CROSS-CULTURAL ANALYSIS OF SELECTED EMOTIONAL QUESTIONS FROM THE CORNELL MEDICAL INDEX

ESKO KALIMO* AND THOMAS W. BICE

Department of Medical Care and Hospitals and Department of Behavioral Sciences, The Johns Hopkins University, Baltimore, Maryland 21205, U.S.A.

AND

MARIJA NOVOSEL

Institute of Public Health of Croatia, Zagreb, Yugoslavia

When considering the problems of the cross-cultural applicability of mental health questionnaires, the influence of cultural factors on the results is of central importance. Zola (1966) has suggested that there are at least two ways in which symptoms usually defined as indicating illness in one population may be ignored in others. First, the estimate of the prevalence of a condition may be misleading, for if the corresponding symptoms are common within the population, they might not be considered symptomatic of any illnesses. Secondly, the form taken by the illness may depend on the dominant value-orientations of the culture. Hunt (1959) has concluded that 'studies provide evidence of extensive cultural variability in the expression of mental disorder'. The cultural relativity and dependence on the values of the society may greatly affect the measurement and definition of mental illness in different societies (Wittkower and Fried, 1958; Opler, 1959; Håkanson, 1968). These selective contextual processes, and not aetiological ones, account for many epidemiological differences in the prevalence of mental disorders between societies and subgroups within a society. This has led Zubin and Kietzman (1966) to distinguish culture-dependent and culture-free aspects of mental disorders. The cultural influences can be seen clearly in studies in which ethnic origin has been found to be highly related to the perception and interpretation of symptoms, to the corresponding reaction, and to the content of disorder (Zborowski, 1952 and 1969; Opler and Singer, 1956; Zola, 1966). The cultural dependence of mental disorders restricts the applicability of universal measurement methods in many cultures. The problem of validity in comparing field studies of mental disorder has proved to be difficult (Dohrenwend and Dohrenwend, 1965).

The Cornell Medical Index (Brodman, Erdmann, Lorge, and Wolff, 1949; Brodman, Erdmann, and Wolff, 1949), abbreviated CMI, is a well-known health questionnaire developed originally in the United States. It can be used for both clinical and research purposes to give fairly comprehensive information about the medical problems of the respondent as perceived by himself. CMI contains 195 questions which can be answered by either 'yes' or 'no'. On the basis of their content, the questions have been divided into 18 sections, six of which (51 questions) refer to mood and feeling patterns. The other questions relate to bodily symptoms, past illnesses, or family history. No detailed rules are given on how to interpret the results.

The total number of 'yes' responses to CMI as a whole or to questions relating to mood and feeling patterns has been found to be a valid indicator of the presence of emotional disorders (Brodman, Erdmann, Lorge, Gershenson, and Wolff, 1952 a and b; Brodman, Erdmann, Lorge, Deutschberger, and Wolff, 1954; Lawton, 1959; Culpan, Davies, and Oppenheim, 1960). With CMI many emotional disorders have been identified which were overlooked in hospital investigation (Brodman, Erdmann, Lorge, and Wolff, 1951). A cutting point of 30 positive responses to the questionnaire is commonly used to indicate a probable serious emotional disorder (Brodman et al., 1949; Abramson, Terespolsky, Broek and Kark, 1965; Abramson, 1966). The sensitivity of this high score may be relatively low, as some emotionally disturbed respondents give only a few, although clinically significant 'yes' responses (Brodman et al., 1952a). The applicability of information about the bodily symptoms to measuring emotional disorders can be seen in a study in which a shortened scale, derived from the physical sections of CMI, was used as a measure of neuroticism (Gibson, Hanson, and West, 1967). An obvious restriction on
the applicability of CMI and other measures derived from it is the fact that most of the responses are self-ratings of the respondent and they therefore reflect perceived ill-health rather than objective ill-health (Suchman, Phillips, and Streib, 1958). Depending on the purpose, this kind of health information may or may not be valid. The validity of CMI may also be lowered by the fact that all the questions are worded in the same direction. The scores based on the 'yes' responses may be contaminated by acquiescence, i.e., the tendency to answer 'yes' irrespective of the content of the question (Fisher, 1967). Acquiescence may be conceptualized as a continuous variable, whose one end consists of the 'yeasayers' and the other end the 'naysayers' (Couch and Keniston, 1960). There is also evidence that responses to mental health questionnaires like CMI may be influenced by another response bias, social desirability, i.e., the tendency to admit to good rather than bad things about oneself (Phillips and Clancy, 1970).

The problems of the cross-national measurement in general centre around the equivalence of the meaning of the measuring instrument in different languages and in different cultures. This problem may be further conceptualized by assessing the equivalence of the instrument as a measure of a concept on three cumulative levels: the technological, the semantic, and the conceptual (Bice and Kalimo, 1970). A measure is accepted as technologically equivalent if formally similar standardized methods, e.g., not standardized methods can be applied in all study populations. A measure is semantically equivalent if it (usually its questions) means the same thing in all study populations. If questions are needed in different languages, the translator usually is instructed to translate the meaning of the question rather than to translate to maintain strict word by word correspondence (Casagrande, 1954). Rather complicated procedures, including back translations, are often necessary to maintain the semantic equivalence of the stimuli (Anderson, 1967). Semantically equivalent questions need not be also conceptually equivalent. The same manifest response to a semantically equivalent question may have different meanings in different societies. The questions can be accepted as conceptually equivalent only if the responses can be shown to be indicators of the same concept in different study populations. On the other hand, conceptually equivalent questions need not be semantically equivalent (Przeworski and Teune, 1966).

