

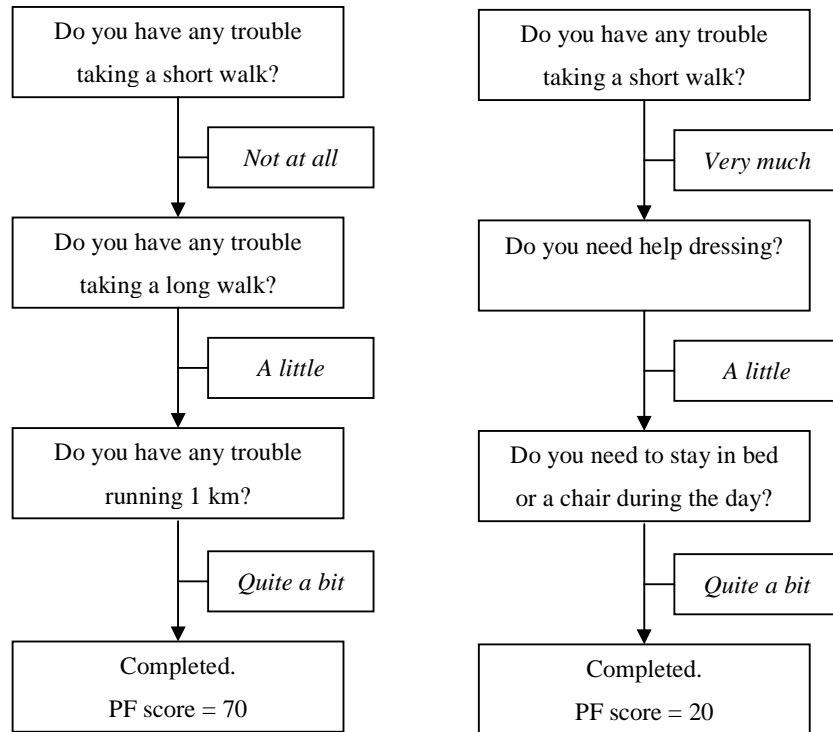
Development of computerized adaptive testing (CAT) for the EORTC QLQ-C30 dimensions

In November 2005 the EORTC Quality of Life Group (QLG) initiated a project aiming at developing computerized adaptive testing (CAT) [1] for the EORTC QLQ-C30 dimensions. The project is coordinated by Mogens Groenvold and Morten Aa. Petersen (Denmark) and the project group comprises several QLG members across Europe* .

The basic idea of CAT is to adapt the questionnaire to the individual patients. This is done by using the responses to the previously asked items to select the most informative next item. The administration of items proceeds until a predefined level of precision has been reached or until a predefined number of items has been asked. Two (fictive) examples of how a CAT measurement of physical functioning (PF) could proceed are shown in Fig. 1. As can be seen, if the respondent reports no problems on an item, the next will concern a more demanding task, while if severe problems are reported the next item will concern a less demanding task.

CAT measurement has several advantages compared to traditional questionnaire measurement including: increased precision, reduced floor and ceiling effects, non-informative questions can be avoided, and the ‘questionnaire’ can be adapted to each study. Furthermore, basing the CAT on item response theory (IRT) [2] ensures that all scores for a dimension (e.g. physical functioning) may be directly comparable across patients and studies even though the scores are based on different subsets of items.

Fig. 1. Two examples of how a CAT measurement of physical functioning could proceed.



The intention is that the ‘QLQ-C30-CAT’ instrument will measure the same concepts as measured with the existing questionnaire, just better. This requires development of new items supplementing the existing items by covering the same concepts and filling out ‘gaps’ (e.g. items for patients in very poor condition may be missing). The development of these so-called item pools consists of the following steps:

- 1) Literature review for each QLQ-C30 dimension to assess which aspects of the dimension the QLQ-C30 items measure and to identify items developed and used to measure the dimension
- 2) Development of new items based on the literature review in step 1. The new items should measure the relevant aspects of the dimension and fit the ‘QLQ-C30 item format’ (e.g., a one week time frame and the response categories ‘not at all’ to ‘very much’)
- 3) Expert evaluations to verify the relevance, appropriateness, and wording of the developed items
- 4) Patient interviews with the same purpose
- 5) Large scale surveys to collect patient responses.
- 6) Psychometric/IRT analyses for the final selection and calibration of items based on the patient responses collected in step 5.

Table 1. The status of the item pool development, fall 2008.

Dimension	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
Physical functioning	Completed	Completed	Completed	Completed	Completed	Dec 08*
Pain	Completed	Completed	Completed	Completed	Dec 08*	
Dyspnoea, insomnia	Completed	Completed	Dec 08*			
Role, emotional, social function, fatigue, lack of appetite, nausea	Completed	Ongoing				
Constipation	Ongoing					

*: Expected start.

Table 1 summarizes the status of the item pool development. As can be seen, we are currently working with 11 of the 15 dimensions in the QLQ-C30. We plan to start the item development for the remaining dimensions by the end of 2008.

As examples of how the item pool development may proceed, Table 2 gives further details of the steps conducted for physical functioning and pain. For both dimensions about 40% of the items identified in the literature measured a relevant aspect of the dimension, i.e. measured one of the aspects measured by the QLQ-C30 items. Many of these ‘relevant’ items asked about the same thing (e.g. ‘climbing stairs’) or were incompatible with the response choices used in the QLQ-C30 (e.g. ‘how many days have you had pain?’). Therefore, compared to the number of identified items, we were able to construct relatively few candidate items for the two pools. The number of candidate items was further reduced based on evaluations internally in the project group and evaluations by experts. This shows that our requirements that the new items should measure the same aspects and have the same item format as the existing QLQ-C30 items limit the number of items that can be constructed. Nevertheless, the item pools for these two dimensions will be huge expansions and thus clear improvements compared to the existing questionnaire. Further, the requirements ensure that the CAT version will be backward-compatible with the parent instrument, will measure within a well-known quality of life structure, and will have a simple and user-friendly format.

