

3 Testing the HACCC National Service Standards Instrument

This chapter examines the validity and reliability of the HACCC National Service Standards Instrument as tested in the pilot project and proposes a method of scoring agencies against the standards. The analysis begins by presenting how agencies scored against each of the standards in the pilot test and which of the 27 standards most often received 'met', 'not met' and 'partly met' ratings.

Section 3.2 examines the face and content validity of the Instrument. It discusses the clarity, desirability, practicality and appropriateness of the standards and their associated performance information in the Instrument. This section draws on feedback from service providers (from the Service Provider Survey and comments written on agency Instruments themselves) and comments from assessors (from debriefing sessions and the Assessor Survey).

The section on internal consistency (Section 3.3) presents the findings of a range of statistical analyses of pilot test data. The results of these analyses indicated that certain changes to the Instrument would be likely to improve its validity and reliability, particularly in relation to scoring. The results of an analysis of pilot test data that incorporates recommended changes to the Instrument are also presented.

The validity of using an Instrument Score, calculated by adding individual ratings against standards, is supported by tests of its concurrent validity. Section 3.4 makes a test of the Instrument's concurrent validity by comparison with the overall appraisals of agency performance. These global measures of perceived quality of service were provided at the same time as the Instrument's ratings. Comparison is also made between agency self-ratings and assessor ratings.

Finally, a key aspect of the reliability of the Instrument concerns the consistency with which different raters would rate the same agency against the standards. Ideally, by using the guidelines and following the performance information contained in the Instrument, each standard would be rated the same when different raters assess the same agency. The results of inter-rater reliability assessments undertaken by consultancy team members for 25 agencies are presented in Section 3.5. These results contribute to an assessment of the adequacy of the Instrument used in the pilot test.

Box 3.1 The HACC National Service Standards

Objective 1: ACCESS TO SERVICES

- 1.1 *Assessment occurs for each consumer.*
- 1.2 *Consumers are allocated available resources according to prioritised need.*
- 1.3 *Access to services by consumers with special needs is decided on a non-discriminatory basis.*
- 1.4 *Consumers in receipt of other services are not discriminated against in receiving additional services.*
- 1.5 *Consumers who reapply for services are assessed with needs being prioritised.*

Objective 2: INFORMATION AND CONSULTATION

- 2.1 *Consumers are aware of their rights and responsibilities.*
- 2.2 *Consumers are aware of services available.*
- 2.3 *Consumers are informed of the basis of service provision, including changes that may occur.*

Objective 3: EFFICIENT AND EFFECTIVE MANAGEMENT

- 3.1 *Consumers receive appropriate services provided through the processes of ongoing planning, monitoring and evaluation of services.*
- 3.2 *Consumers receive services from agencies that adhere to accountable management practices.*
- 3.3 *Consumers receive services from appropriately skilled staff.*

Objective 4: COORDINATED, PLANNED AND RELIABLE SERVICE DELIVERY

- 4.1 *Each consumer receives ongoing assessment (formal and informal) that takes all support needs into account.*
- 4.2 *Each consumer has a service delivery/care plan which is tailored to individual need and outlines the service he or she can expect to receive.*
- 4.3 *Consumers' cultural needs are addressed.*
- 4.4 *The needs of consumers with dementia, memory loss and similar disorders are addressed.*
- 4.5 *Consumers receive services which include appropriate coordination and referral processes.*

Objective 5: PRIVACY, CONFIDENTIALITY AND ACCESS TO PERSONAL INFORMATION

- 5.1 *Consumers are informed of the privacy and confidentiality procedures and understand their rights in relation to these procedures.*
- 5.2 *Consumers have signed confidentiality release forms.*
- 5.3 *Consumers are able to gain access to their personal information.*

Objective 6: COMPLAINTS AND DISPUTES

- 6.1 *Consumers are aware of the complaints process.*
- 6.2 *Each consumer's complaint about a service, or access to a service, is dealt with fairly, promptly, confidentially and without retribution.*

- 6.3 *Services are modified as a result of 'upheld' complaints.*
- 6.4 *Each consumer receives assistance, if requested, to help with the resolution of conflict about a service that arises between the frail elderly person or younger person with a disability and his or her carer.*

Objective 7: ADVOCACY

- 7.1 *Each consumer has access to an advocate of his or her choice.*
- 7.2 *Consumers know of their rights to use an advocate.*
- 7.3 *Consumers know about advocacy services – where they are and how to use them.*
- 7.4 *The agency involves advocates with respect to representing the interests of the consumer.*

3.1 How agencies scored against the standards

This section looks at how agencies scored against the standards, using the performance information and the guidelines provided for the Instrument applied in the pilot test. A list of the 27 standards is provided in Box 3.1. The performance information required from agencies under each standard is part of the HACC National Service Standards Instrument as used in the pilot (a copy is included in appendix A of the supplementary report *Developing Quality Measures for Home and Community Care: Technical Appendixes*). Agencies received specific instructions on how to complete the performance information within the Instrument while assessors (whether government officers or peers) were given additional assistance to help with scoring agencies against the standards in the form of the assessor guidelines.

Agencies were asked to give themselves a rating against each standard and to write this rating on their copy of the Instrument. The assessor for the agency was asked to mark his or her ratings on a separate Ratings Summary Form. Figures 3.1, 3.2 and 3.3 graphically show these ratings for the total sample of Instruments and Ratings Summary Forms received. Figure 3.1 shows the proportion of agencies that believed that their agency met each standard and the proportion of agencies that assessors believed met each standard. Figure 3.2 shows the proportion of agencies that believed that their agency partly met each standard and the proportion of agencies that assessors believed partly met each standard, and Figure 3.3 shows proportions from both these sources giving a 'not met' rating.

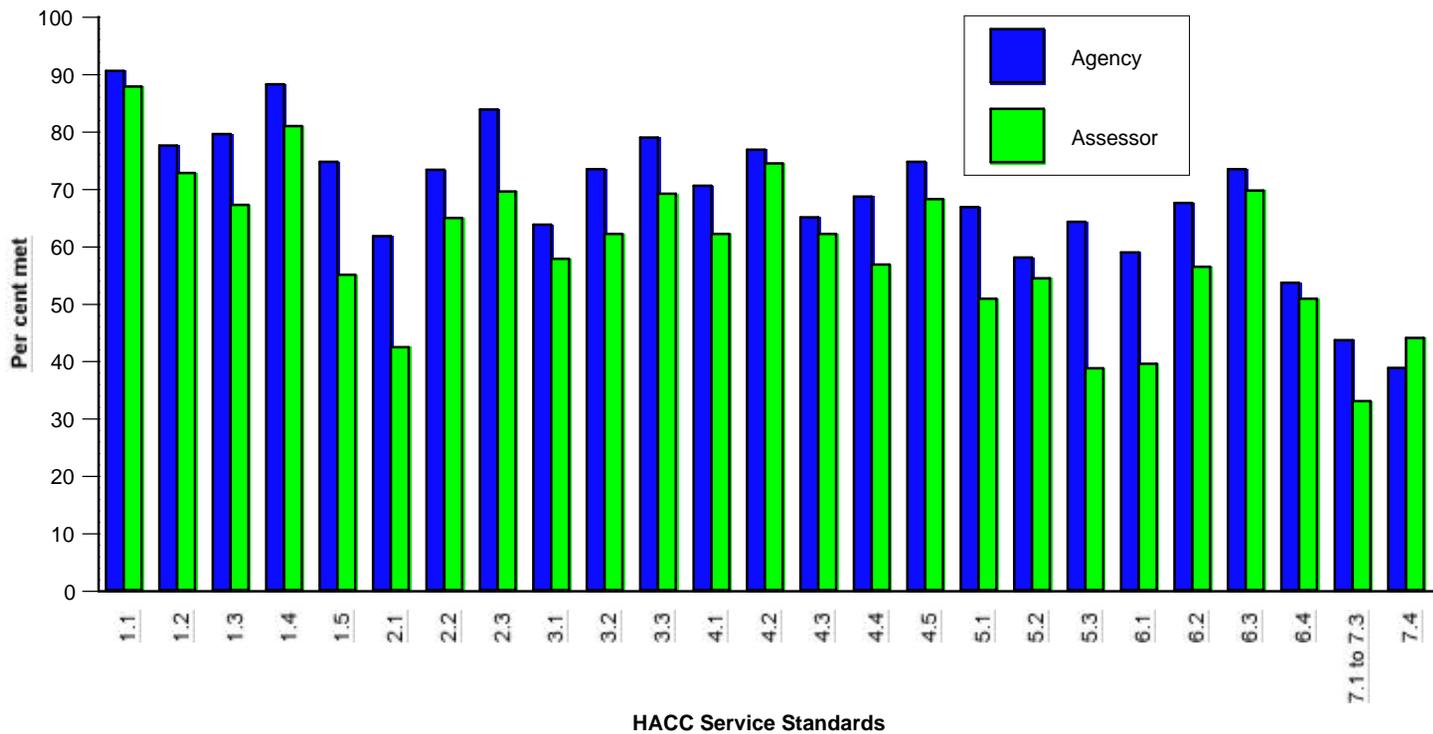


Figure 3.1: Percentage of agencies given a 'met' rating by the agency representative and by assessing person

