ANNEX 6

COMPREHENSIBILITY TESTING METHODOLOGY

Annex 6

COMPREHENSIBILITY TESTING METHODOLOGY[1]

A6.1 This instrument aims to provide a methodology for the assessment of the comprehensibility of labels and Safety Data Sheets (SDS's) for chemical hazards. The tool has been developed with a particular focus on addressing the needs of workers and consumers in developing countries. The emphasis of instrument development has been to provide a tool that is, as far as possible, globally applicable taking into account varied levels of literacy and differences in cultural experience.

A6.2 Overview of the instrument

A6.2.1 The instrument is organized into a number of modules, directions for each of which is covered in this annex. Broadly speaking, the instrument consists of four parts:

(a) Module 1: This is a focus group, whose main purpose is to ensure that the instruments used in Modules 2 to 11 are sensible across diverse cultures and settings. Its use is recommended in all categories of target populations (see Table A6.2 below) but it should be mandatory to commence with this module in groups of workers and community members from cultures different to the settings in which labels and SDS's have been produced;

(b) Modules 2 to 8: These include a general questionnaire (Module 2) and a set of label and Safety Data Sheet questions and exercises (Modules 3 to 8). Depending on whether the subject is a worker and makes use of a Safety Data Sheet, some elements of these modules may not apply;

(c) Module 9: This is a simulation exercise. One version is intended for workers and is applicable to most people involved in production, while the other version (Module 9a) is adapted for a consumer setting;

(d) Module 10: Module 10 contains a final post-test questionnaire. It is applicable to all participants in the questionnaires (Modules 2 to 8) and the simulations (Module 9). It is also administered to participants in the group exercise (Module 11). The questionnaire is focused on training, and past experience, and offers an opportunity for open-ended feedback and comment on the testing process;

(e) Module 11: This is a group exercise for workers that draws on all elements contained in previous modules and is intended to test comprehensibility in the context of group learning. It is designed to complement Modules 2 to 10 but is carried out on different subjects to those in Modules 1, 2 to 8, and 9.

A6.2.3 It is further proposed that follow-up testing be conducted at one and twelve months after comprehensibility testing. This testing should be repeated on the same subjects who underwent initial testing. Depending on resources and logistics, it may be possible to avoid re-testing on all the modules completed at baseline. Repeat testing would be important to gain insight into retention and real benefits of exposure to hazard messages.

A6.2.4 Table A6.1 summarizes the modules in the instrument, the main activities in the modules, and the objectives and outcomes to be derived from each module.

[1] *Developed by a multidisciplinary team at the University of Cape Town, for the International Labour Office (ILO) Working Group on Hazard Communication as part of international efforts to promote a Global Harmonised System (GHS) for hazard communication.*

A6.2.5 Although the testing instrument has been designed as a self-contained package, it may be possible to make use of selective modules from the battery where there are local priorities and needs. Moreover, it is recognized that as global harmonization of hazard communication evolves, new needs for testing may arise. The instrument may be adapted to take account of new testing priorities over time by using adapted testing materials (labels and SDS's) in the same testing formats. For example, if new icons for hazard symbols are under consideration, module 4 can be amended to include new symbols.

A6.3 Use of Annex 6 and of the testing instrument

A6.3.1 Each module is the actual test questionnaire for a specific set of comprehensibility testing objectives. The layout of the modules is such that instructions are clearly marked in the questionnaires for those administering the comprehensibility tests. Accompanying each module, but presented separately, is a set of detailed guidance notes comprising the manual for the particular module. The manuals also outline the different labels and/or SDS's to be used in each module and the outputs and time requirements of each module.

A6.3.2 To avoid rendering the modules to lengthy, instructions on the modules have been kept to a minimum in the text of the modules, reserving the elaboration on instructions for the manual sections. Where key instructions are present in Modules 3 to 11, they are listed in bold text within shaded boxes to improve ease of administration. Italic font is used throughout the modules for all text to be read out to the subject.

