Chapter 20
The Structure and Stability of the Proposed International Classification

J. Rehm, R.T. Trotter, II, S. Chatterji, and T.B.Ustun'

Introduction

The data presented in this chapter had a direct impact on the content and structure of the ICIDH-2 classification, and the development of assessment instruments. The methods were chosen to help to identify the correct placement of items in the classification, and to present clear information on the cross-cultural stability of the structure of ICIDH-2 as a whole.

Methodologically, the data are at the other end of the spectrum from the data produced to investigate the more global issues concerning disability reported in the previous chapter (key informant, focus group and centre description data). While the data have significant qualitative components, the methods used are designed to provide a qualitative-quantitative bridge in the continuum of cross-cultural applicability methods. The primary question the methods were asked to answer was whether the concepts and items in the draft classification and assessment instruments could be constructed to include only items that met both the cultural and the psychometric requirements for the development of the classification and associated disability instruments.

Methods

The CAR methods chosen for this task were procedures that focused on individual elements of the classification, and the relationships of those elements to one another. The data needs identified for this section included:

* Centre for Addiction and Mental Health, Addiction Research Foundation Division, Toronto, Ontario, Canada (Rehm); Department of Anthropology, Northern Arizona University, Flagstaff, Arizona, USA (Trotter); WHO, Assessmem, Classification and Epidemiology Group, Geneva, Switzerland (Chatterji, Üstün).

1. The need to identify linguistic equivalences for conceptual transfer of elements of the classification into local languages, and back into English;
2. To determine whether the proposed structure of the classification has good cross-cultural stability;
3. To provide an item-by-item evaluation of the cross-cultural applicability of each facet of the classification;
4. To collect data on the boundaries between the three domains of the classification system.

In addition, the methods were employed to gather further information on the potential cultural sensitivity of the items, and to compare the relative importance of different types of disabling conditions in different cultures. Table 1 presents the matches between the methods and data collection needs reported in this Chapter. It should be noted as well that the method selected not over covered these needs but also allowed a significant degree of triangulation of results, without unnecessary duplication.

Table 1. Matching methods with relevant data needs

Research methods (*Types of data collected*)	Research issues for project	Data needs by method
Translation/back-translation and linguistic analysis protocols (*qualitative*)	Linguistic equivalences for items or sections off the classification; identification of problematic individual items.	Identifying linguistic equivalences for conceptual transfer.
Pilesorting (*qualitative and quantitative*)	Cross-cultural stability of the classification; identification of problematic individual items; discovery of underlying cultural dimensions within the classification.	Investigating proposed structure of the classification; and item-by-item evaluation of cross-cultural applicability.
Concept mapping (*quantitative, some qualitative*)	Cultural applicability of items; problems with taboo; age and gender bias; socio-economic conditions; linguistic problems with items.	Cultural contexts, practices and values concerning disabilities; investigating proposed structure of the classification; item-by-item evaluation of cross-cultural applicability; data on the boundaries of the classification system.

The CAR research was conducted at 20[1] sites in a total of 16 nations. Table 2 describes the number of informants represented in the data collected by all sites.

Table 2. Numbers of informants or data collection and reporting sessions

Method	Total *N* Numbers of participating centres
Translation and linguistic analysis	12 (3 used original English version)
Pile sorting	450 informants at 19 centres
Concept mapping	441 informants at 18 centres

Item and classification results

The CAR effort was highly successful, resulting in the direct modification of a number of items in ICIDH-2. The CAR data also provided invaluable direction for the development of epidemiological and other assessment instruments being developed in parallel to ICIDH-2. The data will be presented first by the specific data collection method used, followed by triangulated summary data.

Linguistic analysis

The majority of the key ICIDH-2 items were easily transferable from culture to culture and language to language. There are, however, some problematic concepts in ICIDH-2 that exhibit one or more of six types of primary linguistic incompatibility. These incompatibility problems can be grouped by the following categories.