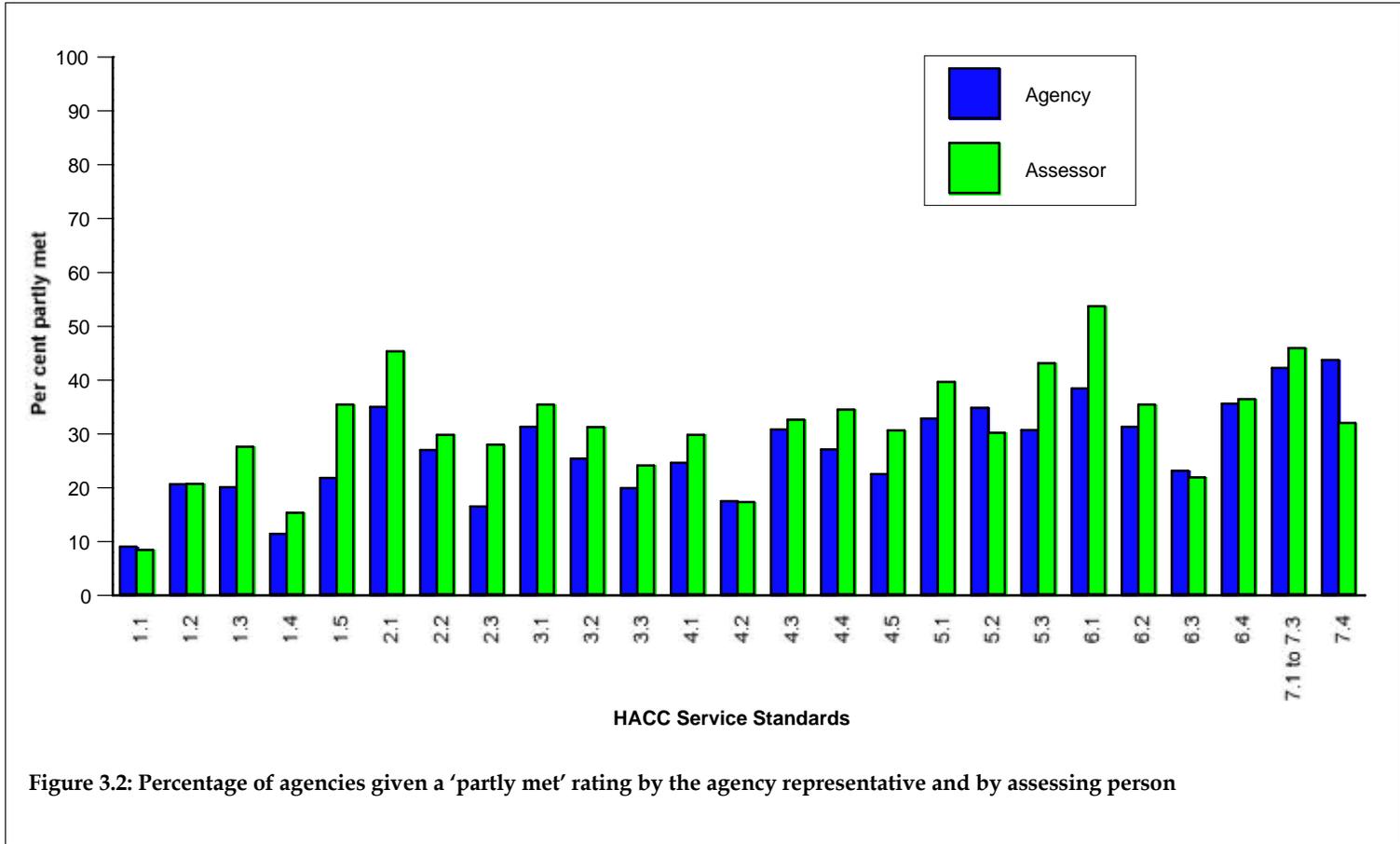


Figure 3.2: Percentage of agencies given a 'partly met' rating by the agency representative and by assessing person

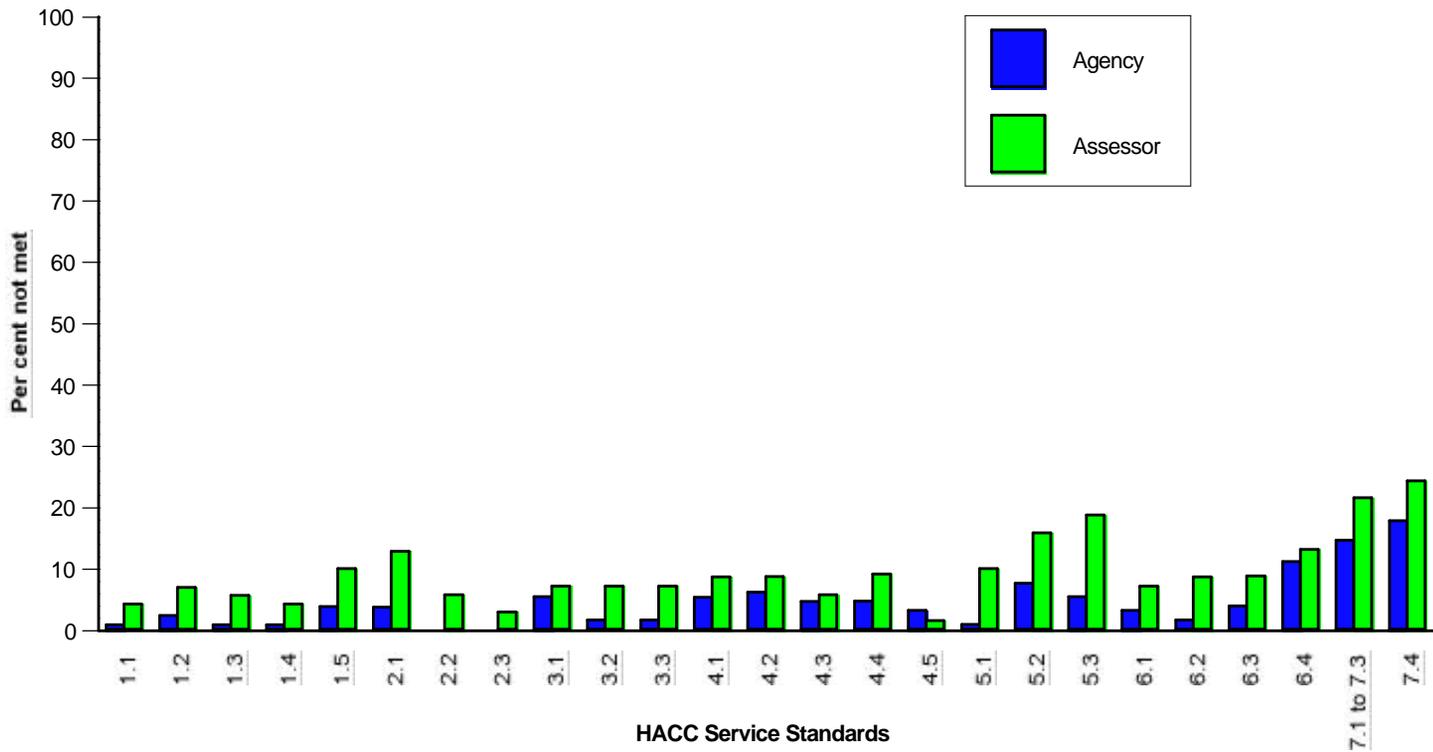


Figure 3.3: Percentage of agencies given a 'not met' rating by the agency representative and by assessing person

3.1.1 Standards against which agencies did not perform well

The lowest proportions of 'met' ratings occurred for the standards listed under Objective 7 (Advocacy). Eighteen per cent of agencies believed that they did not meet the standard regarding the involvement of advocates (7.4), and 15% believed they did not meet the standards regarding consumer information (7.3), consumer knowledge of rights (7.2), and consumer access to an advocate (7.1).

According to assessor ratings, a quarter (24%) of agencies did not meet the standard regarding the involvement of advocates and 21% did not meet the standards regarding consumer information, consumer knowledge of rights, and consumer access to an advocate.

According to agency responses on the Instrument, the reasons for the low rate of 'met' ratings included the lack of policy and standard procedure regarding advocacy, a lack of information available to consumers and a lack of staff training. Rural and remote agencies reported having limited access to formal advocacy services and limited access to information about advocacy. Formal advocacy services specifically for older people were reported to be unavailable in some areas. Service type was also raised as a factor. For a number of agencies providing transport, giving this information to all consumers in the same way as other HACC agencies was considered too costly.

The next lowest proportions of 'met' ratings, according to both assessors and agencies, occurred under Objective 6 (Complaints and Disputes). According to assessor ratings, the lowest proportion of 'met' ratings was recorded for Standard 6.1 (Consumers are aware of the complaints process). Only 39% of agencies were given a 'met' rating for this standard. The most common reason for agencies failing to obtain a 'met' rating was that the explanation to consumers of the complaints process was often only verbal and not written. The guidelines for assessors specified that both should occur to warrant a 'met' rating. Other reasons for agencies not meeting this standard were that they failed to provide reminders, failed to provide brochures in languages other than English, or were unable to provide an adequate policy or description of procedure on the matter.

According to agency ratings, the next lowest proportion of 'met' ratings was recorded for Standard 6.4 (Each consumer receives assistance, if requested, to help with the resolution of a conflict about a service that arises between the frail elderly person or younger person with a disability and his or her carer). (See Section 3.6 for further discussion.) Only 54% of agencies gave themselves a 'met' rating for this standard. Many agencies expressed the belief that it was not their role to become involved in such conflict resolution. Eleven per cent of agencies believed they had not met the standard. Agencies which believed they had partly met the standard indicated that they would attempt some level of resolution followed by referral if there was no successful resolution. These procedures were often not formalised in policy or described and passed on to staff as standard practice in any way. As such, agencies would not agree that the standard was 'not met' but neither would they propose that they had fully addressed the issue.

Apart from the objectives grouped under Objectives 6 and 7, there were three further individual standards against which agencies did not perform particularly well. These were Standard 5.3 (Consumers are able to gain access to their personal information),

Standard 2.1 (Consumers are aware of their rights and responsibilities), and Standard 5.2 (Consumers have signed confidentiality release forms). Nineteen per cent of agencies were scored as 'not met' against Standard 5.3. The most common failure reported by agencies was that no formal advice was given to consumers regarding access to personal information or agency responsibilities. Agencies also reported a lack of appropriate policies, failure to inform consumers of what was kept, or failure to have a policy regarding who was appropriate to access it.

For Standard 2.1 (Consumers are aware of their rights and responsibilities), assessors rated 13% of agencies as 'not met'. This relatively high proportion was due to many agencies failing to make provision for people from non-English-speaking backgrounds to understand their rights and responsibilities. The guidelines to assessors required that this factor be satisfied in order to avoid a 'not met' rating.