In step 5 (data collection) we will need to recruit relatively large numbers of patients. For each dimension the aim is to obtain responses from at least 100 patients from each country and from 1,000 patients in all. For physical functioning the data collection is almost completed. For pain the collection will start when the candidate items have been modified according to the patient

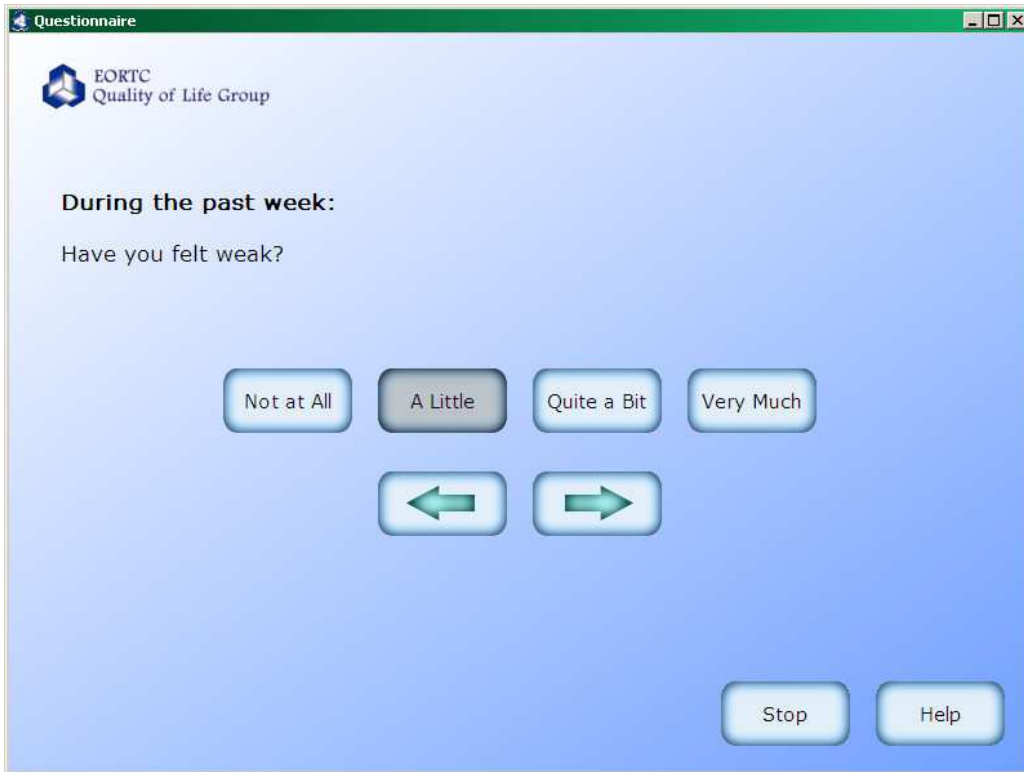
interviews. Anybody interested in participating in this work is welcome to contact Morten Aa. Petersen (map01@bbh.regionh.dk).

Table 2. Number of items after each step in the development process.

		Physical functioning	Pain
Step 1	Identified	975 items	337 items
	Relevant	407 items	140 items
		↓	↓
Step 2	Developed	86 items	39 items
	After internal evaluation	66 items	29 items
		↓	↓
Step 3	Expert evaluations	11 items deleted, 12 reformulated. 55 items for step 4	4 items deleted, 3 reformulated, 1 new added. 26 items for step 4
		↓	↓
Step 4	Patient interviews	43 interviews conducted. 4 items deleted, 12 reformulated. 51 items for step 5	31 interviews conducted, analyses ongoing
		↓	↓
Step 5	Large scale survey	1014 questionnaires collected	(Expected start Dec '08)

In parallel with the item development we are developing the program for conducting the CAT measurement. A first version of this program is ready and we are currently developing version 2. In the second version we have updated the layout (how questions etc. are presented to patient), optimized the computations, and removed bugs. The aim is to make the program as flexible as possible so that it can be adapted to the individual study/patient group. Fig. 2 shows a screenshot example of how the questions are presented in the latest but still preliminary version of the program.

Fig. 2. Screenshot of how the questions are presented in version 2 of the CAT-program.



The initial project grant was for two years. At the EORTC QLG meeting in November 2007 the group decided to continue funding the project for two or possibly three more years (depending on external funding). We expect that this extension of the project will enable us to develop CAT measurement for all dimensions in the QLQ-C30.

*: The EORTC QLG CAT-group consists of the following members: Mogens Groenvold, Morten Aa. Petersen (principal investigators), Neil Aaronson, Thierry Conroy, Wei-Chu Chie, Anna Costantini, Peter Fayers, Jorunn Helbostad, Marianne J. Hjermstad, Bernhard Holzner, Stein Kaasa, Jon H. Loge, Jackie Routledge, Susanne Singer, Martin Taphoorn, Galina Velikova, and Teresa Young.

References

1. Wainer H. (2000). *Computerized Adaptive testing: A Primer*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
2. van der Linden WJ, Hambleton RK. (1997). *Handbook of Modern Item Response Theory*. Berlin: Springer-Verlag.