

## Prior development work

Consultation	Concept mapping workshops (2 with patients, Australia) + patient interviews (Australia)	Nominal group workshops (1 patients, 1 experts, International)
Domains identified	8 health literacy domains	16 health literacy domains
Problem analysis and innovation	<b>Analysis of limitations</b> a) Australian-centric wording b) Only sensitive at the serious problems end of the spectrum c) Provider feedback of limited relevance to some population groups d) Many concepts from initial consultations disappeared	<b>Innovation required</b> a) Validity driven approach b) Use of Bloom's taxonomy to guide item structure and content c) sensitivity across entire range of scale d) tool needs to 'serve' consumers, practitioners and policymakers



### Validity driven process used to develop the Health Literacy Questionnaire

1. Re-analysis      Re-analysis of concept mapping data. Synthesis with OMERACT domains. Identification of candidate items from consultation data. Examination of items in 24 domains identified from prior analyses. Reduction to 11 domains for draft tool after elimination of overlap.
2. Principles and process for writing items      Development of guiding principles for instrument development and framework for hypothesizing item difficulty based on the revised Bloom's taxonomy. Ensure items for a scale rather than a checklist.
3. Consultation and review      Circulation of draft tool to partners in clinical, community health and government agencies, locally and internationally.
4. Domain clarification      Content review to remove redundancies, overlap and checklist concepts
5. Item generation      Use of item writing criteria. Review of all items according to expected difficulty level. Creation of 7 to 10 items per construct. 91 items generated for 11 constructs.
6. Item and scale testing in a calibration sample      Administration of draft items to 634 people across target populations. Psychometric analysis using item difficulty, reliability, confirmatory structural equation modeling and item response theory. Removal of poorly performing items. Cognitive testing, item rewording. 55 items in 10 constructs survived.
7. Item and scale refinement in a replication sample      Administration of items to 405 people. Psychometric analysis using item difficulty, reliability, confirmatory structural equation modeling and item response theory. Removal of poorly performing items. 44 items in 9 constructs survived.
8. Final tool      Refined tool with 44 items across 9 independent scales each with 4 to 6 items, all with clear construct differentiation, high to acceptable reliability, no substantive cross loading or correlated residuals and minimal disordered thresholds.