The pilot test of the Instrument revealed some more general difficulties with appropriately measuring agency performance in relation to special needs groups, and particularly people from a non-English-speaking background. (See Section 2.6 for further discussion and some suggestions for overcoming the apparent undue influence of an agency's failure to meet Standard 2.1 on their scores for other standards.)

Greater clarity about the HACC program's policy on gaining consumers' permission to release personal information would assist with redrafting the guidelines for Standard 5.2 (Consumers have signed confidentiality release forms). Standard 5.2 had a relatively high level of 'not met' ratings given to agencies by assessors (16%). These ratings were largely the result of agencies not obtaining written approval from consumers for the release of information. The guidelines were not sufficiently clear on what would constitute a 'not met' rating against this standard. Hence, some assessors gave agencies that did not obtain written approval a 'partly met'. Clarification of the extent to which a consent form signed by the consumer at one point in time (say, at referral as recorded on the CIARR, or Client Information, Assessment and Referral Record) is sufficient to cover all instances of subsequent information sharing is required.

3.1.2 Standards against which the agencies performed well

The highest proportion of 'met' ratings given by assessors occurred for Standards 1.1 (88%), 1.4 (81%), 4.2 (74%) and 6.3 (70%). The reasons for agencies being given a 'met' rating by assessors were not documented on the Ratings Summary Forms. By viewing agency responses and following the guidelines, it is reasonable, however, to conjecture why this might be so. In some cases, the absence of clear definitions for key elements of standards may have resulted in very high scores. For example, without a definition of what constitutes an 'assessment' or a 'care plan', agencies may have been able to confidently report very high compliance with such standards. The requirements of Standards 1.1 and 4.2 are such that it is relatively easy for agencies to provide tangible evidence.

Standard 1.1 (Assessment occurs for each consumer) requires that agencies calculate the proportion of consumers formally assessed and account for those not assessed. However, no information was given on what constitutes a 'formal' assessment or, in

fact, which type of assessment is the subject of the standard – for example, it may refer to assessment for service eligibility or assessment to determine functional dependency or to service specific assessments of need. This performance information was frequently satisfied with responses of, or close to 100%. However, the more difficult questions regarding the agency’s assessment method followed under Standard 1.2 (Consumers are allocated available resources according to prioritised need). This second standard received a higher proportion of ‘partly met’ ratings (21%) and a lower proportion of ‘met’ ratings (73%) according to assessors.

Similarly, Standard 4.2 (Each consumer has a service delivery/care plan that is tailored to individual need and outlines the service he or she can expect to receive) requires that agencies indicate the proportion of consumers with a service delivery/care plan. Again, for most agencies the response was simply 100%. This standard also required detail of how staff and clients were informed of this plan, which may have accounted for the somewhat lower proportions of ‘met’ ratings than Standard 1.1 (74% compared with 88%).

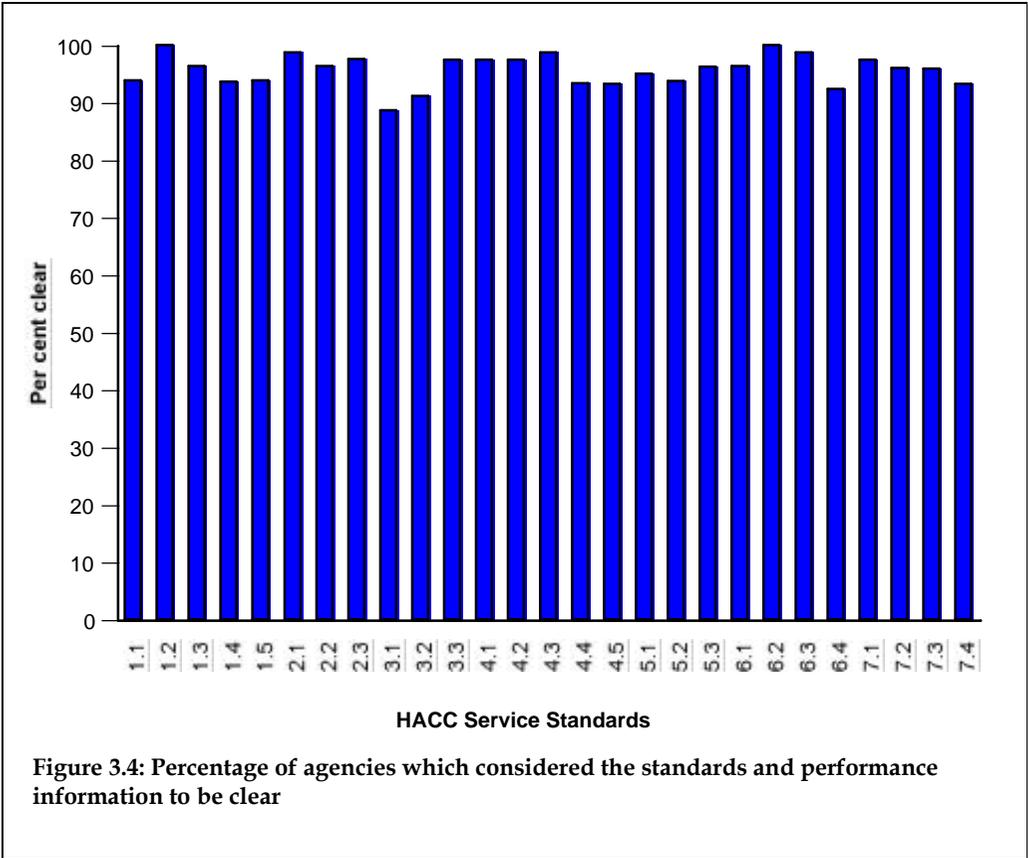
Standard 6.3 (Services are modified as a result of ‘upheld’ complaints) required agencies to document an instance in which this occurred. This standard may have been difficult to assess accurately without directly consulting consumers. The assessor would have been entirely dependent on the agency’s description of successful examples of upholding complaints. The Instrument alone would provide no evidence of complaints that were unfairly not upheld.

Standard 1.4 (Consumers in receipt of other services are not discriminated against in receiving additional services) required the agency to describe how it ensured that this did not happen. Agency responses to this question generally fell into two categories: (1) the types of services provided by other agencies were not of relevance to the service they provided and were therefore not taken into account; or (2) they were able to communicate with other agencies to provide a coordinated service based on assessed client need. In both cases, it could reasonably be claimed that unfair discrimination was not occurring. Again, this may have been another standard where it was difficult for assessors to determine if agencies were discriminating unfairly without asking consumers and assessing individual cases.

3.2 Face and content validity

3.2.1 Clarity of the Instrument

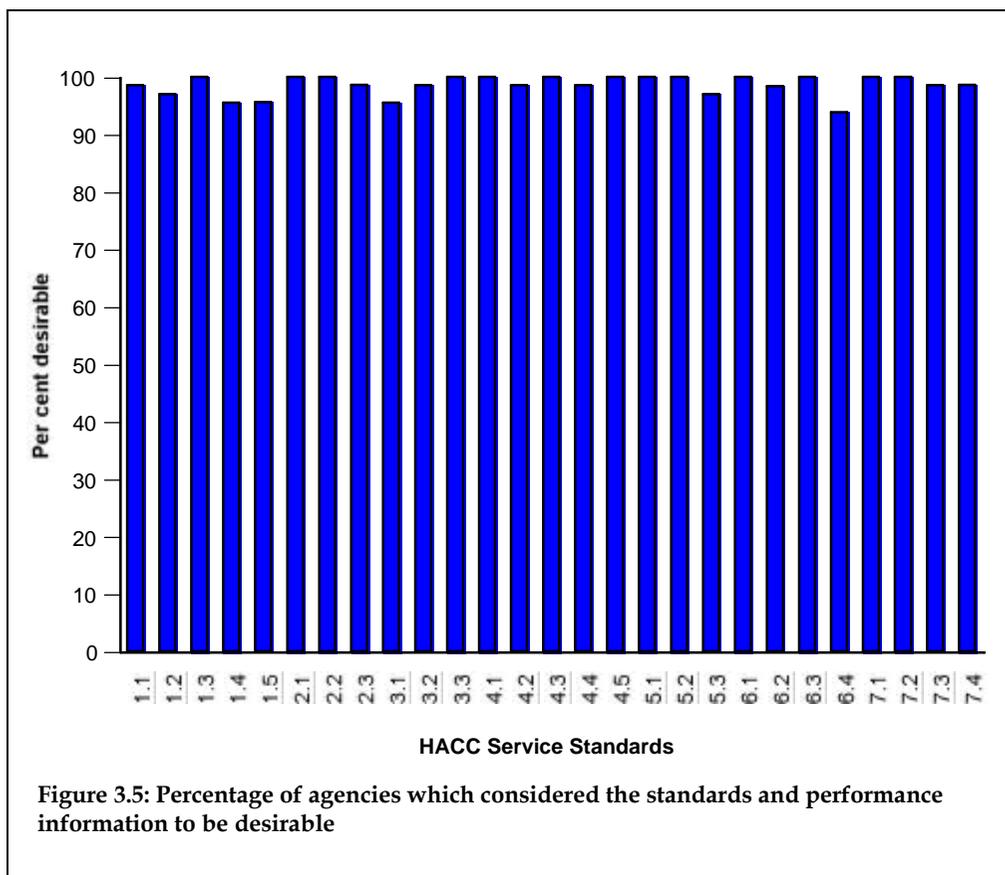
Service provider feedback on the clarity of the Instrument was gained through the Service Provider Survey. Service providers were asked, ‘Are there any standards or performance information items that were unclear?’. Figure 3.4 indicates that the majority of agencies believed the standards and their performance information as described in the Instrument to be clear. Averaging over the 27 standards, 95% of service providers agreed they were clear.