A6.3.3 Some modules (Modules 3, 4, 6, 7, 8 and 9) require random selection of labels and/or SDS's. A box of cards is provided to the interviewer to expedite the selection of a random label/SDS or set of labels/SDS's. The interviewer will have a specific box of such cards marked for every relevant module.

A6.3.4 Labels and Safety Data Sheets are provided but should be to conform to the normative styles and presentations existing in the countries in which the tool is to be applied. The GHS will bring a certain degree of standardization in the content and layout of hazard communication methods but a great deal of variation will still arise in relation to local traditions, styles, size and preferences. Labels and SDS's used in testing must as far as possible reflect the typical local usage patterns. Therefore, although sample labels and SDS's are provided with this manual, users are encouraged to adapt the test materials within the limits of the experimental design requirements so that the materials appear as authentic as possible to local subjects.

A6.3.5 Notwithstanding attempts to simplify the relatively complex testing procedures required to measure hazard communication comprehensibility, the test instrument require careful administration and quality control. Training of interviewers is therefore critical. This is dealt with in more detail in the manuals for Modules 1 and 2.

Table A6.1: Comprehensibility testing: Objectives and outcomes by module

Module	Contents	Objectives	Outcome
Module 1	**Focus groups**	To shape research tool to the context, language, and cultural interpretations of the specific target group. To identify cultural specific definitions of words. To test whether ranking, the use of colour for attributing hazard, and the quantitative estimation of ambiguous variables are culturally transferable. Testing strategies used in subsequent modules are piloted for face validity and identify alternatives. To identify potential biases in the testing situation arising from cultural use of items.	Culturally consistent explanations for difficult words. Appropriate use of colour in local context. Account of cultural factors that would bias comprehensibility tests. Validation of colour blindness test methods. Interpretability of psychometric scales for non-Western populations. Contextual testing. Instruments to capture workers' experience. “Dummy” symbols.
Module 2	**General interview**	To ascertain demographic and other data as a basis for analysis of comprehensibility. To clarify competence in colour and visual acuity necessary for some of the subsequent tests. To collect data on work experience, critical to interpretation of comprehensibility assessments.	Relevant demographic and other data for linking to study results and analysis. Colour and visual acuity assessed. Role work experience plays in comprehensibility.
Module 3	**Recall, reading, and comprehensibility of labels and SDS’s**	To evaluate subjects' familiarity with a label and an SDS. To test subjects' recall of label elements. To evaluate the sequence used to look at label elements. To test the comprehensibility of signal words, colours, symbols and hazard statements. To assess the impact of the label on the subjects': - Ranking of hazard, both to self and to spouse or child, - Intention to use, store and dispose of the chemical. Whether ranking and reporting change after questions on comprehensibility. Can subjects correctly identify the appropriate SDS? Can subjects correctly identify information on chemical name, health hazard, physical hazard and use of protective clothing?	Identify a priori familiarity with labels and SDS's. An assessment of the impact of different label fonts. Identification of poorly understood elements terms. Identify statements with highest comprehensibility. Hazard ranking, and intention to behave as a result of the label. The effect of detailed questions on comprehensibility on subjects' perceptions of hazard as a proxy for training. The impact of the Hawthorne effect will be gauged. Comparison of ranking of hazard to self differs from ranking of hazard to a close relative. Identifying whether subjects can link data from a label to an appropriate SDS in a meaningful way.

