

2 Method

2.1 The HACC National Service Standards Instrument

2.1.1 Development and structure

The HACC National Service Standards Instrument was adapted from the Quality Measures Instrument developed by the HACC Outcomes Working Group. A major alteration was to split the Quality Measures Instrument into two components: one directly addressing outcomes according to the consumers they affect; and the other using agency-provided information to assess that standards were being upheld. The HACC National Service Standards Instrument is the latter of these two and will be referred to simply as the Instrument. Further modifications to the original Instrument were made as a result of pre-pilot testing and were necessary to better evoke useful performance information against each of the standards.

The pre-pilot test involved semi-structured interviews with a wide range of service providers aimed at obtaining feedback about the practicality, clarity, appropriateness, desirability and comprehensiveness of the standards, and their related consumer outcomes. Thirty-six agencies were included in this stage, covering a range of service types, sizes and locations. Successive refinements were made to the Instrument based on the information obtained during discussions with these service providers. Assessor guidelines to accompany the Instrument were also developed during this phase of the project. The pre-pilot testing of preliminary versions of the Instrument and assessor guidelines enabled maximum benefit to be gained from the pilot stage.

The revised Instrument and the assessor guidelines received the approval of the HACC Officials Standards Working Group before the commencement of the pilot test.

The Instrument comprised seven sections, each concerned with one of the seven objectives. Standards relating to the objectives were contained in these sections. There were 27 standards in total. Performance information requests were listed under each of the standards in order to address the relevant quality issues. Thus, for each standard there were one or more performance information requests for data of a quantitative and/or qualitative nature. The agency was asked to supply supporting documentation where possible and appropriate. Notes explaining what sort of information agencies should provide accompanied each performance information request. At the end of each section pertaining to a standard were check boxes for one of three ratings: 'met', 'partly met' and 'not met'. Agencies were to indicate which rating was appropriate for their agency for each standard. The Instrument concluded with a question asking for an overall rating of agency performance against the standards and with an action plan for meeting standards that had not been adequately met. A copy of the Instrument, in the

form tested in the pilot, is included in a supplementary report (*Developing Quality Measures for Home and Community Care: Technical Appendixes* provided on CD in the back of this report).

2.1.2 Supporting Instruments

The development and testing of the HACC Service Standards Instrument required the concomitant development of a number of supporting Instruments in order to:

- facilitate implementation in the pilot phase (the assessor guidelines);
- test the validity, reliability and comprehensiveness of the quality appraisal process (the Ratings Summary Form, the Assessment of Review Process Survey, the Service Provider Survey); and
- test the clarity, desirability, practicality and appropriateness of the standards (the Assessment of Review Process Survey, the Service Provider Survey).

Each of these supporting instruments is described below.

Assessor guidelines

Purpose: To assist assessors in rating agencies.

During pre-pilot testing of the Instrument, guidelines were developed to assist all assessors in deciding on agency ratings against the standards. The guidelines were developed to be applied to the greatest possible range of HACC funded agencies. As such, they did not include detailed and prescriptive checklists of 'things to look for' to make ratings decisions against the standards. Instead, they relied on assessor training in the application of the guidelines and service standards and upon assessors using their experience of quality of service across agencies to arrive at a score for the agency against each standard. By this method the performance indicators against the standards could exist as more general non-prescriptive indicators, allowing the knowledge and experience of the project officer to aid in the application, reliability and generalisability of the appraisal process. This method has been shown to be a reliable measure of compliance with service standards in nursing homes (Braithwaite et al. 1991; Braithwaite & Braithwaite 1995). A copy of the original assessor guidelines as used in the pilot is included in a supplementary report provided on CD in the back of this report (*Developing Quality Measures for Home and Community Care: Technical Appendixes*). In the revised version of the Instrument, the assessor guidelines have been incorporated and no longer exist as a distinct entity.

Assessment of review process

Purpose: To provide assessor input concerning the validity of the Instrument and concerning the effectiveness of the assessment process.

The Assessment of Review Process questionnaire was given to assessors after they had completed their assessment interviews with agencies. In this questionnaire assessors were asked to indicate the level of difficulty they experienced in rating each of the standards and were asked to indicate how appropriate the standards were to the agencies they assessed. Assessors were also asked to comment on the Instrument's comprehensiveness and balance.

Seventeen Assessment of Review Process questionnaires were returned to the Institute from the 53 assessors who participated in the pilot. A copy of the Assessment of Review Process questionnaire is included in a supplementary CD (*Developing Quality Measures for Home and Community Care: Technical Appendixes*).

Service provider survey

Purpose: To provide service provider input concerning the reliability and validity of the Instrument.