For 17 standards at least 95% of service providers agreed that standards were clear. A further nine standards were considered clear by between 90% and 95% of service providers. Standards considered least clear were those concerning efficient and effective management. Standard 3.1 (Consumers receive appropriate services provided through the processes of ongoing planning, monitoring, and evaluation of services) was considered unclear by 11% of service providers. Standard 3.2 (Consumers receive services from agencies that adhere to accountable management practices) was considered unclear by 9% of service providers. Standard 3.1 in particular requires of agencies a relatively high level of organisational acumen and resources, including planning, monitoring and evaluation of services in response to community needs at a higher level. Such planning and service development is difficult to achieve among agencies with limited resources and those still struggling to establish service delivery. Three agencies in rural or remote settings were unclear on how to address this standard. As one remote area assessor put it: 'Often if one can get a service going in an area it is a remarkable achievement'.

3.2.2 Desirability of the Instrument

The Service Provider Survey asked ‘Are there any standards or performance information items that were undesirable?’. Figure 3.5 indicates that the majority of agencies believed the standards and their performance information as described in the Instrument to be desirable. Averaging over the 27 standards, 99% of service providers agreed they were desirable.



Thirteen standards were considered to be desirable by all service providers surveyed. A further eight standards were considered undesirable by only one service provider.

The standard least likely to be considered desirable was Standard 6.4 (Each consumer receives assistance, if requested, to help with the resolution of a conflict about a service that arises between the frail elderly person or younger person with a disability and his or her carer). Six per cent of service providers found this standard to be undesirable. These agencies included a home nursing service, a respite service, a home visiting and dementia support service, and a multi-service agency offering home visiting, transport, shopping, home maintenance and cleaning.

Standards 1.4 (Consumers in receipt of other services are not discriminated against in receiving additional services), 1.5 (Consumers who reapply for services are assessed with needs being prioritised) and 3.1 (Consumers receive appropriate services provided through the processes of ongoing planning, monitoring and evaluation of services) were considered undesirable by three service providers. Standard 1.4 was believed to be undesirable because of the need to discriminate against consumers already receiving a service, particularly when the service is scarce and required by others of greater need. Standard 1.5 was seen as undesirable by agencies for whom it was not appropriate to take into account previous refusals when prioritising need. It was also not considered to be a desirable standard for an agency that never refused service to consumers. Standard 3.1 was not seen as desirable by service providers who believed that it was beyond their charter and capabilities to undertake monitoring of community need. The lack of need to refuse service was also cited as a reason for regarding Standard 1.2 as undesirable. A respite agency and a transport service believed that prioritising consumer need was unnecessarily time- and resource-consuming when all requests could be adequately met.

3.2.3 Practicality of the Instrument

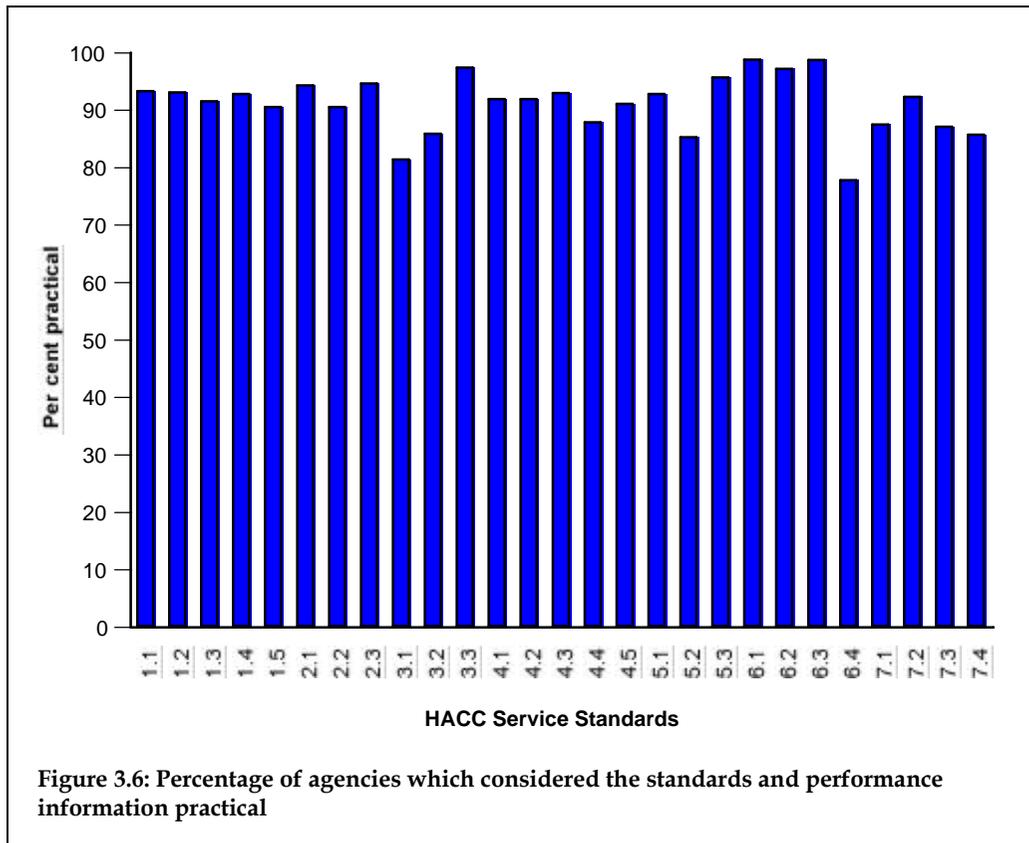
If there was any area in which the standards were seen to be relatively lacking it was in the area of practicality. The Service Provider Survey asked, 'Are there any standards or performance information items that were impractical?'. On average, 91% of service providers agreed that the standards were practical. Figure 3.6 illustrates service providers' beliefs about the practicality of standards and performance information as described in the Instrument.

Only five standards were rated as practical by at least 95% of service providers, with a further 20 being rated as practical by between 85% and 95%. The standards rated as practical by less than 85% of service providers were Standard 6.4 (Each consumer receives assistance, if requested, to help with the resolution of a conflict about a service that arises between the frail elderly person or younger person with a disability and his or her carer) and Standard 3.1 (Consumers receive appropriate services provided through the processes of ongoing planning, monitoring, and evaluation of services).

Twenty-two per cent of service providers surveyed believed Standard 6.4 was impractical. Most commonly this was because agencies saw their role as meeting a specific defined need of consumers, not as determining the appropriate balance of resolved needs between carers and care recipients. To undertake such a role was perceived to be more in the line of case management or advocacy. Nineteen per cent of service providers believed that Standard 3.1 was not practical. Agencies operating out of larger organisations stated that it was difficult to separate planning for one team service from that of an entire organisation, particularly when routine planning occurred at the broader level. It was also considered impractical in so far as agencies felt there was insufficient funding to spend time doing the kind of analysis requested by this standard.

The objective that was most frequently criticised as being impractical was Objective 7 (Advocacy). On average, the standards under Objective 7 were considered to be impractical by 12% of agencies. The difficulties in the practicality of these standards was largely due to the lack of available advocacy services in the agency's area and the tendency for consumers to ask the agency to act as an advocate whether or not it was a

formally recognised function of the agency. Service providers argued that it made more sense to rely on advocacy services to supply the agency with information about their own services, although this was not always forthcoming. It was also considered impractical to supply consumers with information about advocacy at first contact since so many other pieces of information were being exchanged at this time. Some agencies considered it more practical to give this information to consumers as the need arose.



3.2.4 Appropriateness of the Instrument

The Service Provider Survey asked ‘Are there any standards or performance information items that were inappropriate?’. The appropriateness of the standards and their performance information was primarily determined by the type of service provided by the agency completing the Instrument. Other agency characteristics which affected the perceived appropriateness of the Instrument included the type of clients accessing the service, the geographic location of the agency, the structure and maturity of the organisation, the economic and social environment surrounding the agency and the available service provider networks.

Figure 3.7 shows that the majority of agencies considered the standards appropriate; averaging over standards, 94% of agencies believed the standards to be appropriate.

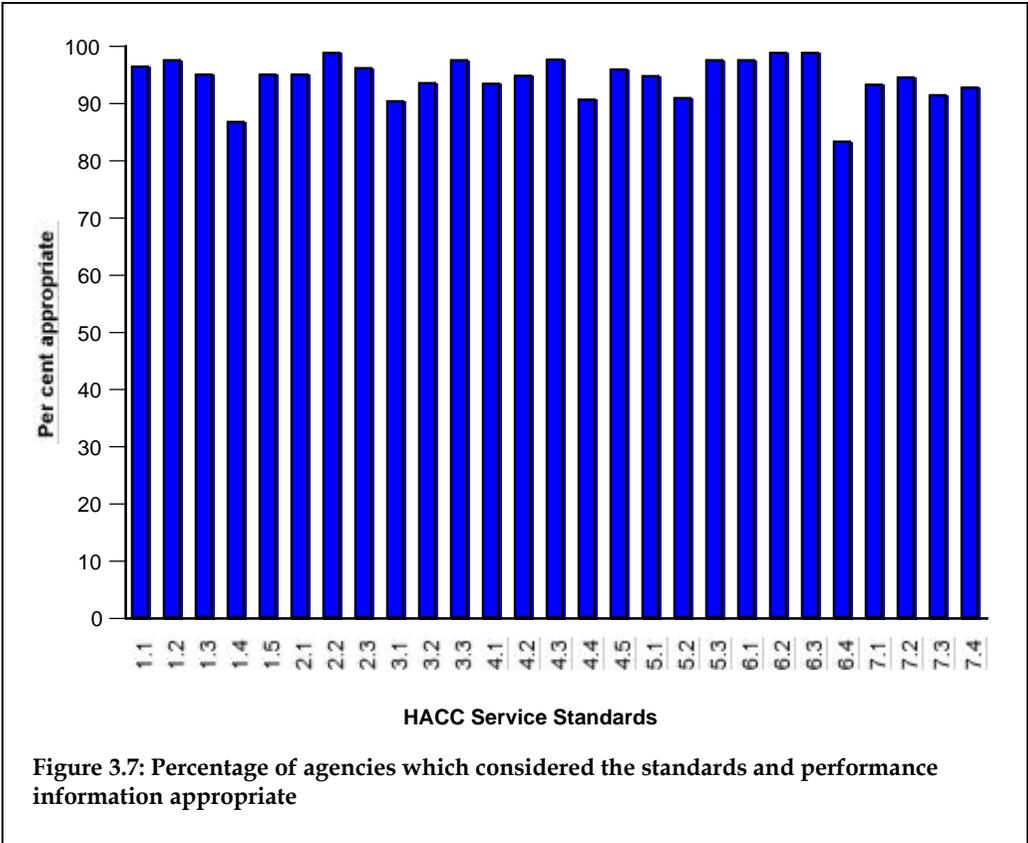
For 25 standards at least, 90% of agencies regarded them to be appropriate. Two standards were considered appropriate by less than 90% of service providers.