The transfer of verbal diagnostic measures from one culture to another can be examined by applying the preceding concepts (see also Ortar, 1963; Anastasi, 1968). Since most diagnostic methods are supposed to be measures of specific concepts, a diagnostic method should preferably be a conceptually equivalent measure, i.e., its construct validity should be the same in all study populations. When a diagnostic method is applied in different social contexts, it is important that the degree of the conceptual equivalence of the measure be assessed. If the method is not found conceptually equivalent, a new method may have to be developed for the purpose. This may lead to the use of different diagnostic methods as measures of the same concept in different social contexts. The evaluation of the scores of diagnostic methods is done usually by reference to norms based on a sample of a population. Norms may be culturally specific, and often they pertain only to subgroups within a nation. The necessity of the norms creates a problem specific to diagnostic methods in discussing problems of cross-national measurement. Although a method is found conceptually equivalent in different cultural settings and in different languages, it is likely that the same norms cannot be applied in all populations. Diagnostic methods which can be used in different cultures with different norms are called culturally ipsativized methods by Straus (1969). Correspondingly, he called conceptually equivalent but semantically different methods, for which different norms are used for different populations, culturally specific methods. Only in rare instances is it possible to apply a same diagnostic method with the same norms, i.e., a culturally universal method, in different cultures. When applying a diagnostic method, a decision is usually made about the relative position of an individual on a latent continuum. In this case, a much greater conceptual equivalence in addition to group-specific norms is needed than would be the case if inferences were made only about group differences by the same method in different societies. Manis, Brawer, Hunt, and Kercher (1963) have argued that a mental health scale similar to CMI was more valid as a measure of group differences than as a measure of the relative positions of individuals. It can be concluded that very good conceptual equivalence is necessary so that a measure of mental disorders would be applicable in many cultures as a comparable diagnostic method for individual cases. The cultural dependence of the mental disorders probably hinders the development of culturally universal and even culturally ipsativized methods for measuring mental disorder.

The applicability of CMI has been examined in
among a group where CMI was developed. Chance (1962) has reported results about the applicability of CMI among a group of Alaskan Eskimos. He found that 'in the predominantly psychiatric section of the questionnaire, Eskimo cultural attitudes directly influenced responses to questions dealing with anxiety symptoms and the decision-making process'. In these areas there was clearly need for culturally specific questions. Conceptual equivalence was greater in questions dealing with physical symptoms, the questions being culturally ipsative. When the applicability of CMI was examined in a Zulu population, two types of bias were found (Scotch and Geiger, 1963–64). First, the validity of some questions was different, since the meanings of many symptoms were not the same for the Zulu samples as for the North Americans. Secondly, environmental inappropriateness was noticed because of differences in objective conditions, e.g., the number of doctors and the nature of common and clinically meaningful symptoms. It was concluded that CMI had some validity as a measure of health profile, although the average number of 'yes' responses among the Zulu samples was much higher than in the United States' samples. CMI was thus a culturally ipsatized measure in the Zulu culture. CMI scores also have been shown to be related to ethnic origin in a group of United States Army inductees, thus homogeneous in age, sex, and language (Croog, 1961). Respondents of Italian and Jewish origin reported significantly more 'yes' responses than respondents of Irish, British, and German origins. This finding probably reflects culture-dependent differences in the perception and expression of symptoms.

Laufer (1953) has reported results about the use of the Cornell Index, a measure of emotional ill-health closely similar to CMI, among Okinawan natives. The construct validity of the questions was found to be quite different, since 'yes' responses indicated courtesy rather than neurosis. Many questions were not meaningful and were not understood. The Cornell Index was thus unusable among the Okinawans, its conceptual equivalence being very low when compared to that of the North Americans. On the other hand, the Maudsley Personality Inventory, originally developed in Britain as a measure of neuroticism and extraversion-introversion, has turned out to be fairly valid for these purposes in Chile (Bolardos, 1964), India (Jaloba, 1964), and the Lebanon (Rafi, 1965). Even the means of the scales in the samples were not very different from the means of the standardization samples in Britain. Kline (1967) found the Cattell 16 P.F. Personality Questionnaire and the Eysenck Personality Inventory to be valid measures among literate, educated Ghanaians. These studies have focused mainly on the construct validity of the measure without paying much attention to the need for using different norms in different societies. Research using the Italian version of the MMPI, a questionnaire measure of mental disorders, showed that according to the American norms the scores of the Italian samples indicated a considerably greater tendency to psychopathology (Rosen and Rizzo, 1961). They also found that the inventory scores vary greatly among geographical areas in Italy, presumably due to material and cultural correlates of place of residence. The results pointed out the need for different norms, i.e., separate standardization of the inventory in each country and probably in different populations within the same country, since the same score had different meanings in different forms and different populations. Glatt (1969) evaluated the French, German, and Spanish translations of the MMPI by using bilingual subjects, each of whom completed the English and one other form. The German and Spanish forms were adequate and clinically useful when compared to the English form. The French form was not conceptually equivalent, probably due to deficiencies in translation. Also, the means of many scales were different in the French and English forms.