After they had completed the Instrument and, where appropriate, their assessment interview, agencies were asked to complete the Service Provider Survey. This questionnaire asked service providers to indicate standards' ratings that they disagreed with and to comment on the reason for their disagreement. Service providers were asked to indicate whether they considered the standards to be clear, desirable, practical and appropriate. They were asked to comment on whether they considered the Instrument sufficiently comprehensive and were also asked a number of questions as to whether they believed that the standards were achievable. The survey also collected some descriptive data concerning the agency, such as the number of clients, the time in operation, and their membership in a chain or group of agencies under one auspicing body.

Of the 145 Service Provider questionnaires sent out, 102 were returned to the Institute. A copy of the Service Provider questionnaire is included in a supplementary CD (*Developing Quality Measures for Home and Community Care: Technical Appendixes*).

Ratings summary form

Purpose: To provide data for analyses of the Instrument's reliability.

When assessors had decided on the ratings the agency should receive against the standards, they were asked to record these ratings on the Ratings Summary Form. These ratings took the same form as those in the Instrument: a choice of one of three ratings, 'met', 'partly met' or 'not met'. One additional summary item asked for an overall appraisal of agency performance. Assessors were also given the opportunity to write comments against standards ratings, indicating what action they thought the agency must take to meet the standards. These actions were then to be used to add to or alter the forward action plan at the end of the Instrument.

Assessors received Ratings Summary Forms for each assessment they undertook. Of the 94 Ratings Summary Forms were sent out to assessors, 74 forms were returned to the Institute. A copy of Ratings Summary Form is included in a supplementary CD (*Developing Quality Measures for Home and Community Care: Technical Appendixes*).

2.1.3 Sample

Given the project's aim of developing an Instrument applicable across the diverse range of HACC service types, the sample of agencies for the pilot was selected to be representative of agencies according to agency size, service type, location and outlet type. Representation of other characteristics was also sought in participating agencies.

These included:

- agencies that were providing services in rural and remote locations;
- agencies that were providing services in lower socioeconomic areas;
- agencies that were providing services specifically to people of a non-English-speaking background or to Aboriginal and Torres Strait Islander peoples; and
- agencies which had been reviewed through quality appraisal processes such as the Disability Services Standards, Community Health Accreditation and Standards Program (CHASP) or Australian Council of Healthcare Standards (ACHS).

The selection of agencies according to these characteristics was constrained by a number of factors. The most influential were, firstly, the ability of each State to provide assessors to conduct the assessment interviews as required and, secondly, conflict with other HACC program activities in particular areas and at particular times. Given these constraints, States were asked to select agencies for the pilot themselves but with certain conditions. Agencies were to fall into the categories determined by the project team (listed in the tables that follow), and they were to represent the range of service quality (that is, they were not to be selected on the basis of their likely performance against the standards in the pilot).

Five agencies withdrew from the pilot prior to commencement, leaving 162 available to participate in the pilot. Completed Instruments were received from 145 of these.

Agencies were selected from the following States and Territories: the Australian Capital Territory, the Northern Territory, Queensland, New South Wales, South Australia, Victoria and Western Australia. Tasmania did not take up the option to be part of the pilot project. The number and proportion of agencies involved in the pilot, selected on the basis of the specified criteria, are shown in the following tables (Tables 2.1 to 25).

Table 2.1: Agencies participating by size

Size	Number	Proportion
Small	37	25.5%
Medium	53	36.6%
Large	34	23.4%
Unspecified	21	14.5%
Total	145	100%

Table 2.2: Agencies participating by outlet type

Outlet type	Number	Proportion
Agency	113	77.9%
Council	22	15.2%
Hospital	10	6.9%
Total	145	100%

Table 2.3: Agencies participating by service type

Service type	Number	Proportion
Multiple services	38	26.2%
Home help/personal care	18	12.4%
Community nursing	13	9.0%
Allied health	5	3.4%
Respite	36	24.8%
Home maintenance/ modification	3	2.1%
Transport	5	3.4%
Meals	14	9.7%
Case coordination/ management	10	6.9%
Social support	3	2.1%
Total	145	100%

Table 2.4: Agencies participating by review type (n=145)

Review type	Number	Proportion
Disability Services Standards	20	13.8%
ACHS	12	8.3%
CHASP	12	8.3%
Unspecified	106	73.1%

Notes

1. Unspecified indicates cases where it was not known where agencies had undertaken a review, or it was not known what type of review they had undertaken.
2. Total number in table is greater than 145 because some agencies had undertaken more than one type of review.

Table 2.5: Agencies participating by other specified criteria (n=145)

Other specified criteria	Number	Proportion
Services to Non-English-speaking background people	8	5.5%
Services to Aboriginal and Torres Strait Islander people	3	2.1%
Low socioeconomic areas	4	2.8%
Rural areas	10	6.9%
Remote areas	11	7.6%
Unspecified	109	75.2%

The sample was further divided according to the method of assessment employed for the service appraisal (see Section 2.2). The breakdown by assessment type and State or Territory is presented in Table 2.6.