Category 1: Concepts that were difficult or impossible to translate into one of the CAR study languages

Some ICIDH-2 terms and concepts were difficult or impossible to translate, because there was no equivalent term or cultural concept in the local

[1] The centres which completed all six data collection tasks were three sites in India (Bangalore, Chennai, and Delhi), Japan, The Netherlands, Nigeria, Romania, and Tunisia. Tunisia was responsible for data collection in Egypt, so there was no separate centre narrative for Egypt. In the end, no data was collected from China. The centres which collected all of the data except the focus group data included Canada, Greece, Luxembourg, Spain, Turkey (with sites in Ankara, Istanbul, and Antalya), and the United Kingdom. The United States site (Flagstaff) did not collect key informant interview data or concept mapping data. The Cambodian site was a later addition to the field group and provided a centre description, key informant interviews, focus group data, and the revised item evaluation data set. The sites that collected focus group data included Cambodia, India (three sites), Japan, The Netherlands, Nigeria, Romania, Tunisia, and the United States (Flagstaff).

language. Table 3 shows examples of these types of ICIDH-2 concepts, the linguistic problem, and at least one language in which the condition occurred.

Table 3 Examples of linguistic equivalency problems

Concept or phrase	Problem	Language
1 Affect	Does not translate	Hindi
2 Affect	No idiomatic equivalent	Kannada
3 Disablement	No generic term in Arabic	Arabic
4. Disablement	Cannot be used as umbrella term	Hindi
5 Executive function	Difficult to translate	Hindi, Tamil

Category 2: Concepts whose translation contained a narrowed connotation of the original term, changing the meaning of the original in one of the CAR study languages.

A number of ICIDH-2 concepts could be translated into a participating centre language, but the translation left out connotations that were present in the original term. This makes the concepts only partially equivalent. In some cases, the problem could be overcome by adding additional words or phrases to expand the concept. In other cases, no satisfactory solution was found for the shift in meaning. This could lead to a problem in interpreting data between two centres, if the concepts were embedded in survey questions or coding of classification items. Table 4 provides examples of concepts and words that typify this condition.

Table 4 Examples of concepts and terms with partial equivalency with English

Concept or phrase	Problem	Language
1 Affect	Refers only to emotional state	Tamil
2 Affect	Partial overlap with sentiment	Romanian
3 Handling everyday physical environment	Partial overlap, but primary meaning is temperature, humidity (domestic environment better)	Romanian
4 Well being	Attached to life, not health	Arabic

Category 3: Concepts the translation of which contained an elaborated set of connotations different from the original tertn, changing its meaning in one of the CAR study languages.

A number of ICIDH-2 concepts could be translated into a participating centre language, but the translated term had a larger range of connotations than the original term. This makes the concepts only partially equivalent. In some cases, the problem could be overcome by adding additional words or

phrases to narrow the concept. In other cases, no satisfactory solution was found for the shift in meaning. Table 5 provides examples of concepts and words that exemplify this condition.

Table 5 Examples of concepts and terms with partial equivalency or with additional multiple connotations that interfere with equivalency

Concept or phrase	Problem	Language
1 Affect	Multiple meanings beyond the English	Dutch
2 Function	Translates as act, work, action	Kannada

Category 4: Distinct concepts in the English of ICIDH-2 that translate into a single term when translated into one of the CAR study languages.

There were a number of concepts that identify distinctly different meanings in English, but lose those distinctions when translated into one of the languages of the CAR centres. If these concepts are used in the classification system to distinguish different conditions, functions, processes or elements of the classification system, then the loss of distinction through overlap has the potential for creating confusion between supposedly distinct categories when the classification system is used in other cultures. Examples of the concepts and phrases which included this type of linguistic problem are identified in Table 6.

Table 6 Concepts and terms that merged dunng translation-back translation

Concept or phrase	Problem	Language
1 Disease	Disease, illness, sickness are interchangeable	Hmdi
2 Disease	Disease and illness are interchangeable	Romanian
3 Disorder	Feeling ill, not healthy, illness	Kannada
4 Disorder	Disease and disorder synonymous	Tamil
5 Disorder	Disease and disorder translate to same term	Hindi

Category 5 Cross-culturally inapplicable definitions or examples

A number of items could be conceptually transferred to the other languages through translation, but the definitions or examples that were used to explain the item contained elements that were cross-culturally inappropriate. There were also a few terms and definitions that were inappropriate for use in one or more cultural contexts, because of local cultural values. These conditions were identified in the linguistic analysis at the various sites. The following are examples of items or contextual conditions that caused problems in translating the definitions of the ICIDH-2 items:

from the same respondents or from individuals in the same respondent categories. This data collection process allowed for concept comparison both within and between cultural groups. It allowed the data to be compared from country to country, language group to language group and between key social positions across cultures (health providers compared with lay individuals, and so on).