These results seem to support the conclusion that culturally similar nations are likely to have similar forms of mental disorder, and in culturally similar nations it may be feasible to apply culturally ipsatized mental health questionnaires or even the same questionnaires with the same norms, i.e. (in these cultures), universal methods. The less similar the cultures, the greater is the probability that culturally specific measures must be applied to take into account possible differences in the forms of mental disorder. The results also suggest that even within the same country, different norms may be necessary for different subgroups so that the diagnostic applicability of the method could be maintained.

The conceptual equivalence of a mental health questionnaire should preferably be measured against a criterion in each study population. If the cultural relativity of mental disorders is accepted, this criterion can vary from one population to another. There are also other, less satisfactory, techniques for measuring the degree of conceptual equivalence. One can, e.g., examine the correlations among the items of the questionnaire in each study population or the relationships of the questionnaire with other variables in each study population (Cronbach and
Meehl, 1955). If these relationships turn out to be similar in each population, the questionnaire is probably a conceptually equivalent measure in these populations. However, these latter methods do not take into account the cultural relativity of the measured phenomenon.

As the total number of ‘yes’ responses to CMI has validity in measuring morbidity in general, it may be assumed to correlate positively with age. This kind of correlation has been found (Brodman, Erdmann, Lorge, and Wolff, 1953), but the correlation was caused only by a difference in reporting bodily complaints among the age groups. Older men and women made more bodily complaints than younger, but a like number of complaints of mood and feelings. In the same study, women made more complaints than men although both had the same average number of diagnosed disorders. Many other studies in the United States and Britain (Herbolzheimer and Ballard, 1958; Culpan et al., 1960; Brown and Fry, 1962; Ryle and Hamilton, 1962) have shown that women get higher scores than men on CMI. Abramson (1966) has shown that the sex differences are greater among ostensibly healthy persons and non-psychiatric patients, since emotionally disturbed men and women tend to differ only little in the proportion with a score of 10 or above on the CMI questions measuring mood and feeling patterns. Hence the relatively marked sex differences found in non-selected populations may indicate a real sex difference in emotional health rather than suggest that women give more ‘yes’ responses because of cultural influences. Nevertheless, the results of Phillips and Segal (1969) provided strong support for the hypothesis that women report more psychiatric symptoms than men with an equal number of physical illnesses. The result suggests a need for different critical scoring levels for men and women on CMI.

**Problem**

The present study stems from the measurement problems in an international study, in which a measure of emotional disturbances was used in study populations with five different languages. The purposes of this paper are to examine the conceptual equivalence and the applicability to the same evaluative norms of a set of emotional questions of CMI in study areas located in seven industrialized countries.

The conceptual equivalence of the questions will be measured by examining the relationships among the questions within each study area. Also, an index of the questions will be compared with age, sex, and perceived morbidity. The marginal distributions of the questions and the index will be compared among the study areas. The role of a different language and country in the data will be analysed by comparing the results among study areas grouped by language and by country.

This study tests the following measurement hypotheses:

1. The set of questions is a conceptually equivalent measure of emotional disturbances in all study areas.
2. The same evaluative norms are not applicable to the set of questions in all study areas.
3. The same evaluative norms are more applicable in study populations with the same language and within the same country than in populations with a different language and in different countries.
4. The index of the questions is positively related to age, female sex, and perceived morbidity in all study areas.

**Method**

**Questions**

For the purposes of the larger study of which this analysis is a part, a brief measure of emotional disturbances was needed. Ten questions were selected from CMI on the basis of their content and estimated 'applicability in cross-national research'. The questions represented various aspects of emotional disturbances and were chosen from seven different sections of CMI. Half of the questions belong to sections M-R, which include questions of mood and feeling patterns. The questions are given below:

**Question**

1. Are you often bothered by thumping of your heart?
2. Does your heart often race like mad?
3. Does pressure or pain in the head often make life miserable?
4. Do you often shake or tremble?
5. Do you often get spells of complete exhaustion or fatigue?
6. Do you wear yourself out worrying about your health?
7. Does your thinking get completely mixed up when you have to do things quickly?
8. Are you constantly keyed up and jittery?
9. Do frightening thoughts keep coming back in your mind?
10. Do you usually feel unhappy or depressed?

**Section of CMI**

C (Cardiovascular system)
C (Cardiovascular system)
G (Nervous system)
R (Tension)
I (Fatigability)
J (Frequency of illness)
M (Inadequacy)
R (Tension)
R (Tension)
N (Depression)

*The assistance of Dr. Constantino Chusqui of Chile in selecting the questions is gratefully appreciated.*
CROSS-CULTURAL ANALYSIS OF QUESTIONS FROM CORNELL MEDICAL INDEX 233

All these questions are found in identical or closely similar form also in the Cornell Index (Weider, Wolff, Brodman, Mittelmann, and Wechsler, 1949), a measure of emotional disturbances. The questions can be divided into two groups on the basis of the content: those concerning bodily symptoms (questions 1, 2, 3, 4, and 5) and those differentiating between persons with serious personality disturbances and persons without such disturbances (questions 6, 7, 8, 9, and 10).