The number of agencies participating in each assessment method was constrained by the availability of assessors to participate. As the participating States and Territories were required to provide assessors, they determined the number of agencies assessed by each method according to their resources. States and Territories also chose to trial only those methods they believed they were likely to implement. The number of agencies in each assessment method reflects these constraints, in combination with the need to obtain enough agencies in each category to allow for the application of statistical tests of significance, and the concern to reflect a range of likely agency quality in each assessment category.

Table 2.6: Method of assessment by jurisdiction

State	Approaches piloted	No. of agencies selected for the pilot	No. of Instruments returned	No. of agencies withdrawn
New South Wales	Self-assessment	79	69	2
Victoria	Self-assessment with verification	8	8	0
	Joint assessment	4	4	0
	Independent assessment	3	2	1
Queensland	Self-assessment with verification	14	13	0
	Joint assessment	8	6	0
	Independent assessment	3	3	0
Western Australia	Self-assessment with verification	10	9	1
	Joint assessment	4	4	0
South Australia	Peer review	18	18	0
Australian Capital Territory	Self-assessment	8	6	0
Northern Territory	Peer review	8	3	1
Total		167	145	5

2.1.4 Procedure

Once State or Territory representatives had selected agencies for inclusion in the sample and confirmed their ability to participate, the contact and postal details were forwarded to the project team.

The project team distributed the Instrument to agencies according to the contact details supplied by the coordinating State and Territory authorities. An accompanying cover letter described the purpose of the pilot and the method of assessment the agency was to follow. Agencies were allowed a minimum of one week to complete the Instrument.

Assessors then completed their appraisal of the agencies according to the pre-determined assessment method. Assessor ratings against standards were recorded on

the Ratings Summary Forms. Assessors returned the completed Instruments and their completed Ratings Summary Forms to the Institute when the assessment interview had been completed.

2.1.5 Methods of assessment

Five methods were tested in which agencies were assessed against the standards using the Instrument. These methods were born of methodological and practical considerations. One method considered to be less methodologically sound (self-assessment) nevertheless was viewed by jurisdictions as practical to implement given likely resource constraints (this method did not involve a visit to the agency during assessment). Rather than dismiss any of the potential assessment methods without adequate testing, it was determined that they should be included in the pilot to facilitate more rigorous comparison. Five approaches to quality appraisal were thus trialled: self-assessment, self-assessment with verification, joint assessment, peer review assessment, and independent/external rater assessment.

Self-assessment

Agencies were required to fully complete the Instrument without the assistance of an assessor. This included the individual standards ratings, overall appraisal (with the exception of the comments and date of next review) and the forward action plan (with the exception of the time frame and the person to complete the task). To assist in this process, a short guide to completing the Instrument was sent to agencies along with the Instrument. This guide outlined what the Instrument required and concisely described how to arrive at scores against the standards. Box 2.1 shows an extract from this guide. When agencies had completed the Instrument according to these instructions they then mailed the completed Instruments to the Institute.

Seventy-nine agencies were selected for self-assessment. Of these, 69 returned Instruments to the Institute within the time frame of the pilot. Two agencies withdrew, one had closed during the time of the pilot and the other failed to receive the Instrument due to incorrect postal addressing.

The cover letter sent to self-assessing agencies indicated that a proportion of agencies would later be selected to undertake a random verification interview. At the time of completing the Instrument the agencies did not know whether they would be a part of this sub-sample. The project team selected these agencies randomly within categories of service type. When Instruments had been returned, the selected agencies were contacted and an assessment interview was scheduled. Assessors were forwarded the agencies' completed Instruments in preparation for this interview.

During the verification interview, the assessor discussed the agency's responses against the Instrument and viewed the relevant documentation. The assessor completed the Ratings Summary Form, and returned the Instrument and the Ratings Summary Form to the Institute.

In the Australian Capital Territory, four agency self-assessments were verified. In New South Wales, 20 agency self-assessments were scheduled for verification, but only 10 of these were completed within the time frame of the pilot.

Box 2.1 Excerpt from the guide to scoring given to agencies undertaking a self-assessment

'Not met' rating

A 'not met' rating is appropriate where your agency does not satisfy the consumer outcome at its most basic level. For example, Consumer Outcome 6.1 states that 'Consumers are aware of the complaints process'. A 'not met' would be appropriate if your agency did not take steps to ensure that consumers received information about the complaints process and a 'not met' would be appropriate if your agency did not take some action to facilitate consumer understanding of this information.