Standard 6.4 (Each consumer receives assistance, if requested, to help with the resolution of a conflict about a service that arises between the frail elderly person or younger person with a disability and his or her carer) was considered appropriate by only 83% of service providers. These service providers generally considered that it was not their role to become involved in such disputes. Standard 1.4 (Consumers in receipt of other services are not discriminated against in receiving additional services) was not considered appropriate by 14% of service providers because the issue of discrimination was considered irrelevant. Agencies either did not think it relevant to take into account any other services provided when determining their own service provision or, alternatively, services were provided according to a case coordination model in which discrimination was appropriate.

Two other standards considered inappropriate across a range of service types and circumstances were Standard 3.1 and Standard 5.2. The performance information requested under Standard 3.1 (Consumers receive appropriate services provided through the processes of ongoing planning, monitoring and evaluation of services) was not considered appropriate to 10% of agencies. A question under this standard sought information about how the agency monitored community need for services. Some agencies believed that it was difficult for individual service providers to accurately assess community need, particularly unmet need. In addition, some agencies believed they were unable to alter service delivery in response to such assessments because of their particular funding agreements. Nine per cent of service providers believed Standard 5.2 (Consumers have signed confidentiality release forms) was inappropriate. This requirement was not seen to be appropriate for some agencies, particularly those providing services where little or no information about clients is recorded. Other objections to the performance information required were that consumers would find such forms difficult to follow. Some service providers believed that it was not always possible to know to whom information would need to be released.

On average, 7% of agencies did not believe that the standards under Objective 7 (Advocacy) were appropriate to them. Some services did not believe it was their role to advise clients of the range of advocacy services available to them. They argued it was inappropriate to do this because clients should obtain their advocates from outside the organisation. It was also felt that to overemphasise a consumer's need for an advocate may implicitly disempower the client.

The following discussion examines the particular types of agencies for whom the appropriateness of standards and performance information was problematic. It identifies agencies for whom either exemption or more specific interpretation of the Instrument may be required. The effect of type of assessment on the perceived appropriateness of the Instrument is also discussed.



Types of service

Community nursing, allied health

Of the agencies responding to the Service Provider Survey, the agencies which most frequently found that standards were inappropriate were those providing community nursing or allied health services: 56% of these agencies found at least one standard inappropriate. The standard of most frequent concern for this agency type was 4.4 (The needs of consumers with dementia, memory loss and similar disorders are addressed). Agencies that reported this standard as inappropriate argued that consumers were not differentiated according to the presence or absence of dementia or related disorders, and that clients suffering such conditions were treated as any other client with their care modified accordingly. One agency argued that the standard was not specific enough for a nursing service since it did not take into account the physical aspects of their care or issues concerning safety within the environment.

For many community nursing agencies, services to clients were provided for a defined period based on the assessment and referral of another body such as a hospital or general practitioner. This episodic nature of service provision meant that Standard 4.1 (Each consumer receives ongoing assessment (formal and informal) that takes all support needs into account) was considered not applicable. Similarly, Standard 1.5 (Consumers who reapply for services are assessed with needs being prioritised) was

not considered appropriate by these agencies since clients, managed by another agency, came to them through referrals for a defined period. Moreover, because care was provided to address a current problem and not an ongoing one, Standard 2.3 (Consumers are informed of the basis of service provision, including changes that may occur) was not seen to be appropriate from the perspective that the service was not altered within the episode of care.

Transport

Standards considered particularly inappropriate for transport agencies included 1.2, 1.5 and 4.2. Regarding Standard 1.2, agencies argued that where all consumer need could be met it was inappropriate, unnecessarily intrusive and time consuming to assess and prioritise consumer need. Regarding Standard 1.5, an agency providing occasional one-off services to a pool of previously assessed consumers did not believe it appropriate to reassess each time these consumers called for assistance. In relation to Standard 4.2, the nature of transport services frequently meant that care plans were not appropriate.

Meals

Agencies that provide meals did not feel it was appropriate for them to take into account the other services received by a consumer (Standard 1.4: Consumers in receipt of other services are not discriminated against in receiving additional services). Given the role of these agencies in the service provider network, meals agencies did not feel it was appropriate for them to ensure that their clients were representative of the ethnic groups within the community by constructing and monitoring community profiles and client profiles (Standard 1.3: Access to services by consumers with special needs is decided on a non-discriminatory basis). Similar views were expressed with regard to Standard 3.2 (Consumers receive services from agencies that adhere to accountable management practices). Due to the limited record holdings of this type of agency, it was considered inappropriate to enter into explanations with clients about privacy, confidentiality and client rights in relation to these (Standard 5.1: Consumers are informed of the privacy and confidentiality procedures and understand their rights in relation to these procedures). They did not consider that a referral role was appropriate, other than accepting those sent to them (Standard 4.5: Consumers receive services which include appropriate coordination and referral processes). Care plans or service delivery plans were not always considered appropriate (Standard 4.2: Each consumer has a service delivery/care plan which is tailored to individual need and outlines the service he or she can expect to receive). Prioritising need was not seen as appropriate, as it was considered more efficient to service all who meet HACC program eligibility criteria without intrusive and time-consuming assessment processes (Standard 1.2: Consumers are allocated available resources according to prioritised need).

Home maintenance and modification

Home maintenance and modification agencies had difficulty with the appropriateness of the Instrument because the services provided by them tended to be more of the nature of 'one-off' services or 'on demand' services. They also felt that the seasonal nature of their work was less well addressed by the Instrument. Standards that were not always considered appropriate for home maintenance and modification consumers were 1.1 and 4.1, since formal assessment and reassessment were not generally appropriate activities for home maintenance/modification staff. Similarly,

regarding Standard 4.2, clients did not receive a care plan, but rather a document stating what modification or maintenance was performed. For 'one-off' jobs these agencies did not generally distribute or explain all of the HACC consumer rights and responsibilities as they did not feel this was appropriate given the nature of the service (Standard 2.1). Standards 4.3 and 4.4 refer to the special needs of consumers with regard to dementia and memory loss and with regard to cultural needs. Home maintenance and modification agencies argued that the nature of their services was such that it was not appropriate to provide services to these groups any differently than to other consumers.

Services for specific types of client

Services for young people with a disability

Standard 4.4 (The needs of consumers with dementia, memory loss and similar disorders are addressed) was not considered appropriate to services for younger people with a disability, especially by those services caring for clients who are mainly children. As such, dementia was not seen as an issue for these services, although behavioural difficulties may be.

Agencies also argued that the disability or illiteracy of some consumers may mean that it is not appropriate to supply them with brochures on rights and responsibilities. This characteristic of clients was relevant to Standard 5.1 (Consumers are informed of the privacy and confidentiality procedures and understand their rights and responsibilities in relation to these procedures). Six per cent of agencies did not believe that this standard was appropriate. Similarly, for Standard 2.3 (Consumers are informed of the basis of service provision, including changes that may occur) 6% of agencies believed the standard to be inappropriate, arguing that consumers may not always understand the basis of the decision, despite repeated explanations.

Dementia-specific services

Informing clients with dementia of the issues covered in Standards 2.1, 2.2, 2.3, 5.1, 6.1, 7.2 and 7.3 was considered to be not always practical. In some cases, clients were considered unable to understand the information being given to them. In others, agencies believed that clients may be feeling overloaded with information, or they may be stressed or else too ill to deal with interpreting the information.

Providing information to and requiring signatures from clients affected by dementia was not seen as appropriate (Standards 4.5, 5.2 and 5.3). From the total sample of agencies surveyed, 7% of service providers believed it was inappropriate to ask their clients to be involved in service management (Standard 3.2: Consumers receive services from agencies that adhere to accountable management practices). This was largely because they believed they were either 'not interested' or 'don't understand'. Services caring for those with dementia were represented in this group.

Standard 4.4 was not considered detailed enough by allied health and nursing agencies to address the issues of dementia as it affected their service provision (see discussion on previous page).

Services for people from a non-English-speaking background

Standard 1.3 concerning access to services by people of non-English-speaking backgrounds was not considered to provide a relevant measure of quality for services specifically targeted to these groups. It was felt that a more appropriate measure was required to indicate how agencies facilitate access for these groups.

A number of agencies working in communities with few people or none from a non-English-speaking background argued that it was not appropriate to be penalised for failing to provide information for potential members of these groups (Standards 2.1, 2.2, and 6.1). It was also argued that written material was sometimes not appropriate as older people from a non-English-speaking background may not be literate in their first language.

Services for Aboriginal and Torres Strait Islander peoples

Agencies that provide services to consumers who are of Aboriginal and Torres Strait Islander descent can face particular challenges in meeting service delivery standards. The transitory nature of service use among these consumers in some regions, high dependency, and levels of illiteracy are among some of the characteristics that require a different approach to assessment against the standards. The demand for 'one-off' and emergency services operated against agency ability to regularly conduct assessments (Standard 1.1) and keep care plans (Standard 4.2). Illiteracy among clients meant that Standard 5.2 regarding signed confidentiality release forms was not meaningful, and the right to access information (Standard 5.3) was not often seen as an issue for these consumers.