Table A6.1: Comprehensibility testing: Objectives and outcomes by module

Module	Contents	Objectives	Outcome
Module 4	**Rating and understanding of hazards:** **Signal words, colours and symbols**	To test subjects' relative ranking for severity of hazard for: - signal words, colours and symbols; - combinations of symbols and multiple symbols; - selected combinations of symbols, colour and signal words. To test understanding of signal words, colours, and symbols. To test opinion on the ability of signal words, colours, and symbols to attract attention. To test whether subjects' perception of the label will influence their reported intention to use, store or dispose of the chemical. To explore subjects' views as to why hazard elements are present on a label.	Signal words, Colours and Symbols will be rated for ability to denote level of hazard, and for comprehension both separately, and for selected combinations of elements. Quality control assessment of face validity of ranking. Ability of label elements to attract attention. Label rated highest for attracting attention will be explored for its ability to: Prompt the subject to identify further information, particularly health hazard information. Influence reported intention to behave in safe ways.
Module 5	**Comprehension of hazard symbols with and without text**	To test subjects' understanding of symbols representing hazard classes. To test subjects' understanding of concepts of hazard classes. To identify whether adding text words improves understanding of selected symbols representing hazard classes: reproductive, carcinogenic, and mutagenic. To identify whether adding signal words improve understanding of symbols representing classes.	Ability to identify the correct symbol for a hazard class. Identification of hazard classes for which symbols perform poorly; and of symbols which perform poorly as indicators of a hazard class. Identify symbols with ambiguous interpretations. The effectiveness of adding text to symbols for reproductive, carcinogenic and mutagenic hazards. The effectiveness of adding signal words to symbols denoting hazard class.
Module 6	**Size, placement background colour and border of symbols/pictograms**	To test the impact of varying symbol size, border and placement. To test the impact of varying background colour and varying icon size in a pictogram relative to border.	Impact of the symbol size, border and placement: - ability to identify chemical name; - perception of risk,; - recall of symbol as proxy for attention to symbol; - recall of hazard statement as proxy for attention to hazard statement; - reported intention to behave; - sequence of reading; Comparison of whether ranking of hazard to self differs from ranking of hazard to a close relative.

Table A6.1: Comprehensibility testing: Objectives and outcomes by module

Module	Contents	Objectives	Outcome
Module 7	**Pictogram comprehension – additional testing (Pesticides)**	To test subjects' ability to identify information: - chemical name; - health hazards; To assess subjects' rating of hazard. To test subjects' understanding of pictograms. To assess subjects' sequence of reading.	Comprehensibility of pictograms: understanding, ranking of hazard, attention, access to key information. Comparison of whether ranking of hazard to self differs from ranking of hazard to a close relative.
Module 8	**Comprehensibility of safety data sheets (SDS's) by organization of data**	To test subjects' ability to identify safety information from an SDS. To test the understanding of hazard information on an SDS. To evaluate what the subject reads on an SDS and the sequence in which subjects report reading the elements of the SDS. To assess what information is useful, appropriate and understandable. To assess whether SDS information is related to intention to behave in safe ways. To evaluate the impact of different organisation of SDS information on the above.	Comprehension of SDS hazard information assessed from different aspects: (a) Interpretation of health hazard information; (b) Self-assessment of understandability to others; (c) Scoring of how the subject explains a hazard statement to a third party; (d) Reported intention to behave Agreement between these four measures of understanding will be estimated. The impact of different ways to organise SDS information will be estimated. Subjective assessment of the usefulness and appropriateness of sub-elements to identify areas for further review of SDS development.
Module 9	**Simulation exercise: impact of the use of labels and SDS's, and of symbols and signal words on labels on safe chemical practices**	To assess safety practices in relation to a simulated exercise in which a chemical is handled. To evaluate whether safety practices are improved by the presence of the signal word "Danger" and/or by the size of the hazard symbol "Skull and Crossbones". To identify whether past experience in relation to chemicals plays a significant role in both safety practices, and in the impact of signal words and symbols on safety practices.	Measures of actual behaviour observed and related to use of labels, SDS's prior to, and during the task. Safety behaviours include use of PPE and other preventive hygiene practices. The impact of varying label elements (with or without "Danger"; with different size hazard symbol) and SDS layout (explicit heath hazard heading versus health hazard data under regulatory information). Relationship between understanding, practice and experimental conditions to be explored.