The raw data for each respondent were typed as ASCII files and imported into a computer program, ANTHROPAC 3.0 (Borgatti 1996). The program contains a data conversion routine that transforms raw pile sort data into an item-by-item similarities matrix. The matrix data were then analyzed using multivariate statistics. One of the analytical processes was to create a Johnson's hierarchical cluster analysis dendrogram of pile sort data (Johnson 1990), and another process was a correlational analysis algorithm (Qualitative Analysis Programme, QAP) that allowed us to compare the classification solutions between cultures and groups. The cluster analysis solution for the total data set was created using the Johnson's hierarchical algorithm in ANTHROPAC. The resulting data points and icicle plots were evaluated for cut points, both through visual inspection and by removing random sets of informants from the data set, to identify stable clusters. A stable grouping of 12 clusters of items was identified.

This raw data set resulted in the identification of the following stable clusters:

Cluster 1: Items relating to participation in community life
community

- Civic and community life
- Citizenship responsibilities
- Following (showing interest in) events that take place outside of the direct environment
- Leisure
- Cultural activities
- Religious activities

Cluster 2: Interaction with other persons

- Keeping appropriate physical contact, and maintenance of social space
- Interacting with an equal, co-worker, peer
- Showing tolerance in relationships
- Managing close personal relationships
- Interpersonal and social relationships

- Managing relationships with friends
- People sharing living space
- Keeping rules, abiding by decisions

Cluster 3: Sexual activities

- Sexual functions
- Dating and forming a sexual relationship
- Performing consensual sexual acts

Cluster 4: Taking care of work and economic responsibilities

- Economic self-sufficiency
- Activities related to fulfilling of financial obligations and services work
- Work acquisition and retention behaviours

Cluster 5: Daily living activities (two sub-cluster)

- Maintaining physical environment
- Handling everyday physical environment

- Taking care of ones health
- Dressing
- Self care
- Keeping self clean and appropriately groomed
- Washing oneself
- Eating and drinking
- Cooking, baking, frying solid
- Planning and organizing meals
- Taking care of meals
- Procurement and care of necessities
- Taking care of pets and domestic animals
- Taking care of household or family members
- Organizing daily routine
- Monitoring and evaluating of performance of activities and tasks

Cluster 6: Mobility

- Motor coordination

- Mobility
- Transferring oneself
- Moving around
- Changing a body position
- Maintaining a body position
- Handling technical devices, aids for locomotion
- Handling body attached technical aids
- Using public transport

Cluster 7: Communication
- Understanding specific signs
- Use of humour
- Following written instructions
- Following verbal instructions
- Visual sensory perceptions
- Using special means of communication
- Use of communication devices
- Non-verbal means of communication
- Written communication
- Communication activities
- Communication content
- Conversation processes and structure
- Responding to conversational cues

Cluster 8: Learning activities (two sub-clusters)
- Problem solving
- . Arithmetic activities
- Acquiring and applying knowledge
- Abilities related to learning
- Study behaviours
- Education

Cluster 9: Sensory activities
- Experience of pain

- Seeing
- Visual sensory functions
- Hearing functions
- Hearing

Cluster 10: Attention, thought, and memory
- Thought, abstraction, judgement, and related executive functions
- Intellectual development and function
- Recognizing directions in space and time
- Orientation
- Perception
- Memory
- Attention
- Consciousness
- Recognizing
- Psychomotor activity

Cluster 11: Emotionally related activities
- Energy and drive
- Performing an activity for an extended period
- Managing general psychological demands
- Expressing empathy
- Temperament and personality
- Affect
- Managing personal behaviour

Cluster 12: Responding to danger
- Managing a dangerous environment
- Responding to dangers

These data indicate that the structure of the classification system, as a whole, represents a cross-culturally stable organization of the relationships among these items, when viewed within a cross-cultural context. While some individual items are not yet cross-culturally applicable, or stable, the overall structure of the classification (i.e., the organization of the items into specific

clusters within the classification) meets the needs and represents the composite viewpoint of both health professionals and individuals directly affected by disabilities across very diverse cultures, language groups, and levels of economic development at the participating centres.

The overall stability of the cluster pattern was checked by randomly removing 30% of the informant responses from the total data set, and then both regenerating the cluster solution with the modified data set and comparing the modified data set with the total data set, utilizing the QAP routine in ANTHROPAC. Second, each separate site data set was compared to the total data set, utilizing the QAP routine. The finding was that the cluster solutions demonstrate a high level of stability within and between or across sites.