While many Aboriginal and Torres Strait Islander-specific services are appropriately assessed using the current Instrument, it is recommended that an Instrument be developed which specifically addresses the characteristics of this group, particularly in more remote regions where Indigenous culture more strongly affects service provision. The use of resources such as the *Optional Standards of Care for Frail Aged Aboriginal People* (Harrison 1995) may assist in this endeavour.

The challenges faced by services to Aboriginal and Torres Strait Islanders are often compounded by those faced by agencies in remote locations. These difficulties are discussed below.

Rural and remote agencies

In very remote regions the Instrument was found to be of limited applicability. In these areas the more basic requirement was simply to keep services operating. Many of these services run under adverse conditions and time spent in administrative quality assurance tasks, which were considered to be of limited assistance to service delivery, were not considered useful. In these locations the decision to apply the Instrument may be based, in part, on the level of funding to the agency and its ability to invest time in quality assurance issues.

The isolation of clients, the remoteness from other services, the lack of availability of staff, limited sources of funding and adverse weather conditions are seen as some of the factors which affect the ability of these agencies to meet many of the standards. The ability of services to attract adequately skilled staff is limited in remote and rural locations, presenting difficulties for agencies to meet Standard 3.3. Similarly, the

continuing training and education of staff is limited by the agency's resources and access to trainers and appropriate training programs. The lack of staff resources to spend time on administrative tasks also affects these agencies' ability to meet Standard 3.1.

In small communities in remote and rural locations the processes by which information is distributed are different to those in metropolitan areas. Community knowledge of people and services assists with such standards as 2.2 (regarding information about services), 2.3 (regarding information about the basis of service provision and changes that may occur) and 3.3 (regarding the screening of staff and volunteers). Access to formal advocates may not be possible or desired by consumers of remote services, but friends, family and council members were more likely to undertake this role.

Standard 3.2 and its performance information presented additional difficulties for remote agencies. The performance information under this standard requested agencies to show how consumers are involved in service management. The remote environment, distance and the disabilities of clients were argued to act against active consumer involvement in service management. The isolation of clients was also considered to affect the agency's ability to conduct regular reassessments (Standard 4.1).

Type of assessment

The type of assessment also had an effect on the likelihood of agencies indicating that certain standards were not appropriate to them. Of the agencies that had undertaken a joint assessment, 100% indicated that every standard was appropriate to them. Of the agencies that had undertaken a self-assessment, 71% indicated that all of the standards were appropriate. Fifty-eight per cent of agencies who had undertaken a self-assessment with verification indicated that all of the standards were appropriate. Agencies that were assessed by peer review were most likely to indicate that not all the standards were appropriate to them. Only 47% of peer-reviewed agencies indicated that all of the standards were appropriate to their agency.

These results do not indicate a clear relationship between the characteristics of the assessment methods and the degree of perceived appropriateness of the standards. It can be concluded however, that the most collaborative method of assessment, joint assessment, results in the greatest acceptance of the standards by agencies as appropriate to their service.

3.3 Internal consistency

3.3.1 The interrelationship of standards and objectives

This section addresses the issue of whether the standards provide satisfactory information against the seven objectives and whether the seven objectives are sufficiently related to one another to justify adding scores across them. A range of possible uses for Instrument Scores, as well as appropriate methods for calculating scores to answer particular questions potentially of interest to different stakeholders in the appraisal process, are presented.

The first approach to the issue of whether the standards provide a valid and reliable indication of performance against their objectives, sufficient to justify their inclusion in a scale of measurement, is to ask agencies and assessors to comment on them. This was the subject matter of Section 3.2 in which service providers' opinions of the clarity, desirability, practicality and appropriateness of the standards were discussed.

A second approach is to analyse the reliability of the standards in terms of whether different raters interpret the standards in the same way. If different raters interpret standards and performance information differently, then it is impossible to be sure that standards are adequately providing information about their objective. This approach is covered under Section 3.5 (Rater reliability).

A third approach, to examine the interrelationships among standards within objectives and the interrelationships across objectives, is the subject matter of this section.

If all of the objectives are related to one another, such that an agency that performs well on one objective will tend to perform well on the others, then there is a case for adding scores.¹ In this case, all of the objectives combined produce a coherent measure of quality of service in HACCC agencies. However, if there were a one-to-one correspondence between performance on one objective and performance on another, then it would be necessary to measure only one objective; hence, adding scores across objectives would be meaningless. If agency performance on one objective is consistently unrelated to agency performance on others, then it may be more informative and fairer to report performance against this objective separately. For example, if agency performance on the Complaints and Disputes Objective is consistently unrelated to performance on all other objectives then it would be beneficial to report this objective separately.

Before objectives can be added together, it must be established that standards within objectives can be added together. To do this, each standard under the objective should be related to other standards within the objective, in such a way that it both informs and confirms our knowledge of the objective. Standards that represent a particular objective should have something in common with each other and have less in common with other standards representing other objectives. The following discussion examines the relationship between standards within objectives and, in the light of comments by agencies and assessors, addresses the question of whether each standard provides a satisfactory measure against its objective. The ratings of assessors are the data used for the quantitative analysis performed here, since these individuals were given guidelines and training to support the reliability and validity of assessment.

Objective 1: Access to services

To summarise comments by agencies and assessors on this objective: all of the standards under Objective 1 were considered clear by at least 94% of agencies and close to 100% considered these standards desirable.

1. Standards are rated in the following way: 0 (*not met*), 1 (*partly met*), 2 (*met*). Hence qualitative ratings are translated to numerical scores that can be added across standards.

Concerns were raised about the appropriateness of Standard 1.4 (Consumers in receipt of other services are not discriminated against in receiving additional services), with 14% of service providers not considering it to be appropriate. The focus of this standard appears contradictory to the current aim of HACCC to provide coordinated services with other agencies, that is, to provide services in a manner that discriminates appropriately according to need. The content of this standard was also mainly covered by Standards 1.2 and 1.5 ($r = 0.54$ and 0.52 respectively, see table 3.1). However, these correlations were not large enough to consider that Standard 1.4 provided no new information regarding service quality. Definitional improvement of the performance information under this standard may improve its contribution to the measurement of this objective.

Table 3.1: Correlation coefficients for standards under Objective 1

	1.1	1.2	1.3	1.4
1.2	0.58			
1.3	0.32	0.52		
1.4	0.30	0.54	0.48	
1.5	0.33	0.41	0.44	0.52

Standard 1.5 (Consumers who reapply for services are assessed with needs being prioritised) was seen as impractical by 10% of agencies and unclear by 6% of agencies. These agencies requested clearer guidelines on what was required of them. These were agencies for whom full assessments were conducted by another agency, for whom services were provided in single occasions of need, or for whom consumer eligibility is not likely to change (such as belonging or not belonging to the category of young person with a disability). An analysis of the correlations among the standards relating to Objective 1 show no evidence, however, that the issues associated with Standard 1.5 caused it to operate differently to the other standards or to bring up issues unrelated to the other standards.

Table 3.2: Rotated factor loadings for the HACC National Service Standards as measured by the Instrument

Standard	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
1.1	-0.03	0.04	0.46	0.37	0.71	0.00
1.2	0.45	0.30	0.29	0.10	0.47	-0.03
1.3	0.47	0.48	0.24	0.03	0.19	0.35
1.4	0.66	0.29	0.21	-0.03	0.15	0.23
1.5	0.48	0.18	0.40	0.21	0.22	0.21
2.1	0.36	0.14	0.25	0.55	0.10	0.44
2.2	0.19	0.83	0.16	0.28	0.11	-0.04
2.3	0.33	0.30	0.09	0.08	0.64	0.22
3.1	0.24	0.41	0.53	-0.08	0.04	0.29
3.2	0.14	0.64	0.37	-0.08	0.12	0.30
3.3	0.14	0.07	0.57	0.13	0.52	0.22
4.1	0.23	0.07	0.72	0.13	0.10	0.15
4.2	0.25	0.17	0.06	-0.03	0.70	0.26
4.3	0.22	0.86	0.16	0.06	0.09	0.13
4.4	0.21	0.64	-0.19	0.23	0.33	0.09
4.5	0.26	0.23	0.67	0.16	0.18	-0.02
5.1	0.49	0.14	0.02	0.62	0.21	0.27
5.2	-0.07	0.02	0.45	0.73	0.16	0.00
5.3	0.22	0.19	-0.03	0.76	0.01	0.08
6.1	0.58	0.34	0.28	0.34	0.02	0.27
6.2	0.79	0.17	0.12	0.19	0.20	0.13
6.3	0.81	0.13	0.15	0.15	0.14	0.05
6.4	0.15	0.45	0.17	0.30	0.32	0.34
7.1, 2, 3	0.16	0.08	0.13	0.31	0.14	0.82
7.4	0.17	0.19	0.10	0.00	0.20	0.82

The correlations among the standards of Objective 1 are all positive and, with the exception of Standard 1.1 (Assessment occurs for each consumer), indicate a high degree of cohesiveness. A factor analysis,² presented in Table 3.2, identified

2. A factor analysis was performed to verify that the associations between standards reflected their objectives. The 25 ratings were factor analysed using principal components factor analysis. A varimax rotation was applied to the factor solution for the purpose of uncovering subsets of standards that were highly related but relatively independent of other subsets.

Standard 1.1 as operating least well as a predictor of this objective. As noted earlier, 88% of agencies received a 'met' rating on this standard. The ease with which agencies could satisfy the requirements of this standard contributed to its failure to operate as a strong predictor of performance against the objective. Standard 1.2 concerns the allocation of agency resources according to the prioritised need of consumers accessing the service. The critical content of Standard 1.1 may be covered by Standard 1.2, with which it is highly correlated ($r = 0.58$).