'Partly met' rating

A 'partly met' rating would be appropriate where the consumer outcome is satisfied at its most basic level but where a number of other factors should be changed to achieve a better outcome under the standard. The Instrument provides prompts for what these factors may be for each standard on the page opposite to the one where you complete your performance information. Using Consumer Outcome 6.1 as an example again, if your agency satisfies the outcome at its most basic level but no action is taken to periodically remind consumers of the complaints process and inadequate provision is made for special needs groups in informing consumers of the complaints process, then a 'partly met' rating would be appropriate.

'Met' rating

A 'met' rating is appropriate where your agency has satisfied the consumer outcome and has been able to respond positively to each of the points listed in the Instrument. The 'met' rating is not, however, intended to be the equivalent of a score of 100% against a standard, nor is it intended to reflect the notion of best practice. Rather, an agency may have satisfied the standard and the majority of points under the consumer outcome, but may still be able to make changes to improve the quality of service. For example, under Consumer Outcome 6.1 your agency may have been able to answer positively against each point but may have found some scope for improving the awareness of the complaints process for a particular non-English-speaking background group.

Two measures of your agency's overall rating against the standards are requested on this form. The numerical score is obtained by adding the scores made against each standard. As the Instrument indicates, a 'met' rating scores 2, a 'partly met' scores 1, and a 'not met' scores 0.

The second measure of overall performance requires that you make a judgment about how well you think your agency has performed against the standards; whether it meets the standards to an exemplary level, to a good standard, to a minimal or basic standard, or whether it fails to meet the standards. Your decision about this rating should reflect your own opinion about your agency's performance against the standards, irrespective of the numerical score calculated in the previous question. Your responses to this question will be used as a check on the validity of the standards scoring system.

Self-assessment with verification

The self-assessment with verification method also required agencies to fully complete the Instrument including ratings against the standards, overall appraisal and action plans, without the assistance of an assessor. They did not receive the short guide to scoring distributed to agencies undertaking self-assessment. The ratings they gave themselves were considered to be draft ratings. When they had completed the Instrument the agency received a visit from an assessor. Using the assessor guidelines, assessors discussed the agency's responses against the Instrument and viewed the agency's relevant documentation in order to reach final ratings for each standard. Assessors recorded the ratings they believed the agency should receive on the Ratings Summary Form. Where assessors and agencies continued to disagree after discussion this was simply recorded as conflicting entries on the Instrument (containing the agency's self-ratings) and the Ratings Summary Form (completed by the assessor). After the assessment interview, the assessor returned the Instrument and the Ratings Summary Form to the Institute.

Thirty-two agencies were selected for self-assessment with verification in Victoria, Queensland and Western Australia. One agency in this group withdrew from the pilot due to management changes and another did not complete the self-assessment with verification process within the time frame of the pilot.

Joint assessment

Joint assessments did not require agencies to complete the ratings against the standards, the overall appraisals, or action plans prior to an assessment visit. They were required to write answers against the performance information requests and to gather together relevant documentation. Making use of the assessor guidelines, assessors were to interview agencies and, in discussion with the service provider, come to a joint decision about the ratings that it had achieved under the standards. These jointly determined ratings were recorded on the Instrument by the service provider and the Ratings Summary Form by the assessor. Where the agency and the assessor continued to disagree about ratings after discussion, they were instructed simply to record their differing ratings on their respective forms. The assessor returned the Instrument and the Ratings Summary Form to the Institute after the interview.

Sixteen agencies were selected for joint assessments. Two Instruments were not returned within the time frame of the pilot.

Independent or external assessor

This assessment method differed from the others tested in that it used assessors who were independent of or external to the regional HACC program. These assessors thus did not have prior knowledge of the agencies they assessed, but did have some HACC program knowledge. The two independent raters involved in the pilot were State government officers whose duties did not normally involve dealing with regional HACC agencies. Both had extensive experience in the area of quality appraisal.

Independent raters carried out assessments as either self-assessments with verification or joint assessments. Independent raters used the Ratings Summary Form to record

their ratings of agency performance. This form was completed during the assessment interview.

Six agencies were selected for independent rater assessments. One agency withdrew from the pilot.

Peer review assessment

Peer review assessments were conducted in both the Northern Territory and South Australia. The methods employed were, however, quite different in the different States. In the Northern Territory, peer review assessment differed from the other methods tested in that it used assessors who were staff members of other HACC agencies. Agencies to be assessed completed the Instrument following either the joint or self-assessment with verification method and discussed their responses with peer assessors in an interview. Instruments completed by the agencies, and Ratings Summary Forms completed by the assessors, were then sent on to the Institute.

Of the eight agencies selected to have a peer review assessment in the Northern Territory, one withdrew from the pilot. Seven assessments were conducted with agencies. Three Instruments were returned to the Institute.

Peer review assessments in South Australia were conducted as desk audits of completed agency Instruments. Agencies fully completed the Instrument including ratings and forward action plans. This was done without the assistance of an assessor and without the short guide to scoring distributed to agencies undertaking self-assessment. Completed Instruments were then sent to a peer review panel.