The questions were translated from English to Finnish, Polish, Serbo-Croatian, and Spanish. The translators were instructed to maintain the semantic equivalence of the questions. A procedure suggested by Schachter (1954) was used to evaluate the translations. Two translators translated the questions independently back into English. These were compared with the original English questions.

Questions included three response alternatives since, in addition to the 'yes' and 'no' responses suggested by the CMI Manual, 'don't know' responses were allowed. Questions were dichotomized for the analysis by incorporating 'don't know' responses in the 'no' responses. The percentages of 'don't know' responses were negligible, the median percentage over all questions and study areas being 0% and the highest percentage for any question in any study area 4%. An index of emotional disturbances was created by counting the numbers of 'yes' responses to the questions.

Two measures of perceived morbidity were used, a measure of disability days and an indicator of chronic health problem. The disability days variable is defined as the number of days within the two weeks before the interview that the respondent either stayed in bed all or part of any day or restricted his usual activities because of ill health. The chronic health problem index is based on a series of questions concerning the presence and effects of impairments and chronic illnesses. Three categories are defined: (1) persons who reported no impairments and no chronic illnesses, (2) those with either an impairment or chronic illness, neither of which causes permanent disability, and (3) persons who are disabled by either an impairment or chronic illness.

Samples

The data were gathered by means of household surveys in 12 areas located in seven countries. The areas are as follows: Canada (Grande Prairie, Alberta; Saskatchewan; Fraser, British Columbia; Jersey, British Columbia), the United Kingdom (Liverpool), Argentina (Buenos Aires), Finland (Helsinki), Poland (Lodz), and Yugoslavia (Banat and Rijeka). The populations of these areas include, in most cases, rural and urban segments. More detailed information about the populations and study areas is given elsewhere (WHO/ICS-MCU, 1970d).

Within each region four independent probability samples of households of non-institutionalized, civilian persons were selected. Since the field work was planned to extend over the 12-month period from 1 June 1968 to 31 May 1969, the subsamples were selected quarterly, each comprised of households to be interviewed within a three-month period. In the selected households all permanent, civilian residents were eligible to be interviewed. Those 15 years of age and older responded to questions in an adults' questionnaire, and parallel forms were completed for all children in interviews with, in most cases, one of their parents. Sampling plans are described in more detail in a manual (WHO/ICS-MCU, 1970a).

Most samples being self-weighting, stratified sampling plans with unequal sampling fractions were used in two areas, Saskatchewan and Lodz. Although these samples are not self-weighting, in the analyses reported in this paper all samples were considered additive, and strata are not weighted. The resulting error in the statistics is probably inconsequential.

Throughout the planning and execution of the study standardized procedures intended to achieve maximum comparability were employed. Questionnaires were developed through collaboration and pretesting involving participants from all study areas; interviewers' training and supervisory procedures as well as coding instructions were standardized and made available to field work directors (WHO/ICS-MCU, 1970b, c, e).

The data reported in this paper were collected in interviews of adults selected in the first quarterly subsample, i.e., over the period from June 1, 1968, to August 31, 1968. The number of respondents in each area is as follows: Grande Prairie (716), Saskatchewan (624), Fraser (569), Jersey (539), Northwestern Vermont (522), Baltimore (559), Liverpool (433), Buenos Aires (725), Helsinki (941), Lodz (865), Banat (1,023), and Rijeka (515). Response rates were generally high, ranging from 90 to 99% over all areas.

Statistical Techniques

Each of the analytical objectives of this paper is approached via a different analytical technique. Our measure of conceptual equivalence of the 10

---

*Fraser and Jersey are separate, non-contiguous school districts located in the Lower Fraser Valley, east of metropolitan Vancouver.*
questions across the study areas is based on comparisons of the correlation structures among the questions. Therefore, the basic data for these comparisons are matrices which include each area's all pairwise associations among the 10 questions. From these matrices the ranges and median associations are compared as well as the rank orders of the associations. Because we wish to compare average levels of association among the areas, it is necessary that the value of the statistics used in the matrices should not depend on the questions' marginal distributions. Goodman and Kruskal's (1964) gamma, a measure of ordinal association, was selected because it satisfies this requirement. In the fourfold case, gamma reduces to Yule's Q.

In order to test the appropriateness of the same evaluative norms in all areas, the mean scores of the index of emotional disturbance are compared among the areas. Categorizing each area in terms of the primary language spoken and country, the Mann-Whitney U test (Siegel, 1956) is applied to test the null hypothesis that the distribution of mean scores does not differ between language groups and countries. In applying this test, we assume that it is unlikely that average scores would vary systematically with these variables unless the responses to questions were influenced by cultural factors. Because the 12 study areas do not constitute a probability sample of a defined population of areas, the probability statement associated with the Mann-Whitney U test is not interpreted from the normal sampling error perspective. Rather, the probabilities refer to the chance of observing a difference between means of ranks as large as or larger than the observed difference when all haphazard combinations of areas are considered. For example, with seven English-speaking and five non-English-speaking areas, there are \( \binom{12}{2} = 792 \) haphazard samples in which the 12 areas are divided into groups of seven and five. Associated with each of these samples is a difference between means of the ranks of mean scores which gives a distribution of differences. One of these haphazard samples is the one actually observed. The observed difference is compared to the distribution of differences to determine how frequently it would be expected to occur by chance (Winch and Campbell, 1969).