A number of agencies indicated that Standard 1.3 (Access to services by consumers with special needs is decided on a non-discriminatory basis) was not well measured by the performance information used. Nine per cent of agencies believed this standard, and its performance information, were impractical. The performance information listed under this standard required that agencies compare the profile of their client base with the profile of likely need in the community. The reason this was considered impractical was that it was too time consuming for agencies to get statistics regarding population profiles. Government HACC officers did not always have these statistics available for agencies and agencies believed that they should be assessed against this standard in some other way. A high correlation ($r = 0.73$) was found between Standard 1.3 and 4.3 (Consumers' cultural needs are addressed), indicating that Standard 1.3 did not supply any more information than 4.3 in terms of addressing issues associated with special needs groups. (A table of correlations among all standards can be found in Appendix D.)

Objective 2: Information and consultation

At least 96% of agencies considered the standards under Objective 2 to be clear, and close to 100% considered them to be desirable. The most common complaints were that Standards 2.1 (Consumers are aware of their rights and responsibilities) and 2.2 (Consumers are aware of services available) were impractical from the perspective of overloading consumers with information. Six per cent of service providers argued that Standard 2.1 was impractical and 10% of service providers argued that Standard 2.2 was impractical.

Table 3.3: Correlation coefficients for standards under Objective 2

	2.1	2.2
2.2	0.36	
2.3	0.39	0.46

The correlations between the standards in Objective 2 indicate that the standards are cohesive in measuring information and consultation issues but, among themselves, are not so highly correlated as to be redundant. Standard 2.1 correlates very highly with several other standards ratings. With Standard 5.1 (Consumers are informed of the privacy and confidentiality procedures and understand their rights in relation to these procedures) the correlation was $r = 0.65$. With Standard 6.1 (Consumers are aware of the complaints process) the correlation was $r = 0.57$, and with Standards 7.1,

7.2, and 7.3 (concerning advocacy), the correlation was $r = 0.57$. These standards also deal with awareness of rights in relation to privacy, confidentiality, complaints and advocacy. The high correlations between these items suggest that they could be more efficiently and just as effectively measured under one performance information request.

Standard 2.2 (Consumers are aware of services available) shows higher correlation with standards from other objectives than any other standard. Its correlation with Standard 1.3 (Access to services by consumers with special needs is decided on a non-discriminatory basis) is $r = 0.62$. With Standard 4.3 (Consumers' cultural needs are addressed) its correlation is as high as $r = 0.82$, and with Standard 4.4 (The needs of consumers with dementia, memory loss and similar disorders are addressed) its correlation is $r = 0.62$. For these standards, the assessor guidelines emphasised the need to facilitate service provision to those of non-English-speaking background and Aboriginal and Torres Strait Islander background. As such, for each of these standards a 'met' rating could not be obtained unless consideration was given to these groups. These correlations are the likely result of such an emphasis. The internal consistency of the standards under this objective may be improved and the degree of repetitive measurement across objectives reduced if the major issues relating to service provision to Aboriginals and Torres Strait Islanders and people of non-English-speaking background are drawn together so that they constitute minimum criteria for a 'met' rating on one particular standard.

Standard 2.3 (Consumers are informed of the basis of service provision, including changes that may occur) shows a strong correlation ($r = 0.56$) with Standard 4.2 (Each consumer has a service delivery/care plan which is tailored to individual need and outlines the service he or she can expect to receive). Informing consumers of their care or service delivery plan very often was said to occur during the same process as the explanation of the basis of service provision. The size of the correlation suggests the information gained from these two standards may be complementary rather than redundant.

Objective 3: Efficient and effective management

While the standards of Objective 3 were seen as desirable by at least 96% of service providers, of all the standards they were most likely to be considered unclear, with, on average, 92% of service providers rating the standards of this objective as clear. These standards were also seen as impractical and inappropriate by a substantial minority (13% and 7% on average, respectively). The standard that was rated highest under this objective on clarity, desirability, practicality and appropriateness was Standard 3.3, with at least 97% of service providers agreeing that it fulfilled each of these requirements.

Table 3.4: Correlation coefficients for standards under Objective 3

	3.1	3.2
3.2	0.51	
3.3	0.44	0.47

Assessors reported difficulty combining the two pieces of performance information into one rating against Standard 3.2. Nevertheless, these standards correlated strongly with one another within the objective, indicating that they appear to be each contributing to the measurement of quality with regard to efficient and effective management.

Results of the factor analysis indicated that the three standards tended to load primarily on one underlying factor, with some cross-loading on only one other factor. Standard 3.3 (Consumers receive services from appropriately skilled staff) showed a strong correlation ($r = 0.65$, see Table 3.2) with Standard 1.1 (Assessment occurs for each consumer). This possibly reflected the tendency for more comprehensive assessments to be undertaken by agencies with highly skilled and qualified staff. The presence of this correlation adds weight to the argument for eliminating the performance information required by Standard 1.1, since an important associated quality factor is measured by Standard 3.3.

Objective 4: Coordinated, planned and reliable service delivery

All of the standards under Objective 4 were considered clear by at least 93% of agencies and close to 100% considered these standards desirable. They were somewhat less likely than standards under other objectives to be considered practical and appropriate, however – with an average of 91% believing the standards under this objective to be practical and an average of 94% believing the standards under this objective to be appropriate.

The standard in Objective 4 that showed least coherence in measuring the domain of coordinated, planned and reliable service delivery was 4.4 (The needs of consumers with dementia, memory loss and similar disorders are addressed). This standard had a correlation of only $r = 0.22$ with Standard 4.5 (Consumers receive services which include appropriate coordination and referral processes) and a correlation of only $r = 0.25$ with Standard 4.1 (Each consumer receives ongoing assessment (formal and informal) that takes all support needs into account). Standard 4.4 was also considered impractical and inappropriate by a minority of agencies (8% and 7% respectively).

The information on the performance of agencies against Standard 4.4 was, in many of cases, the same as that gained from Standard 4.3 (Consumers' cultural needs are addressed). This is evidenced by the high correlation between the two standards ($r = 0.65$). This may be partly due to the emphasis in Standard 4.4 on addressing the needs of dementia sufferers from a non-English-speaking background. Standard 4.4 also correlated highly with Standard 2.2 (Consumers are aware of services available) ($r = 0.62$). The overlap between these two standards can be explained by the reference in both to the need for the agency to be aware of other services available, including those providing support to dementia sufferers as specified in Standard 4.4. Both standards also require agencies to make special provision for those from Aboriginal and Torres Strait Islander backgrounds and non-English-speaking backgrounds.

The factor analysis shown in Table 3.2 reveals that the standards of Objective 4 do not load consistently on any one factor. Standard 4.3 (Consumers' cultural needs are addressed) is correlated with standards from three other objectives: Standard 1.3 (Access to services by consumers with special needs is decided on a non-discriminatory basis, $r = 0.73$); Standard 2.2 (Consumers are aware of services available, $r = 0.82$); and Standard 3.2 (Consumers receive services from agencies that adhere to accountable management practices, $r = 0.60$). Again, these correlations are likely to be the result of an emphasis in each of these standards on providing services in such a way as to facilitate service provision to those of non-English-speaking backgrounds and Aboriginal and Torres Strait Islander backgrounds.

Table 3.5: Correlation coefficients for standards under Objective 4

	4.1	4.2	4.3	4.4
4.2	0.34			
4.3	0.39	0.39		
4.4	0.25	0.46	0.65	
4.5	0.40	0.37	0.37	0.22

Notwithstanding the cross-loading of Standard 4.3 with other objectives, and excluding Standard 4.4, the standards of Objective 4 show an appropriate level of intercorrelation to suggest that they relate together sufficiently to be considered as measuring the domain associated with coordinated, planned and reliable service delivery.

Objective 5: Privacy, confidentiality and access to personal information

On average, the standards under Objective 5 were considered desirable by 99% of service providers. At least 94% of service providers considered these standards to be clear. The poorest performance of these standards was in the area of practicality and appropriateness. At least 85% of service providers considered the standards under Objective 5 to be practical and at least 91% of service providers considered the standards under Objective 5 to be appropriate.

The standard of most concern to agencies under this objective was Standard 5.2 (Consumers have signed confidentiality release forms). The impracticality and inappropriateness of obtaining signed confidentiality release forms was an issue for a range of agencies. Nevertheless, the concern of this standard with access to and control of personal information related it strongly to the other standards under this objective.

The correlations among the standards under this objective were moderate, suggesting that they each tap a different aspect of the objective. The factor analysis confirmed that these items covered a common domain as evidenced by each standard loading on a single factor.

Table 3.6: Correlation coefficients for standards under Objective 5

	5.1	5.2
5.2	0.34	
5.3	0.55	0.46

As would be expected, Standard 5.1 correlated with two other standards concerned with consumer rights. The correlation between Standard 5.1 (Consumers are informed of the privacy and confidentiality procedures and understand their rights in relation to these procedures) and Standard 2.1 (Consumers are aware of their rights and responsibilities) was high at $r = 0.65$. Standard 6.1 (Consumers are aware of the complaints process) was also highly correlated with Standard 5.1 ($r = 0.59$).

Objective 6: Complaints and disputes

The complaints and disputes standards were considered practical, clear, appropriate and desirable by between 99% and 100% of service providers for all of the standards under this objective except Standard 6.4 (Each consumer receives assistance, if requested, to help with the resolution of conflict about a service that arises between the frail elderly person or younger person with a disability and his or her carer). This role was seen to be quite different from that required from agencies in dealing with complaints about their own service. The role of advocate or mediator that this standard was seen to request was not seen as appropriate. The correlations reported in Table 3.7 suggest that Standard 6.4 was also rated differently to the other standards belonging to this objective: the correlations for this standard are among the lowest in the group. Indeed, inspection of the correlation of all standards presented in Appendix D reveals that, compared with its correlation with all other standards, Standard 6.4 is least associated with the complaints standards under Objective 6. Standard 6.4 was most strongly correlated with Standard 3.2 (Consumers receive services from agencies that adhere to accountable management practices, $r = 0.59$). This suggested that Standard 6.4 may be more concerned with well-organised policies and comprehensive guidelines for accountability than with the resolution of complaints.