Peer review panels were three-member teams composed of staff members of other HACC agencies. Each three-member peer review team conducted six agency assessments. Panels met together away from the agency being assessed to view and discuss the completed Instruments and provided documentation. Using the assessor guidelines, the peer review panels reached final ratings for each standard. Panel members recorded the ratings they believed the agency should receive on the Ratings Summary Form. The panel then contacted the agency to provide feedback on their appraisal. Instruments and Ratings Summary Forms were then sent on to the Institute. Agencies were not given the opportunity to change their ratings.

Eighteen agencies were assessed by peer review in South Australia. All Instruments were returned to the Institute.

2.1.6 Training of assessors

Prior to their first assessment interview, assessors participated in a one-day training session to become familiar with the Instrument, the assessor guidelines and the assessment process. The assessor guidelines and Ratings Summary Forms were distributed to assessors in this session.

The method of using the assessor guidelines to arrive at scores for agencies against each of the standards was discussed during training. While the assessor guidelines provided criteria against which agencies could be assessed, assessors were also explicitly instructed to use an '80/20 rule'. According to this rule, 'met' was not presented as the equivalent of 100% or the best possible performance, but rather that 80% or more of the

listed criteria had been achieved by the agency. The agency must also have met any specified basic criteria for the standard but, in other respects, the assessor was to follow the 80% rule, that is, they may observe that the agency has four out of five things in place. 'Not met' was appropriate where an agency failed the basic criteria where these were specified as such in the guidelines. Where basic criteria were not specified, a 'not met' rating was appropriate where 80% of criteria listed under the standard were not achieved, that is, the agency had failed to achieve four out of five criteria. A 'partly met' was advised where an agency satisfied the specified basic criteria but perhaps only half of other criteria, that is, the agency achieved somewhere between 20% and 80% of the listed criteria. This rule was devised with the intention of allowing assessors some leeway to exercise their judgment and knowledge of the agency's circumstances when interpreting the requirements of the standards.

A role-play and assessment exercise using a fictitious agency's response to the Instrument provided a medium for introducing both the Instrument and the scoring method. Group discussion after each of these exercises clarified the scoring method, the assessor's role and the purpose of the pilot. It was emphasised that, although the assessor guidelines would provide some indication of appropriate ratings, the assessors themselves were to use their judgment and knowledge of the agency's circumstances to come to a scoring decision. In this way, the Instrument could be applied to a wide variety of HACC agencies.

The process of each of the relevant assessment methods was described, highlighting what was required of assessors in each.

2.1.7 Feedback regarding the pilot

Feedback regarding the Instrument and the assessment process was sought from both assessors and agencies. As noted in Section 2.1.2, when the assessment process was completed Service Provider Surveys were distributed to agencies. Of the 145 Service Provider questionnaires sent out, 102 were returned to the Institute.

The main source of feedback from assessors was from debriefing meetings. During these meetings, assessors in each State met with members of the consultancy team to discuss their experiences during the pilot. Almost all assessors were able to attend these debriefing sessions. An Assessment of Review Process survey form was also distributed during these meetings to obtain quantitative feedback from assessors regarding the Instrument and assessment process. Fifty-three assessors participated in the pilot test of the Instrument. Of these, 17 returned an Assessment of Review Process Survey. Part of the reason for this low response rate was that peer review teams provided their own comprehensive written reports of the peer review process. Of the 38 non-peer review assessors attending debriefing sessions, 45% returned the Assessment of Review Process Survey.

2.1.8 Reliability study

A reliability study was conducted to ensure that the Instrument would produce ratings for agencies that did not vary when different people conducted the assessment. To test this, members of the consultancy team conducted a second assessment on a sample of agencies. Two methods of testing inter-rater reliability were used. Desk audits involved

assessing agencies using their provided documentation and written answers on the Instrument only. Reliability assessment visits involved a member of the consultancy team undertaking a visit to agencies after the assessor had conducted their assessment interview with the agency. In this way the veracity of two assessment methods could be tested: assessment by visit and assessment by documentation alone.

Desk audits were conducted for 10 New South Wales agency Instruments. These agencies had undertaken self-assessments (see Section 2.2) that had been verified by assessors.

Reliability visits to agencies were conducted for 15 agencies in South Australia, Queensland, Victoria and Western Australia. Of these 15 agencies, five had been assessed using the self-assessment with verification method, five with peer review assessment, and five had undergone a joint assessment.

These reliability studies were conducted after assessors had returned all agency Instruments and accompanying documentation to the Institute. In both tests of inter-rater reliability, the second rater did not have knowledge of the ratings given to the agency by the assessor. That is, they did not view the Ratings Summary Forms for agencies. The agency's self-ratings had been recorded on the Instrument that the reliability raters were using as the basis of their assessment. As such, they were aware of the agency's own opinion of what ratings they deserved, thus placing the reliability raters in possession of the same written information as the assessors who conducted the assessment interview with the agency. The important aspect of the reliability study was that reliability raters should decide on ratings for agencies independently of the previous assessor.