Finally, product-moment correlations are computed between the index scores and age, sex, and perceived morbidity.

**RESULTS**

The first hypothesis of this study is that the 10 selected questions from CMI are conceptually equivalent indicators of emotional disturbances in the 12 study areas. The data pertaining to this hypothesis are given in Table I, which contains the ranges of the gammas among the questions for each area. It can be seen that all associations in all matrices are positively signed. This result indicates that in each of the areas the 10 questions may have tapped a single underlying dimension.

As the questions were divided into two groups on the basis of their substantive content, average levels

*The full gamma matrices for all study areas are available from Dr. Bice.

**TABLE I**

<table>
<thead>
<tr>
<th>Study Area</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grande Prairie</td>
<td>37-94</td>
</tr>
<tr>
<td>Saskatchewan</td>
<td>25-81</td>
</tr>
<tr>
<td>Fraser</td>
<td>49-94</td>
</tr>
<tr>
<td>Jersey</td>
<td>13-93</td>
</tr>
<tr>
<td>N. W. Vermont</td>
<td>17-98</td>
</tr>
<tr>
<td>Baltimore</td>
<td>43-90</td>
</tr>
<tr>
<td>Liverpool</td>
<td>20-97</td>
</tr>
<tr>
<td>Buenos Aires</td>
<td>29-90</td>
</tr>
<tr>
<td>Helsinki</td>
<td>58-96</td>
</tr>
<tr>
<td>Lodz</td>
<td>45-96</td>
</tr>
<tr>
<td>Banat</td>
<td>27-96</td>
</tr>
<tr>
<td>Rijeka</td>
<td>20-96</td>
</tr>
</tbody>
</table>

**TABLE II**

<table>
<thead>
<tr>
<th>Study Area</th>
<th>All Questions</th>
<th>Bodily Symptoms</th>
<th>Personality Disturbances</th>
<th>Body Symptoms</th>
<th>Personality Disturbances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grande Prairie</td>
<td>68</td>
<td>57</td>
<td>78</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>Saskatchewan</td>
<td>68</td>
<td>62</td>
<td>77</td>
<td>62</td>
<td>66</td>
</tr>
<tr>
<td>Fraser</td>
<td>69</td>
<td>68</td>
<td>77</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>Jersey</td>
<td>59</td>
<td>56</td>
<td>72</td>
<td>54</td>
<td>54</td>
</tr>
<tr>
<td>N. W. Vermont</td>
<td>67</td>
<td>52</td>
<td>78</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>Baltimore</td>
<td>66</td>
<td>67</td>
<td>62</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>Liverpool</td>
<td>71</td>
<td>71</td>
<td>74</td>
<td>67</td>
<td>67</td>
</tr>
<tr>
<td>Buenos Aires</td>
<td>63</td>
<td>66</td>
<td>62</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>Helsinki</td>
<td>72</td>
<td>72</td>
<td>80</td>
<td>71</td>
<td>71</td>
</tr>
<tr>
<td>Lodz</td>
<td>68</td>
<td>68</td>
<td>63</td>
<td>63</td>
<td>63</td>
</tr>
<tr>
<td>Banat</td>
<td>69</td>
<td>72</td>
<td>63</td>
<td>68</td>
<td>68</td>
</tr>
<tr>
<td>Rijeka</td>
<td>64</td>
<td>70</td>
<td>58</td>
<td>63</td>
<td>63</td>
</tr>
</tbody>
</table>
of association were computed for each total matrix and for the two subsets as well as for the associations among questions in different subsets. The median gammas in Table II show that, in general, the subgroup medians differ little from the total medians, and there are no consistent differences between the sub-groups. Thus there are no empirical grounds for considering bodily symptom and personality disturbance questions separately. To test the dimensionality of the gammas, principal component analyses of the matrices of two areas were carried out. In the Baltimore data the first eigenvalue accounted for 80% of the sum of all positive eigenvalues, and in the Helsinki data the corresponding proportion was 83%. Because the average associations among the questions were high and the matrices were so similar, principal component analyses were not carried out for other areas. These results clearly indicate that the questions measure a single dimension.