The following conditions apply to the overall data cluster solution. The clusters correspond closely to many of the chapters of the ICIDH-2 classification. This indicates that the classification has significant cross-cultural stability in its structure, since the piles that created these clusters were unconstrained (i.e. they were not created using cues about how to cluster them). The items highlighted by asterisks are items that are very stable cross-culturally, and are likely to be the best points for creating stable questions for cross-cultural questionnaires. The weakest items (those that are only loosely tied to a specific cluster) are the items that have the least cross-cultural consensus about their placement. This indicates some form of cross-cultural ambiguity that needs to be resolved by changing the item, or defining it more clearly.

The cross-culturally problematic phrases and concepts identified by the pile sort process were those that individual respondents had significant amounts of difficulty in placing within the classification system. Each of these items contains one or more conceptual elements that make them difficult to translate and to be understood in several cultures, or elements that cannot be easily associated with the other items in the classification system, because of extreme cross-cultural variation in the activity, or in the assumptions attached to the condition. These items have been examined for either significant revision or removal from the concepts used in ICIDH-2.

The problematic items include: "using public transportation," "problem solving," "understanding specific signs," "psychomotor activity," "energy and drive," "managing personal behaviour." Some items, such as "religious activities" and "pain," were often placed in a separate, single-item category. Others, as well as items, such as "following events that take place outside the direct environment" were often consistently placed in two or more clusters. Several items were not understood cross-culturally, such as "sharing living space" or "use of humour." Finally, there were a number of concepts and phrases that could be classified within clusters, but showed only very marginal association with the clusters.

Qualitative analysis of pile sort data

The pile sort data were analyzed by qualitative methods as well as the quantitative approach described above. As a second test for the presence or absence of cross-culturally ambiguous items, those that individuals had difficulty in placing in any specific pile were analyzed on the basis of the information given in the "reasons for pile sort and pile label" cards accompanying each pile.

The pile sort technique nearly always produces a "residual or miscellaneous" category as part of the classification process. These items can be separately identified and interpreted as items problematic for the classification system as a whole. An analysis of these items for the CAR study demonstrated two types of difficulty for the classification: within culture ambiguity and cross-cultural ambiguity. Some items were found to be hard to associate with other items by the people doing the pile sorting at certain sites, but not at others. These, and a second set of items that were identified in "miscellaneous" piles in multiple cultures, are being utilized to target domains for change in the revision process. These items provide a cross-check with the data derived from the mathematical models of the pile sorts, and from the statistical analysis of the concept mapping data analysis conducted for the CAR study. The results confirmed that the items listed above as problematic match those that were ambiguous for the pile sorting, and those that occasioned the most cross-cultural difficulties in the concept mapping exercise.

The pile sort technique is ideal for comparing cross-cultural views on classification systems. The items that show a clear cross-cultural stability and high saliency in the cluster analysis may be strong candidates for anchors for the development of instruments to assess the ICIDH-2 domains, or to use in a short form for assessment of the population. Items for the ICIDH-2 epidemiological and assessment instruments have subsequently been drawn from items that demonstrate high stability in the pile sort cluster analysis. Those items that show the least stability in the pile sort analysis were therefore poor choices as core items for instrument development. The underlying dimensions demonstrated by the multi-dimensional scaling of the pile sort data indicate that a) the overall structure of the classification includes the concept of simple to complex domains, based on complexity of task ; and b) the domains recommended for instrument development capture the majority of the underlying principles embedded in the classification system

Concept mapping results

The concept mapping analysis identified a list of items that were problematic on any of the 10 questions asked about (see Chapter 3 for details): clarity

of the item; clarity of the underlying concept; usability of the item for the culture; usability of the item for all age groups; usability of the item for both sexes; usability of the item for all socioeconomic groups; usability of the item for ethnic and minority groups; cultural sensitivity of items; classification of the items into the levels of impairment, activity, or participation; and importance of the item for ICIDH-2 as a whole.

Table 8 gives examples of items that showed problems either in clarity of the item or its definition. All items that caused problems for at least 20% of the respondents are listed in order from the most problematic to the least problematic.

Table 8. Items and definitions with most problematic scores in the concept mapping

Item name (classification)	Problems %	
	Item	Definition
Community (Participation)	28.6	21.9
Thought, abstraction, judgement, and related Executive functions (Impairment)	24.3	21.9
Orientation (Impairment)	21.4	9.4
Performing an activity for an extended period (psychological endurance) (Activity)	21.1	15.1
Affect (Impairment)	20.7	12.5
Managing general psychological demands (Activity)	20.6	13.9
Keeping appropriate physical contact, and maintenance of social space (Activity)	20.5	12.0

Note: Percentages of problems indicate proportion of respondents who had problems with either the clarity of the item or the clarity of the concept or definition.