For both methods, reliability raters recorded their ratings on a Ratings Summary Form. Quantitative analyses later compared Ratings Summary Form responses of reliability raters and assessors to establish the inter-rater reliability coefficient and per cent agreement.

2.1.9 Testing for validity and reliability

Testing and refining the HACC National Service Standards Instrument required an assessment of its reliability and validity. The validity of the Instrument refers to the how effectively the Instrument measures what it is intended to measure – quality consumer service in HACC-funded agencies. Reliability refers to whether the Instrument will produce the same results for agencies under different conditions of administration. Tests were made of the face and content validity of the Instrument, its internal consistency, and its inter-rater reliability.

Face and content validity

Content validity is concerned with the extent to which the Instrument adequately covers the domain of service quality that it is intended to cover, that is, the extent to which it is sufficiently comprehensive and balanced. The face validity of the Instrument refers to the extent to which 'on the face of it' the Instrument provides a measure of quality in HACC-funded services. Hence its measures should have some meaning to those in the community care sector. Prior to pilot testing, discussions were held with service providers to gain their opinions on ways in which the Instrument could be

made more comprehensive and useful for the assessment of quality in their agency. After pilot testing of the Instrument a more quantitative examination was conducted, via the Service Provider Survey, of service providers' opinion on whether the service standards and their associated performance information requests were clear, desirable, practical and appropriate. At this time, assessors were asked to indicate how appropriate they found the Instrument to the agencies with which they conducted assessments and were asked which standards they found difficult to rate (the Assessment of Review Process Survey). Service providers who had had assessors rate their agencies were also asked which standards ratings they were critical of and why. Just as the Instrument must appear to collect valid indicators of quality performance, the assessment process must also be free from apparent flaws in its validity. Qualitative data, obtained from assessors during the debriefing sessions, were used to evaluate the face validity of each of the assessment methods.

Internal consistency

The HACC National Service Standards Instrument is intended to provide a measure of quality. Internal consistency addresses whether each of the various components of the Instrument contributes to a sensible and coherent measure of this quality. Internal consistency is assessed by statistical methods. The results of factor analyses of the ratings against the standards were examined, as were alpha reliability coefficients for the groups of standards that related to each of the seven objectives. Correlations between objectives indicate whether each of the objectives is sufficiently related to another to be considered as contributing to the measurement of a single construct.

Concurrent validity

When two different measures of the same or similar construct agree they are said to have concurrent validity. Their agreement provides evidence to confirm that the measures are tapping into the same underlying factors. The concurrent validity of the Instrument was tested by comparing the ratings against the standards with an overall appraisal of agency performance as perceived by assessors and agencies. The correspondence between agency self-ratings and assessor ratings also provided a measure of concurrent validity.

The degree to which agencies' and assessors' ratings converged provided evidence of the concurrent validity of the Instrument. Examination of this agreement within and across assessment methods provided an indication of the degree to which each of the methods supported this concurrent validity. Similar comparisons were made across assessment methods for the concurrent validity of Instrument Scores and the overall appraisal of agency performance.

Inter-rater reliability

The HACC National Service Standards Instrument tested in the pilot contained performance information that required that a third party, an assessor, use his or her judgment and knowledge of the service to decide on ratings against the standards appropriate for the agency. An important question arising from this circumstance is

whether the ratings an agency received would depend more on the assessor than on the quality of service in the agency. Assessors received training on the use of the Instrument and guidelines to assist them in making ratings decisions. These methodological processes were designed to support inter-rater reliability. Nevertheless, some assessors could have been perceived as tougher than others or some more sophisticated in their approach to assessment. Assessors varied in their level of experience in dealing with or working in HACC-funded agencies: some were government officers, others were peers of the service providers. An important aspect of the study, therefore, was to determine the level of reliability between raters. Two methods of reliability assessment were used: one involved a second appraisal by a reliability rater with 15 agencies; the other involved 10 desk audits of agency-completed Instruments. In both cases, a second set of ratings were generated and then compared with those given to the agency by the first assessor (see Section 2.2.3).

Inter-rater reliability refers to the tendency for the Instrument to be applied consistently by different raters. The method of assessment may also have affected this reliability. The use that assessors made of the Instrument and the tendency for their own biases and assessment styles to enter the assessment process may have been affected by assessment method. Although the sample size was small, comparisons of inter-rater reliability were made across assessment methods. These comparisons were based on the results of the second interview conducted by the reliability assessor with 15 agencies. Five of these agencies had previously undergone peer review, five had undergone self-assessment with verification, and five had cooperated in a joint assessment.