The medians for the total gammas vary from 0.63 in Buenos Aires to 0.72 in Helsinki, which suggests that the average levels of association among the items are fairly similar across the areas. However, the patterns of associations vary considerably among the areas and are related to the primary language spoken and the nation in which the area is located. Within each area the gammas were rank ordered, and coefficients of concordance were calculated across all areas, across areas in which English is the major language and all others, and across the Canadian, United States, Yugoslav, and other areas. The coefficient of concordance across all areas is a rather low 0.33. Across the seven English-speaking study areas its value is 0.50, and across the non-English-speaking areas the coefficient of concordance is 0.51. Hence, there is greater similarity in the patterns of associations among the English-speaking areas, in which the same, untranslated set of questions was used, and in the non-English-speaking areas, in which translated versions of the questions were employed, than among all the areas taken together. Similarly, the patterns of associations are more similar within nations than for the set of areas as a whole. The coefficients of concordance for the rank orders of gammas is 0.55 across the four Canadian areas, 0.72 across the two United States areas, 0.91 across the two Yugoslav areas, and 0.41 across the other four areas.

The findings that the 10 questions are consistently positively correlated in all study areas and that the empirical associations do not warrant the questions being divided into subsets support the interpretation that all 10 questions measure a single latent dimension within each area. However, differences among patterns of associations among the questions suggest that differences in the relationships between individual questions and the latent dimension exist among areas which used the same untranslated English questions and those which employed translations and also among areas located in different countries.

The data in Table III show that few individuals gave affirmative replies to the questions in all areas, but the median percentages of 'yes' responses and patterns of 'yes' responses to questions vary considerably among the study areas. Although only one of the 120 percentages in Table III exceeds 50%, the range of median 'yes' responses across the study areas is from 6 in Helsinki to 23 in Banat. Within each area the questions were rank ordered on the basis of the proportion of 'yes' responses, and coefficients of concordance were calculated across all areas, across English-speaking and non-English-speaking areas, and across areas within the same countries. Over all the areas the coefficient of concordance for rank orders is 0.55. Across English-speaking areas and non-English-speaking areas the values are, respectively, 0.85 and 0.62. Across the Canadian, United States, and Yugoslav study areas the corresponding values are 0.90, 0.94, and 0.97; for the remaining four areas the coefficient of

<table>
<thead>
<tr>
<th>Item</th>
<th>Grande Prairie</th>
<th>Saskatchewan</th>
<th>Fraser</th>
<th>Jersey</th>
<th>N.W. Vermont</th>
<th>Baltimore</th>
<th>Liverpool</th>
<th>Buenos Aires</th>
<th>Helsinki</th>
<th>Lodz</th>
<th>Banat</th>
<th>Rijeka</th>
<th>Median % 'Yes'</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>6</td>
<td>17</td>
<td>6</td>
<td>5</td>
<td>11</td>
<td>10</td>
<td>14</td>
<td>5</td>
<td>6</td>
<td>10</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>9</td>
<td>16</td>
<td>7</td>
<td>5</td>
<td>14</td>
<td>8</td>
<td>14</td>
<td>10</td>
<td>8</td>
<td>15</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>12</td>
<td>18</td>
<td>6</td>
<td>10</td>
<td>9</td>
<td>17</td>
<td>14</td>
<td>12</td>
<td>12</td>
<td>18</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>12</td>
<td>9</td>
<td>9</td>
<td>12</td>
<td>7</td>
<td>17</td>
<td>17</td>
<td>12</td>
<td>13</td>
<td>18</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>12</td>
<td>8</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td>17</td>
<td>19</td>
<td>17</td>
<td>14</td>
<td>13</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>13</td>
<td>13</td>
<td>7</td>
<td>13</td>
<td>9</td>
<td>19</td>
<td>19</td>
<td>20</td>
<td>20</td>
<td>14</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>13</td>
<td>13</td>
<td>14</td>
<td>11</td>
<td>11</td>
<td>13</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>14</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>13</td>
<td>14</td>
<td>11</td>
<td>10</td>
<td>13</td>
<td>20</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>14</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>14</td>
<td>15</td>
<td>12</td>
<td>10</td>
<td>13</td>
<td>20</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>14</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>19</td>
<td>17</td>
<td>15</td>
<td>11</td>
<td>11</td>
<td>13</td>
<td>20</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>14</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>Median</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>22</td>
<td>10</td>
<td>7</td>
</tr>
</tbody>
</table>

TABLE III

PER CENT. 'YES' RESPONSES, BY QUESTION AND STUDY AREA
concordance is 0.44. As was the case with the rank orders of gammas, the patterns of ‘yes’ responses to questions were more similar among English-speaking areas and among areas located in the same countries than among all areas taken together.

An index of emotional disturbance was constructed for each respondent by counting his number of ‘yes’ responses. Distributions of the index scores, means, and measures of the index’s internal consistency based on Kuder-Richardson formula 20 (Gulliksen, 1950) appear in Table IV.

The second hypothesis of this study states that the same evaluative norms cannot be applied to all study areas. Assuming that the prevalences of emotional disturbance are equal across areas, one would expect to find similarities among the average index scores, and, if the prevalences in fact differ, then average scores should not be systematically related to whether or not translations of the questions are employed. The data in Table IV indicate that the distributions of index scores differ among the study areas: the mean scores range from 0.7 in Helsinki to 2.8 in Rijeka.