Table A6.1: Comprehensibility testing: Objectives and outcomes by module

Module	Contents	Objectives	Outcome
Module 10	**Post interview/post simulation interview**	To ascertain past history of contact with chemicals and training. To test the effect of a brief explanation of symbols, signal words, colours and hazard statements on ranking for severity of hazard, and comprehension. To identify chemical information needs from subjects.	Variables derived from training and past experience for stratified analysis of responses to modules 3 to 9. Results will help to indicate whether training should be the subject of more detailed evaluation in the long term. Responses to questions on needs for chemical information can be useful to GHS efforts on chemical safety.
Module 11	**Group exercise - comprehension**	To test whether learning about hazard communication happens differently in a group context than with individuals. To test whether subjects working as a group come up with significantly different answers than when individual subjects are asked a questionnaire.	A quality control assessment on the affect of group versus individual learning. Groups coming up with significantly different responses from individuals indicate that the testing model needs to be revised. Implications for how training should be addressed in future as an element of hazard communication.

A6.3.6 *Consent*: Before conducting any of the modules in this instrument, participants should first give informed consent. To do so, the purpose of the exercises should be explained to them as well as the procedures that will be asked of them. Participants should not be coerced into participating and should know that they have the right to withdraw their participation at any time. The nature of the information provided in the consent procedure is sufficiently generic so as not to give away the explicit hypotheses being tested.

A6.3.7 *Consent procedures* are outlined in the opening sections of Modules 1 (focus group), 2 (commencement of interviews) and 10 (simulation exercises). Irrespective of whether the same subjects complete all modules or not, all three consent procedure should be applied when required. The consent procedure for the simulation is by necessity more of an explanation to obviate the obvious bias to be introduced by alerting the subject to the purpose of the exercise.

A6.3.8 *Policy on rewards or compensation to participants:* Each participating respondent in this study is to be given some form of compensation or incentive for participating in the study. Participating respondents should be told in consenting to the testing that at the end of the study some form of compensation will be presented to them. Compensation may vary from country to country depending on what is culturally appropriate and locally available. Some suggestions (based on other studies) are food (lunch), hats/caps, mugs, food (sugar, rice, mealie, meal), certificates, etc. It is up to the countries applying the tool to develop an appropriate policy on compensation for participants.

A6.4 Sampling

A6.4.1 *Target populations*

A6.4.1.1 Target populations are outlined in Table A6.2 below. These are largely adult working populations, typical of groups who use, distribute or manage chemicals, either directly or indirectly. Children are another important potential audience. However, although the ability to provide understandable safety messages to children is recognised as critically important, it has not been possible to address this area in this manual because of the specialised methods required for evaluation. Further development at some future point may be able to extend the comprehensibility testing to methods suitable for children.

A6.4.1.2 Proposed methods for attaining representative samples are outlined in the Manual sections for Modules 1 and 2. University students should not be used as they have been extensively used in previous hazard communication studies and are not considered representative of the target populations identified in this study.

A6.4.2 *Focus groups*

A6.4.2.1 Given the aim of the focus groups to ensure that the instruments used in Modules 2 to 11 are sensible across diverse cultures and settings, participants for focus groups should be as far as possible typical of the target groups to be evaluated. Emphasis should be placed on targeting groups of workers and community members from cultures different to the settings in which labels and SDS's have been produced. This will mainly apply to farm workers, non-agricultural workers and community/residents/consumer groups, both literate and non-literate, groups whose cultural and linguistic backgrounds may make hazard communication complex. Categories for focus groups are recommended in the Table A6.2 below.

A6.4.2.2 At least 2 focus groups are recommended per category. However, where results from a focus group in one category (e.g. non-literate farm workers) appear highly similar to an analogous group (e.g. non-literate non-agricultural workers), it may be possible to dispense with further groups. This should only be done if the testers are confident that no different results would be anticipated from additional testing. In general, once findings from different focus groups are consistent, it is recommended to proceed directly to the main evaluation (Modules 2 onward). Where findings appear vastly discrepant, or where inadequate information to inform the rest of the instrument has been obtained, it is recommended to continue assembling focus groups until such information is obtained. Under such circumstances, testing until results are consistent or clarity is achieved may require more groups than the 2 per category recommended.