Inferential tests of the difference between methods

Agencies were selected for the pilot using criteria which sought to involve agencies of the full range of service quality. The choice of assessment method to be used with agencies also was not to be determined by the agency's expected performance but was intended to vary across the full range of agency quality.

By this selection method it can be assumed that when agencies are grouped by assessment type, the true service quality of agencies in each group should not, on average, vary substantially. If the assessment methods are each equally effective at reflecting the true service quality of agencies and the Instrument is reliable, then the Instrument should produce the same average performance scores for the group of agencies using each different assessment method (within a degree of error to be expected by chance).

All other things being equal, it can be assumed that if differences occur between the average performance of the agencies in each assessment type, then this difference is likely to be the result of factors associated with the assessment method.¹ Significance tests were conducted on the differences between the mean scores for each group of agencies according to assessment method.

1. This assumption is somewhat compromised by the absence of a properly randomised sample.

2.2 The consumer survey

2.2.1 Development and structure

Pre-test

The original draft Quality Measures Instrument developed by the Outcomes Working Group did contain items requiring consumer feedback, but these were not constructed to form part of a consumer survey. It was therefore necessary to design a consumer questionnaire and to devise a method for its implementation.

An editorial sub-committee for the HACC Officials Standards Working Group drafted a set of consumer questions. This group consisted of a consumer representative, a service provider representative, members of the HACC Standards Working Group, and a member of the consultancy team.

The items were tested and refined in three iterations conducted in the Australian Capital Territory: the first test consisted of five face-to-face interviews, the second of five telephone interviews, and the third of 10 mailed questionnaires. After each test, the questionnaire was modified to better collect consumer views of agency service quality. Formatting and layout of the mailed version were altered to aid readers in replying to the questions. Where appropriate, open-ended questions were replaced with fixed multiple-choice options. Some changes were made to language to avoid technical or bureaucratic terminology which may have compromised clarity and comprehension for general service users. Additional information was supplied for issues that were found to be confronting or confusing to consumers. Redundant questions were eliminated. A small number of items were added to collect demographic data on respondents.

The agreement of the HACC Officials Standards Working Group was received for the revisions. The final Consumer Survey Form also received the clearance of the Australian Institute of Health and Welfare Ethics Committee before distribution to consumers.

2.2.2 Instrument

The consumer survey tested in the pilot contained questions listed under five sections: Provision of Services; Rights and Information; Satisfaction with Services; Advocacy; and General Information. Each question in the first four sections was specifically designed to measure agency performance as it related to a particular standard. In this way, consumer appraisals could be matched directly to agency performance against the standards. Four questions, listed under General Information, sought information on characteristics of the respondents, including their age, sex, carer status, and membership of a special needs group.

Two methods of receiving consumer feedback regarding agency performance were tested and compared: telephone interviews and mailed surveys. There were some minor differences between the questionnaires trialled in the telephone interview and mailed survey. To assist consumers, the format and layout of the mailed survey differed from that of the telephone interview schedule. The mailed survey also contained three additional questions. These additional questions aided clarity, sought

further information and aided the translation of the telephone interviews into the less interactive medium of the mailed survey. One repetitive question was also dropped from the mailed survey.

The telephone interview schedule contained a total of 47 questions related to performance against the standards. The mailed questionnaire contained 49 performance-related questions. These questions are listed in Chapter 5. A copy of the questionnaire used in the mailed survey is included in a supplementary CD (*Developing Quality Measures for Home and Community Care: Technical Appendixes*).

The HACC Officials Standards Working Group member in each State and Territory determined the method of collecting consumer feedback in his or her own jurisdiction. Potentially, four methodologies for using the questionnaire were available for pilot testing – focus groups, individual face-to-face interviews, telephone interviews and mailed questionnaires. Only telephone interviews and mailed questionnaires were tested during the project.

2.2.3 Sample

Telephone interviews

Telephone interviews were conducted with HACC consumers in South Australia only. The Government of this State employed a research agency to undertake interviews with consumers. The data obtained from these interviews were then provided to the project team. Of the 18 agencies in this State that had participated in the pilot test of the Instrument, five volunteer agencies were sought, and obtained, to undertake a trial of the consumer survey.

Consumers were randomly selected from the participating agencies. Interview data were obtained from a total of 75 consumers from five different agencies, representing a response rate of 94%.

Mailed surveys

Mailed consumer surveys were trialled in Queensland, Western Australia, New South Wales, Victoria and the Australian Capital Territory. Five agencies were to be selected in each State. These five agencies were to be chosen to satisfy two general criteria. Where different assessment methods were used within the State, consumer feedback was to be obtained from at least one agency using each different type of assessment method. Where possible, agencies selected for consumer feedback were to represent a range of service types. Using these criteria, the government officials responsible for coordinating the pilot in each State selected the agencies to participate.