The areas’ mean index scores are rank-ordered and the areas are classified as English-speaking (E) and non-English-speaking (N) in the following series:

\[
\text{Lowest Mean} \quad \text{Highest Mean} \\
\text{N} \quad \text{N} \quad \text{E} \quad \text{E} \quad \text{E} \quad \text{E} \quad \text{E} \\
\text{N} \quad \text{N} \\
\]

Applying the Mann-Whitney U test, we find that the probability of getting by chance the observed ordering or one more extreme is 0.05. As language is similar only among the Canadian and United States areas, the nation effect on the mean scores controlled for language can be examined only among these six areas. Applying the Mann-Whitney U test to the ordering of Canadian (C) and United States (U) areas, all of which employed untranslated, English questions, shows that the probability of getting the following series by chance is 0.27. The corresponding probabilities for the English-speaking areas categorized as United States’ areas and others including Liverpool are 0.43 and as Canadian areas and others including Liverpool 0.11. These results indicate that the tendency to respond affirmatively to CMI questions is highly associated among the 12 study areas with whether or not translations were used but less so with the countries in which the English-speaking areas are located. While the differences among mean scores may reflect true variations in the prevalences of emotional disturbance, a more plausible interpretation of these results is that response patterns to these CMI questions are influenced by cultural factors. If this is the case, it is inappropriate to apply the same evaluative norms in different cultures to distributions of emotional disturbance indices based on CMI.

The diagnostic cutting point of 30 for CMI is approximately 15% of the total 195 questions. If we assume that the 10 questions we used constitute a random sample of all CMI questions, which of course they are not, then the 15% cutting point is a score of approximately two. The mean scores of three areas, Lodz, Banat, and Rijeka, all non-English-speaking, exceed this cutting point. Rank ordering the 12 areas in terms of the proportions of individuals above the cutting point of two and classifying them as English-speaking and non-English-speaking, we observe a series identical to that observed when mean index scores were ranked. Hence we conclude that the distributions of proportions of individuals scoring above the cutting point differ among language groups.

An effort was made to examine the influence of the acquiescence tendency on the responses. If respondents were to answer ‘yes’ and ‘no’ in runs, then we might expect a negative relationship between the correlation between pairs of items and their distances in the questionnaire, i.e., responses to adjacent questions would be more highly correlated than would be those separated by two or more questions. However, the product-moment correlations between question distances and gammas shown in Table V are generally low and evenly divided between negatively and positively signed correlations. This evidence, while not conclusive, suggests that differences in tendencies to yea- and nay-say
do not account for the between-area variations in response patterns.

The final objective of the present study is to relate the emotional disturbance index scores to external criterion variables which have been shown in other studies to be associated with CMI scores. Table VI presents product-moment correlations between the emotional disturbance index scores and age, sex, disability days experienced within a two-week period, and the presence or absence of a chronic health problem. The data indicate that the index score is positively associated with sex, disability days, and chronic health problem within each of the study areas. Age is variously related to the index across the areas. Within each of the areas females and persons suffering chronic and acute health problems are more likely to have high index scores than are males and those who claim to be less ill. Judging from the across-area median correlations among these criterion variables presented in Table VII, it appears that females' higher index scores are not due to greater prevalences of perceived illness among females. Overall there is a very weak relationship between sex and perceived morbidity and a stronger relationship between the two perceived morbidity measures. Hence we may conclude that sex and perceived morbidity contribute to the observed index scores.

**DISCUSSION**

The results of these analyses support the hypotheses of the study. The set of 10 CMI questions appears to measure a single dimension in each of the 12 study areas, and the average scores vary among study areas with whether or not translations were employed. Finally, the index score is related within each area to the external criterion variables, sex and perceived morbidity. There is no consistent relationship across the areas with age, a finding which agrees with other results based only on the CMI mood and feeling questions (Brodman et al., 1953). Based on our limited analyses, we conclude that the CMI questions used in this study are identical indicators of emotional disturbance in the 12 study areas: they are both semantically and conceptually equivalent questions. However, the conventional evaluative norms are probably not appropriate in all study areas. The index is thus a culturally ipsatized evaluative method in these areas.

This situation is rather common when subjective variables are measured in cross-cultural research. Although the indicators tap the same dimension their locations on the dimension vary among populations. In effect, cultural factors define thresholds resulting in different proportions of endorsements and overall average scores. One such factor could be due to variations in the perceived social desirability of the CMI questions across the study area. In this study, for example, respondents in Buenos Aires, Lodz, Banat, and Rijeka gave more 'yes' responses, regardless of the content of questions,
than respondents in other areas. Hence we conclude that individuals within each of the areas are rank ordered with respect to the same concept; however, we cannot assume that numerically identical index scores across areas indicate identical degrees of emotional disturbance. Therefore, the index scores may be employed in within-area correlational analyses to test hypotheses in which individuals are units of analysis, but the scores should not be aggregated and used in direct comparisons of average scores or in ecological correlations across areas. The aggregated scores will have meaning only when appropriate evaluative norms are defined for all areas or in comparing areas for which conventional norms are considered reasonable.

The generalizability of our conclusion that the index of emotional disturbance based on 10 CMI questions can be used as a culturally ipsatized method is restricted by features of this study. The samples in this study were selected from regions within countries included in the sphere of urbanized and industrialized culture and, therefore, are more homogeneous than would have been the case had we studied lesser urbanized and industrialized populations such as the Zulus or Eskimos. We noted earlier that it is probably feasible to employ culturally ipsatized measures of mental health in countries which are culturally similar. Furthermore, strict adherence to standardized interviewing procedures in all areas may have enhanced the comparability of measurement.