A6.4.2.3 Focus group participants should preferably not be the same workers included in the testing under Modules 2 to 11 as some learning will take place through the focus group itself. Groups should aim, wherever possible, to be homogenous for language, inasmuch as all participants should be able to communicate in at least one common language.

A6.5 Questionnaire and experimental design

A6.5.1 Different sub-populations of working and non-working people will have different experiences that influence their comprehension of hazard communication messages. Modules 2 to 8, and Module 10 will test comprehension under different experimental conditions. Sample size calculations combined with considerations of logistical ease suggest that the minimum numbers of subjects to be tested are those contained in the Table A6.2 below. Modules 6 (effect of label font and layout on comprehensibility) and 9 (simulation test) include comparisons of different label types (8 and 11 strata respectively). Thus, larger numbers are needed for these modules to generate sufficient cases within each stratum. The other interview Modules (3, 4, 5, 7 and 8) have fewer strata (vary from one to four maximum) and thus can be managed with fewer subjects. Users of this instrument may choose to apply all the modules to all participants, in which case the minimum number of participants recommended is as for Modules 6 and 9 in Table A6.2. Modules 2 and 10 must be completed by all participants as indicated.

A6.5.2 In view of the length of the full battery of tests (see Table A6.3), it may be necessary for logistic reasons to break up the instrument by having different subjects complete only some of the modules. In this way, more participants are recruiting to the study but they complete only some parts of the evaluation. If this is the case, remember that all subjects must complete Modules 2 and 10, irrespective of how many of the other modules they complete. For example, the battery of modules could be sub-divided into sets consisting of:

(a) Modules, 2, 3, 8 and 10;

(b) Modules 2, 4 and 10;

(c) Modules 2, 5, 6, 7 and 10;

(d) Modules 2 and 11;

(e) Modules 9, 2 and 10.

However, it is preferable that, if possible, participants are given the full battery of tests contained in the instrument, and are adequately compensated for their effort.

Table A6.2: Sample size - recommended numbers

Category	Sub-category		Focus group: Module 1	Interviews: Modules 2, 6 and 10; Simulation: Module 9	Interview: Modules 3, 4, 5, 7 and 8.
Target Group 1: Workplace a) Management	**Population 1**: Production Managers, engineering, technical		Optional	30-50[a]	25
	Population 2: Supervisory Managers in industry, agriculture		Optional	30-50[a]	25
b) Workers	Population: Farm workers	**3**. Literate	At least one group	100	50
		4. Non-literate	At least one group	100[a]	50
	Population: Workers other than in agriculture	**5**. Literate	At least one group	100	50
		6. Non-literate	At least one group	100[a]	50
Target Group 2: Transport	**Population 7**. Transport workers		Optional	30-50	25
Target Group 3: Community Residents/ Consumers/general public	**Population 8**: Literate		At least one group	100	50
	Population 9: Non-literate		At least one group	100[a]	50
	Population 10: Retailers and distributors		Optional	30-50[a]	25
Target Group 4: Emergency Responders	**Population 11**: Health Professionals, Technical Extension staff and Emergency Responders		Optional	30-50[a]	25
Target Group 5: Other	**Population 12**: Enforcement / Regulatory		Optional	30-50[a]	25

[a] *Recognizing the practical difficulties in organizing a simulation test, it is suggested that in these groups simulation testing only be carried out where resources are available and where practically feasible.*

A6.5.3 As far as possible, the selection of sub-groups should be done an as representative a sample as possible, using random selection of the population for participation. This is critical for generalizability of the results. Even where different participants are chosen from the same sub-group to complete different parts of the instrument, for reasons of length of the battery, selection of participants should emphasize representativeness. However, it is recognized that random selection may be very difficult to achieve in practice. Nonetheless, it should be borne in mind that whatever, selection is used, it should seek to generate a sample as representative as possible.

A6.5.4 Note that within the modules, randomization of subjects within the groups is essential and cannot be compromised on. Randomization is necessary for internal validity of the comparisons and is not the same as random selection of the sample, which is needed for generalizability of the study results.

