Disorders, 37, 35-52.

Zavecz, T. & Halasz, P. (2001). Changes in public attitudes toward epilepsy in Hungary: Results of survey conducted in 1994 and 2000. Epilepsia, 42, 86-93.

Kenneth O. St. Louis, Ph.D.

West Virginia University

Bobbie Boyd Lubker, Ph.D.

University of North Carolina, Chapel Hill

J. Scott , Ph.D.

University of Pittsburgh

Trisha A. Adkins, M.S.

Jackson County Board of Education, Ripley, West Virginia

Jaan Pill, B.Ed.

Canadian Stuttering Association, Toronto, Ontario

Table 1: Demographic data.

Number in sample	165
Percent return rate	55%
Age in years	31.5 yr
Percent males	29%
Percent females	71%
Total years in school	15.3 yr
Living Arrangement	
Percent single and never married	50%
Percent engaged	5%

Percent married	33%
Percent divorced or separated	5%
Percent other	6%
Working Arrangement	
Percent students	36%
Percent working for salary	64%
Percent working in own business	5%
Percent working at home	3%
Percent volunteering	3%
Percent unemployed	21%
Percent retired	4%
Race	
Percent Caucasian	93%
Percent other (African-American, Hispanic, Asian, Biracial, or Arab)	7%
Religion	
Percent Christian (Catholic, Protestant Orthodox)	85%
Percent Other (Jewish, Muslim Agnostic, Atheist, Other, or None listed)	15%
Percent English as native language	99%
Language Known Besides English	
Percent Spanish	10%
Percent French	6%
Percent German	5%
Percent Other (Russian, Italian, Latin, Chinese, Greek, Slavic or Arabic)	5%
Job satisfaction rating (0-100 / 1-9)	74.90
Physical health rating (0-100 / 1-9)	68.80
Mental health rating (0-100 / 1-9)	74.60
Ability to learn new things rating (0-100 / 1-9)	79.50
Speaking ability rating (0-100 / 1-9)	75.80
Percent with stuttering disorder	3%
Percent with men illness	5%
Percentusing a wheelchair	2%
Percent overweight	26%
Percent old	4%
Percent left-handed	12%
Percent multilingual	10%
Percent good talkers	30%
Percent intelligent	48%

Table 2: Respondent Distributions in the Survey Nonprobability Sample Compared with Census Distributions in the Host and Other WV Counties on Variables of Race, Sex, and Education

		Race		Sex	
Country & State	% Sample from County	White	Other races (n=165)	Female	Male
--	--	93% (b)	7% (b)	71% (b)	29% (b)
		US Census Comparison Data			
Monongalia,					

Monongalia, WV (c)	27% (d)	92% (a)	8% (a)	50% (a)	50% (a)
Brooke, WV	9% (d)	98% (a)	2% (a)	52% (a)	48% (a)
Tyler, WV	7% (d)	99% (a)	1% (a)	51% (a)	49% (a)
Ohio, WV	4% (d)	95% (a)	5% (a)	53% (a)	47% (a)
Mineral, WV	3% (d)	96% (a)	4% (a)	51% (a)	49% (a)
Jefferson, WV	3% (d)	91% (a)	9% (a)	51% (a)	49% (a)
Wetzel, WV	3% (d)	99% (a)	1% (a)	52% (a)	48% (a)
Other WV Counties	12% (d)	--	--	--	--
State of WV	--	95% (a)	5% (a)	51% (a)	49% (a)

Education

Country & State	% with Bachelor's Degree & >25 years old (a)	Mean Years of School (b)	% with [greater than or equal to]16 years School & [greater than or equal to]25 old (b)
Monongalia, WV (c)	--	15.3 (b)	22% (b)
Brooke, WV	32% (a)	--	--
Tyler, WV	13% (a)	--	--
Ohio, WV	9% (a)	--	--
Mineral, WV	23% (a)	--	--
Jefferson, WV	12% (a)	--	--
Wetzel, WV	22% (a)	--	--
Other WV Counties	10% (a)	--	--
State of WV	15% (a)	--	--

(a) Data from the US Census (2000).

(b) Data from demographic section of POSHA-E.

(c) Host county.

(d) Percentage of respondents from county.

Table 3: Overall Means for Three Different Item Orders of Stuttering in the General Section and for Three Different Triad Orders of the Detailed Sections for Stuttering.

Item Order of Stuttering Questions in the General Section

	First (1st of 9)	Middle (5st of 9)	Last (9th of 9)
Mean for 3 General Questions on All Attributes	54.10	32.30	36.60

Mean for all Detailed Stuttering Questions	42.90	43.00	43.7

Triad Order of Stuttering Section Among Three Detailed Sections

	First	Second	Third
Mean for all Detailed Stuttering Questions	43.30	43.90	42.30

Table 4: Overall Frequencies of Identifying Personal Experience with Nine Attributes According to Three Different Item Orders of Stuttering in the General Section.

Order of Stuttering Questions in the General Section

Attribute	First (1st of 9)	Middle (5th of 9) (a)	Last (9th of 9) (b)	Total/3
Mental Illness	58.7	92.0	77.0	72.7
Stuttering	43.8	92.0	69.0	66.0
Wheelchair	44.9	101.0	64.0	67.0
Overweight	93.3	209.0	126.0	137.3
Old	102.8	184.0	131.0	133.3
Left-handed	70.1	146.0	91.0	98.0
Multilingual	63.0	108.0	85.0	82.3
Good Talking	94.4	163.0	126.0	123.0
Intelligent	118.1	251.0	175.0	173.7

(a) Data corrected to 1/3; see text.

(b) Already exactly 1l3 of the respondents.

Gale Document Number:A180592546