Overall, the concept mapping showed a high degree of cross-cultural compatibility of the items in the draft revision of ICIDH-2. This can be seen from the fact that only a very small proportion of the items tested caused problems for more than 20% of the respondents with clarity of either the item or the concept (see Table 8 above). Some items, however, required reconceptualization.

The classification of items into the three levels - Impairments, Activities, and Participation - was not sufficiently clear to all participants: 54.2% of the participant classifications did not correspond to the actual classification in ICIDH-2. As a result, the introduction of the subsequent Beta-2 draft of ICIDH-2 was amended to better explain the new structure and its underlying assumptions. The general results of the concept mapping by disability dimension of ICIDH-2 are summarized in Table 9.

Overall, the Participation disability dimension showed the most problems, especially with regard to usefulness of the items for different subgroups in a culture (age groups, gender, social group, ethnic or minority groups). On the other hand, the Participation items were clearest in terms of their classification within ICIDH-2 - i.e., people made fewer errors in classifying them into Participation.

Table 9. Descriptive results of concept mapping by ICIDH-2-level

		Impairments	Activies	Participation
Clarity of item (% problems)		14.8	12.6	15.0
Clarity of definition (% problems)		12.0	9.6	14.0
Usable in culture (% no)		5.3	5.6	7.1
Usable for all age groups (% no)		8.5	15.0	19.9
Usable for both gender (% no)		3.1	5.8	8.4
Usable for all social economic groups (% no)				
Usable for ethnic or minority groups (% no)		3.8	5.1	8.4
Culturally sensitive (%)		8.7	7.9	6.9
Misclassified (%)		58.7	55.5	34.4
Importance*	Mean	2.4	2.4	2.41
	Standard deviation	1.23	1.14	1.13

(1=not important - 4=very important)

There was most variability between items with respect to questions about usefulness for all age groups and about taboo. Items that were judged to pose most problems with respect to application in all age groups (at least 25% of the sample indicated problems with usability in all age groups) were the following:

Economic self-sufficiency
Work
Activities related to fulfilling financial obligations and services
Cooking, baking, frying solids
Taking care of meals
Dating and forming relationships
Sexual functions
Performing consensual sexual acts

Not surprisingly, the last three items on this list were the only ones seen as prone to be culturally sensitive or taboo (again 25% was taken as the cut-off point).

Overall there was some cultural variability as the results in the country reports suggest (see Chapter 19 above). An important result of the concept mapping analysis was that the dimensions represented by the 10 questions asked were not independent and could be reduced to five dimensions. This was shown by factor analytic techniques taking into account that the underlying data structure was only ordinal (i.e., problem vs. no problem; see Joreskog 1979). The five dimensions become possible because the questions on item and concept clarity loaded on one factor (with factor loadings of 1.0 and 0.86 respectively), all the usability items loaded on a second factor (useful in culture 0.81; useful in all age groups 0.77; useful in both genders 0.83; useful in all socioeconomic groups 1.0; useful in ethnic/minority groups 0.91), and the other items had dimensions (factors) of their own. The emerging factors correlate only to a small degree with each other (between 0.01 and 0.44), with an average correlation of 0.15 (detailed results of the confirmatory factor analysis are available from the first author). As a result, the subsequent field test used only five dimensions in the concept mapping exercise, thus reducing the workload by 50%.

Conclusion

A mixture of quantitative and qualitative techniques was used to study the cross-cultural applicability of the items and concepts in ICIDH-2. A formal translation/back-translation protocol was used to establish equivalences of linguistic terms. Pile sorting and concept mapping identified problems with different structures in different cultures as well as problematic items and concepts. The results of these methods converged to a considerable degree and triangulation was used to identify the most problematic items and concepts. These were changed for the new ICIDH-2 field test version. Moreover, it was found that the classification of items into the three dimensions of disability, namely Impairments, Activities, and Participation was not sufficiently clear to all participants. As a result, the new structure of ICIDH-2 and its underlying assumptions had to be better explained in the introduction of the next, Beta-2, version of ICIDH-2. Finally, the results were used to improve the methodology of the subsequent field tests.