Agencies were asked to randomly select 50 consumers by selecting every '*n*th' consumer from their list of current consumers, calculating '*n*' by dividing the total number of consumers by 50 and rounding to a whole number. For agencies with fewer than 50 clients, all consumers were to be selected.

The overall response rate for the mailout survey was 34%. The following table shows the number of agencies which participated in pilot testing the consumer survey by mailed questionnaire in each jurisdiction, and the number of responses obtained.

Table 2.7: Response rate to the mailed survey

	NSW	Vic	Qld	WA	ACT	Total
Number of agencies	10	2	5*	3	6	26
Number of forms sent out	481	100	230	150	300	1261
Number of responses	117	47	90	69	108	431

* No responses received from clients of one of these agencies, a transport service.

2.2.4 Procedure

Telephone interviews

In South Australia a subcontracted research company, sponsored by the South Australian Government, conducted telephone interviews. Selected consumers were first contacted by the agency from which they were receiving services. Agreement was sought from each selected consumer to being questioned about the quality of the HACC services that they were receiving by an independent telephone interviewer. If the consumer agreed, his or her telephone number was supplied to the research company. Interviewers recorded consumer responses on the questionnaires provided by the project team. The research company then entered these responses into a data file and forwarded them to the Institute.

Mailed questionnaires

Queensland, Western Australia, New South Wales, Victoria and the Australian Capital Territory chose to trial the mailed questionnaire, on the basis that this method was the only one likely to be possible within resource constraints in any future full-scale implementation of the National Service Standards appraisal process. The government officials responsible for coordinating the pilot in each State contacted the selected agencies and sought their agreement to participate. The contact details of these agencies were then passed on to the project team.

The project team forwarded to each of the participating agencies 50 packages for consumers. These packages contained a reply paid envelope (addressed to the Institute), a survey form, and a covering letter explaining the purpose of the pilot, the voluntary nature of the survey, and contact details of the project team. These packages were then sent out by the agencies to the consumers in the randomly selected sample.

Consumers returned their anonymous forms direct to the Institute. On receipt of these forms, the project team undertook quantitative and qualitative analysis of responses.

2.2.5 Testing for validity and reliability

Representativeness

The reliability and validity of a method for collecting consumer feedback is dependent on whether it facilitates the involvement of all HACC consumers. In other words, respondents should be representative of the total HACC population. A particular

method should not put any group of consumers at a disadvantage in providing their feedback. The representativeness of the sample obtained was assessed in two ways: by comparison of the profiles of survey respondents to those of the total HACC population; and by calculation of the overall response rate to the two collection methods.

Item response

The usefulness of the consumer survey as a tool for gaining performance information is limited by the degree to which respondents reply to the questions asked of them. High rates of non-response may indicate that the questions are inappropriate to particular respondents, or it may be a symptom of the lack of clarity in the questions asked. The validity of the consumer survey to its target population was thus tested by examination of the response rate to particular items.

Comparison of the item response rate across survey methods allowed examination of the relative validity of each method. Where survey items are themselves ineffective, poor responses will be noted regardless of the method employed. Where the method of collecting consumer feedback was ineffective, non-responses would be more frequent even for questions that may otherwise have effectively elicited responses.

Concurrent validity

The consumer feedback Instruments were devised as measures of agency performance from the viewpoint of the consumer. The concurrent validity of both the consumer surveys and the Instrument is supported when all measures converge on the same performance appraisals for agencies. Concurrent validity indicates that the assessment tools are measuring the same thing, in this case, agency service quality. Examination of correlations between Instrument ratings and consumer appraisals indicate the relative concurrent validity of the respective measures. Consumer feedback from both mailed surveys and telephone interviews was compared with agency self-ratings and assessors ratings against the Instrument.

2.3 Comparison of quality appraisal mechanisms

Some Home and Community Care agencies have opted or been required to undertake quality appraisal processes, such as Disability Services Standards reviews, or processes associated with accreditation programs such as the Community Health Accreditation and Standards Program (CHASP) or the Australian Council of Healthcare Standards (ACHS). It was therefore desirable to explore the similarities and differences between these processes to determine whether a HACC National Service Standards review of agencies that have already undergone review through one of these other processes would constitute unnecessary duplication and an inefficient use of resources.

A comparative content analysis of the standards contained in the review processes listed above was undertaken. This analysis has been separately published as:

- Butkus E 1997. Home and Community Care National Standards: Comparison with the Disability Service Standards, Community Health Accreditation and Standards Program, and the Australian Council of Healthcare Standards. Canberra: AIHW (Welfare Division Working Paper no. 14).

This material compares the content of the standards, but does not compare their processes of implementation. Section 6 of this report both summarises some of the key findings and outlines the difficulties that arise in attempting to extend a comparison of quality appraisal mechanisms beyond the content of the standards.