A6.5.5 *Simulation studies:* Because simulations studies are relatively resource intensive exercises, it is proposed that the simulations only be conducted with limited target populations - workers, both agricultural and non-agricultural, transporters, and consumers. However, where resources permit, these simulations can easily be applied to other strata as desired.

A6.5.6 *Contamination and co-intervention*

A6.5.6.1 The testing design requires control circumstances. For this reason, the situation should be avoided where a participant is able to see or be told of the experimental materials of another participant. This will invalidate the comparisons being made where manipulation of the independent variable is key to the evaluation. Such events occurring in an experimental set up are called contamination.

A6.5.6.2 To avoid contamination, participants should avoid contact with each other whilst testing is being conducted. This may require considerable effort on the part of the testing team to ensure that chance meetings of subjects does not occur. Although difficult, every effort should be made to minimise the probability of contamination.

A6.5.6.3 A distinct but related problem is co-intervention, where both experimental groups are subjected to an intervention occurring independent of the experimental situation. This would occur when, for example, every worker in factory received detailed hazard safety training in the week before the testing was done. It may result in a masking of the effect of the different hazard communication elements and may lead to an under-estimation of the effect of different formulations of the label and SDS. Where this is not preventable, note should be taken of the possibility that co-intervention took place.

A6.5.7 *Group learning*

Module 11 is included to test comprehensibility in the context of group learning. It is applied only to workers (populations 3 to 6 in Table A6.2 above) and will need a sample separate from workers completing Modules 2 to 8. Ten groups should be tested in total including 5 groups of factory workers and 5 groups of farm workers. Groups should aim to be homogenous for literacy level and approximately equal numbers literate and non-literate groups. Each group should not be larger than 10 and not smaller than 6.

A6.5.8 *Context*

A6.5.8.1 The context under which comprehensibility testing is carried out is crucial to the accurate evaluation of meaning and understanding. This is particularly so amongst workers with little formal education who use contextual cues to improve their understanding of hazard messages. For this reason, the bulk of testing in this instrument makes use of complete labels rather than elements of a label or SDS. While well-educated subjects may find it conceptually easier to respond to the isolated elements, the interpretation of such elements may have little bearing to real world learning situations. For this reason, all testing is to be conducted using realistic labels and SDS's.

A6.5.8.2 To maximize realism, an in-site label attached to a container will be used. To attach a different label to each container may pose an unnecessary burden on the tester, so it is proposed that the label be attached to a standard container, and removed after testing. This procedure may require an assistant to the interviewer if overly burdensome for the interviewer. It is important that every visual cue be offered to subjects to maximise their possibilities of comprehension, particularly for workers with low levels of formal education who rely on contextual information to a greater degree. Therefore, the labels should be presented attached to container at all times. A Velcro strip attached to the container may make the procedure relatively simple.

A6.5.8.3 To standardize opportunities for comprehension, the actual chemicals identified in the labels will be spurious chemicals, although made to look as if they could be genuine agents. This aims to retain context, while not disadvantaging those unfamiliar with a particular chemical.

A6.5.8.4 As indicated above, users are encouraged to adapt the test materials within the limits of the experimental design requirements so that the materials appear as authentic as possible to local subjects so as to maximise context.

A6.5.9 *Sample sizes for sub-studies*

Sample sizes for the sub-studies have been calculated based on a two-sided alpha error of 0.1 and a power of 0.8, but have also been tempered by considerations of logistical feasibility. Preliminary piloting of the instrument confirms these estimates. In particular, the simulation exercise has been considered relatively selectively for a smaller number of subjects and target groups, largely because of anticipated logistical constraints.

A6.5.10 *Translations*

A6.5.10.1 Language is key to much hazard communication. Although the instrument seeks to take account as far as possible of language differences, poor and unstandardized translation may introduce considerable error into the testing. For this reason, careful attention needs to be paid to accurate translation. The following procedure should be followed:

(a) Two persons fluent in English (the language of the current instrument) independently translate the questionnaire into the index language (the language of the target group);

(b) Both translations are then translated back into English by a further pair of translators independent of each other and of the original translators.