**SUMMARY**

The major objective of this study was to examine the conceptual equivalence of a set of Cornell Medical Index questions as indicators of emotional disturbance in areas located in different countries. Ten questions were selected from eight sections of CMI. On the basis of the questions’ contents, two subsets of five were identified, one mentioning bodily symptoms and the other personality disturbances. It was hypothesized that these 10 questions are conceptually equivalent indicators of emotional disturbances in all areas, that the same evaluative norms are not applicable in all areas, and that the index score based on the questions is related to sex and perceived morbidity.

Data were gathered in household surveys conducted in 12 study areas of seven countries: Argentina, Canada, Finland, Poland, the United Kingdom, the United States of America, and Yugoslavia. Several statistical procedures were employed to test the hypotheses. First, the associations among the questions were expressed as gammas (Yule’s Q) in each study area, and the median gammas for the two subsets of five questions were compared. The patterns of rank orders of gammas and rank orders of question difficulties were compared among the areas by means of Kendall’s coefficient of concordance. The Mann-Whitney U test was employed to test the null hypothesis that ranks of average index scores are not differently distributed between English-speaking and non-English-speaking areas. Finally, relationships between index scores and age, sex, and perceived morbidity were expressed as product-moment correlations.

In all areas all gammas among the CMI questions were relatively high and positively signed. The differences between the median gammas for the two subsets of questions were generally small, indicating that all 10 questions tapped the same underlying dimension.

In general, few individuals gave affirmative replies to the questions: only one of the 120 question difficulties exceeded 50%. However, the median percentages of ‘yes’ responses, patterns of question difficulties, and mean index scores varied considerably among the areas. The patterns of difficulties and distributions of ranks of mean index scores were generally more similar within areas grouped by language spoken (English-speaking v. non-English-speaking), and mean index scores tended to be higher in non-English-speaking areas.

The relationships of the index score with sex, and two perceived morbidity variables were similar across the areas, but the relationship with age differed. In all areas females and individuals who reported having experienced disability days or a chronic health problem were more likely than others to have high scores on the index.

The implications of these results for cross-cultural research using CMI questions are discussed.

The data reported in this paper were drawn from the World Health Organization/International Collaborative Study of Medical Care Utilization. The major purposes of this study are to describe, compare and explain variations in the use of health services in 12 areas in seven countries, and preparation of the main report of the study is in progress.

The Chairman of WHO/ICS-MCU is Dr. K. L. White, and the Project Directors are: Dr. S. Greenhill, Edmonton, Canada; Dr. V. L. Matthews, Saskatchewan, Canada; Dr. D. O. Anderson, Vancouver, Canada; Dr. J. H. Mabry, Burlington, Vermont, U.S.A.; Dr. D. L. Rabin, Baltimore, Maryland, U.S.A.; Dr. R. F. L. Logan, London, England; Dr. J. M. Paganini, Buenos Aires, Argentina; Dr. T. Purola, Helsinki, Finland; Dr. J. Indulski, Lodz, Poland; Dr. C. Vukmanovic, Belgrade, Yugoslavia; and Dr. I. Brodarec, Zagreb, Yugoslavia.

The WHO/ICS-MCU has received support from the governments of each of the participating nations, the World Health Organization, and several foundations.
The Study Groups were supported by the following grants or agencies: Edmonton, Canada—Department of Community Medicine, University of Alberta; Saskatchewan, Canada—USPHS Contract PH 110-117; Vancouver, Canada—Government of Canada PH Research Grant No. 609-7-218; Burlington, Vermont, U.S.A.—Contract HSM 110-69-79 and USPHS Grant HS00377; Baltimore, Maryland, U.S.A.—USPHS Grant No. 8-R01-HS00110; London, England—USPHS Grant No. 8R01HS00563-03; Buenos Aires, Argentina—Health Survey, Hospital Escuela 'Jose de San Martin'; Helsinki, Finland—National Pensions Institute, National Board of Health and City of Helsinki; Warsaw-Lodz, Poland—PL480 Project No. BHS-MCA-P-1; Belgrade, Yugoslavia—PL480 Project No. BHS-MCA-Y-1; and Zagreb, Yugoslavia—PL 480 Project No. BHS-MCA-Y-4. The computing for the present study was supported by funds from the National Center for Health Services Research and Development of the Health Services and Mental Health Administration, U.S. Public Health Service, Grant No. 8-R01-HS00110.

REFERENCES


—, —, —, —, (1952 b). The Cornell Medical Index Health Questionnaire IV. The recognition of emotional disturbances in a general hospital. Ibid., 8, 289.


—, —, —, —, — (1951). The Cornell Medical Index Health Questionnaire II. As a diagnostic instrument. Ibid., 145, 152.


GIBSON, H. G., HANSON, R., and WEST, D. J. (1967). A questionnaire measure of neuroticism using a shortened scale derived from The Cornell Medical Index. Ibid., 6, 129.