A6.5.10.2 Back-translations should aim to achieve less than 5% errors on first round. Clarification of the errors in the translation should be conducted to correct ambiguities. Where possible, a combined translation should try to include all elements correctly translated and back translated from either questionnaire.

A6.5.10.3 If the latter is not possible, the translation with the lower rate of errors should be taken as the translation of preference. A second round of back translation will be necessary if errors exceed 5%.

A6.5.11 *Timing of interviews and focus groups*

A6.5.11.1 Interviews and focus groups must be set up at a convenient time for both the interviewee and their employer (when this applies). Farm workers should not be requested to attend an interview during a crucial and busy period for farmers (e.g. planting, ploughing, spraying, or harvest). Workers should be interviewed during working time and should not suffer financial loss for their participation. It is not recommended that workers participate in their own time (lunch or after hours) without adequate compensation. If workers agree to participate during lunch break, the time must be adequate and suitable recompense provided (time back, lunch provided, etc).

A6.5.11.2 Table A6.3 gives the estimated time needed for completion of individual modules based on preliminary piloting with two South African factories. Depending on the module and how skilled the administrators of the modules are, total testing time could vary from 20 minutes to 2 hours. Testing times will be prolonged with non-literate workforces.

Table A6.3: Approximate testing times for hazard communication comprehensibility testing

Module	Time (minutes)
1	60 – 120
2	30 – 45
3	45 – 75
4	75 – 105
5	20 – 30
6	20 –30
7	20 –30
8	45 – 75
9	30
10	30 - 45
11	120 – 180

A6.5.12 ***Rating and coding of responses***

A6.5.12.1 Rating of responses to comprehensibility testing requires expert judgement as to the correctness of the response. Previous experience in Zimbabwe has shown that content analysis of open-ended responses may be feasible where observers are carefully standardized in their approach.

A6.5.12.2 This instrument requires the presence of a set of experts to conduct the rating required for comprehension. The panel of experts should be identified before commencing the study in a process outlined below:

(a) Select a panel with a range of experience, including (one or more) employees, employers and practitioners, as well as researchers skilled in the field of coding and rating;

(b) Convene a workshop with the panel to review the nature of potential responses to questions in each of the modules listed. Review the documentation of the GHS process and aim to arrive at consensus as to what responses would constitute the following categories:

(i) Correct: Meaning is identical, or fully consistent with intention of the GHS construct. This includes responses which are not 100% the same as the GHS meaning but would suffice as the basis for a safety action or precaution;

(ii) Partly correct: Some element of the meaning is correct but it would be insufficient to ensure adequate safety action or precaution;

(iii) Incorrect: Meaning given is either completely wrong, or has very poor relation to the GHS intended meaning;

(iv) Opposite meaning (critical confusions): Meaning given is not only incorrect but indicates an understanding opposite of the intention of the GHS system. Such a critical confusion may result in a dangerous behaviour or action;

(v) Cannot answer/does not know;

(c) Pilot the questionnaire amongst 5 or 10 subjects. Review the results in relation to the criteria selected;

(d) If the results show significant discrepancy, iterate the process above until agreement reached about criteria.

A6.5.12.3 Further coding of responses to questions in the different modules is discussed under each module, where appropriate.

A6.5.13 ***Analyses***

Analyses proposed for these modules are simple computations of proportions and means in relation to different strata. More complex analyses may be undertaken and are indicated in the different modules. An overall estimate for comprehensibility may be attempted by combining results from subjects in the different strata, but should be adjusted for weightings by stratum and by other demographic factors known to affect comprehensibility.

A6.5.14 ***Feedback and follow up***

All subjects should be offered the opportunity of seeing the results of the comprehensibility evaluations, and to give feedback on the interview and testing procedures.

A6.5.15 ***Follow up evaluation***

Subjects participating in these evaluations should be re-interviewed after 1 month and 1 year to assess retention and the medium and long-term benefits of exposure to the GHS hazard messages. Depending on resources and logistics, it may be possible to avoid re-testing on all the modules completed at baseline.