Table 3.7: Correlation coefficients for standards under Objective 6

	6.1	6.2	6.3
6.2	0.58		
6.3	0.55	0.76	
6.4	0.33	0.41	0.33

Standard 6.3 (Services are modified as a result of ‘upheld’ complaints) correlates very highly with 6.2 (Each consumer’s complaint about a service, or access to a service is dealt with fairly, promptly, confidentially and without retribution, $r = 0.76$), suggesting that information may be more efficiently taken using only one of the standard’s performance information. Standard 6.3 received a high level of ‘met’ ratings and was

noted by assessors to be difficult to rate accurately without consumer feedback. The performance information for Standard 6.2 may be sufficient for these two standards.

Objective 7: Advocacy

While agencies believed that the advocacy standards were desirable (at least 99% of service providers believed these standards were desirable), not all saw them as practical or clear, particularly with regard to Standard 7.4 (The agency involves advocates in respect to representing the interests of the consumer). On average 93% of service providers believed these standards were appropriate while, on average, 88% believed they were practical. A difficulty faced by agencies with regard to these standards concerned their own role as advocates for their consumers. Issues of practicality concerned the absence of advocacy services in the area covered by the agency or failure of these advocacy agencies to disseminate information about themselves.

In the Instrument tested in the pilot, one rating was given for the three standards: 7.1, 7.2 and 7.3. There was a high correlation between the rating of Standard 7.4 and the single rating given for Standards 7.1 through 7.3 ($r = 0.67$), suggesting that both ratings addressed the objective of advocacy. The size of this correlation suggests that advocacy may just as efficiently be covered by one of these pieces of performance information. The ratings for Standards 7.1 through 7.3 were also strongly correlated with Standards 2.1 (Consumers are aware of their rights and responsibilities, $r = 0.57$), Standard 5.1 (Consumers are informed of the privacy and confidentiality procedures and understand their rights in relation to these procedures, $r = 0.55$) and 6.1 (Consumers are aware of the complaints process, $r = 0.53$) – all standards concerned with consumer information and rights. It is likely that little performance information would be lost if the assessment of Standards 7.1 to 7.3 was incorporated into another standard to which they were related, such as Standard 2.1. The redundant questions could then be dropped from the Instrument.

3.3.2 Overall scores

An alpha reliability coefficient represents the degree of cohesiveness of the standards representing each objective. The highest alpha possible is an alpha coefficient of one, which would be obtained if standards were answered identically in every instance. These coefficients are presented for each objective in the diagonal of Table 3.8.

The alpha coefficients show a high level of cohesiveness among the standards representing each objective. Therefore, to obtain a single measure for each objective, the scores for each group of standards can be summed. So, for instance, to obtain a single measure for Objective 1, scores across the five standards would be summed. Single measures for each objective were calculated for each agency in this way and these measures were correlated. The results are presented below the diagonal in Table 3.8.

A comparison of the size of the correlations between objectives and the alpha reliability coefficients indicates that, in some cases, the relationship of standards across objectives is stronger than within objectives. This confirms that there is considerable repetition within the Instrument. The same measures of quality have been used to assess different standards across objectives. This means that the Instrument contains more items than it needs to in order to make a measurement of service quality. It suggests that standards

do not comprehensively cover the domain of their own objective without unduly overlapping with the domains of other objectives. Reducing overlap across objectives and comprehensively assessing quality within objectives is critical to forming an accurate additive scale across standards.

When the objectives were examined in the previous section, a number of standards were identified as contributing little to the assessment of performance against the standards and some standards were seen to be more closely related to standards in other objectives. The question to be addressed is whether the performance information required under some standards should be eliminated, or whether some standards would be better grouped or subsumed under different objectives.

Table 3.8: Correlations and alpha reliability coefficients for the seven objectives

Objectives	1	2	3	4	5	6	7
1. Access to services	0.80						
2. Information and consultation	0.74	0.66					
3. Management	0.71	0.58	0.73				
4. Service delivery	0.74	0.78	0.78	0.75			
5. Personal information	0.49	0.61	0.33	0.45	0.71		
6. Complaints and disputes	0.72	0.66	0.56	0.68	0.58	0.79	
7. Advocacy	0.51	0.53	0.50	0.46	0.36	0.52	0.80

Note: Alpha reliability coefficients in the diagonal; correlations between objectives below the diagonal.

Standards found to contribute little to the measurement of service quality once other standards were taken into account included: 1.1 (Assessment occurs for each consumer), 1.3 (Access to services by consumers with special needs is decided on a non-discriminatory basis), 4.4 (The needs of consumers with dementia, memory loss and similar disorders are addressed) and 6.3 (Services are modified as a result of ‘upheld’ complaints). Standards 2.1 (Consumers are aware of their rights and responsibilities), 5.1 (Consumers are informed of the privacy and confidentiality procedures and understand their rights in relation to these procedures) and the first three standards under Objective 7 all deal with consumer awareness of their rights and may be more effectively measured together rather than individually in their respective objectives.

Standard 6.4 did not relate strongly with other standards within its own objective. The performance of agencies against this standard is not associated with their performance on other measures of quality service regarding complaints. It was more closely related to issues of accountability and responsibility. This standard may require industry education to incorporate it into the quality assurance process of all HACC agencies. It can be considered as a factor in the measurement of service quality, but should not be included as an indicator of performance against Objective 6, since of all the standards it is least related to those concerned with complaints.

3.3.3 Recommended changes to scoring

Summary of recommended changes to facilitate scoring of the Instrument:

1. Objective 1 should be limited to the performance information of Standards 1.2, 1.4 and 1.5.

2. Objective 2 should be extended to include performance information covered by 5.1 and 7.1, 7.2, and 7.3 under Standard 2.1.
3. Objective 3 should remain as it is.
4. Objective 4 should be limited to the performance information of Standards 4.1, 4.2, 4.3 and 4.5.
5. Objective 5 should be limited to the performance information of Standards 5.2 and 5.3.
6. Objective 6 should be limited to the performance information of Standards 6.1 and 6.2 with separate consideration given to Standard 6.4.
7. Objective 7 should be limited to the performance information of Standard 7.4.

Table 3.9 shows the alpha reliability coefficients and the intercorrelations for objectives when these adjustments are made. Standard 6.4 is listed separately because it did not correlate strongly with standards within its own objective and it was more highly correlated with a standard in another objective. The balance of the internal consistency of objectives to cross association with other objectives is improved by the changes listed above: the alpha coefficients now tend to be higher than the correlations between objectives. This means that the standards with these adjustments are more likely to represent quality associated with their objective rather than quality associated with another objective.

Table 3.9 Correlations and alpha reliability coefficients for the seven revised objectives

Objectives	1	2	3	4	5	6	7
1. Access to services	0.75						
2. Information and consultation	0.66	0.66					
3. Management	0.63	0.58	0.73				
4. Service delivery	0.64	0.76	0.75	0.70			
5. Personal information	0.32	0.48	0.26	0.36	0.63		
6. Complaints and disputes	0.64	0.65	0.46	0.57	0.38	0.73	
7. Advocacy	0.46	0.45	0.48	0.49	0.11	0.42	*
Standard 6.4	0.50	0.54	0.61	0.53	0.44	0.42	0.53

* Coefficient cannot be calculated as there is only one standard under this objective.

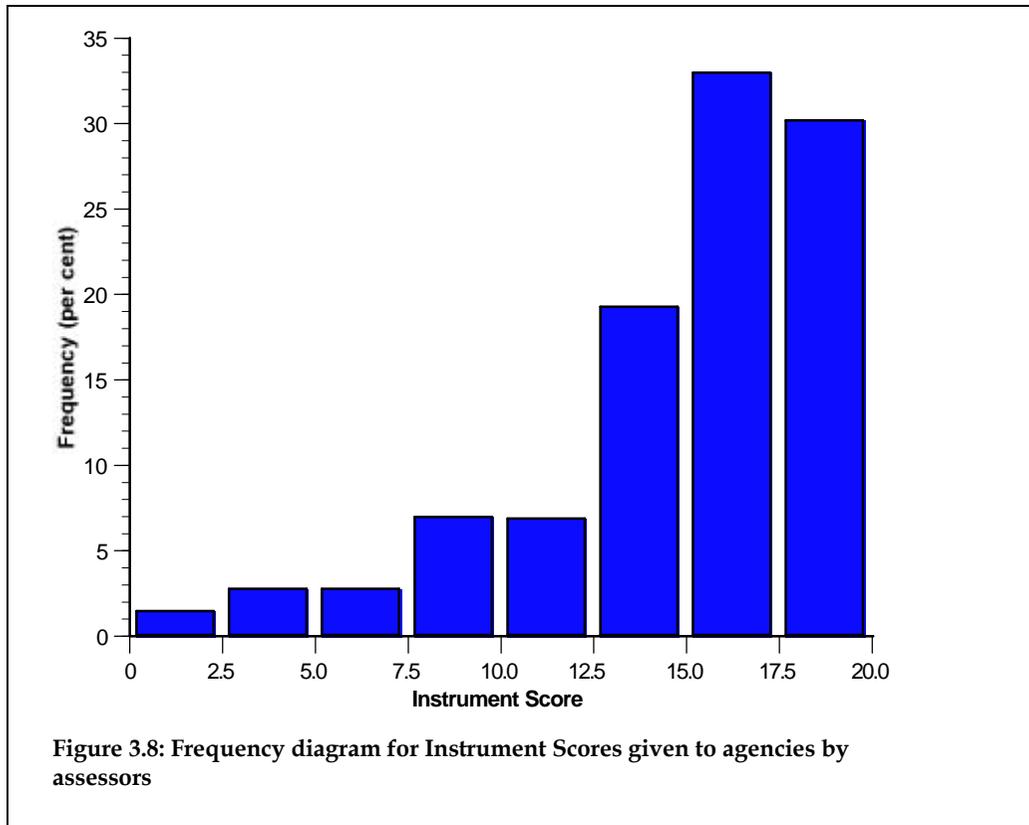
All of the analyses presented so far in this chapter converge on the conclusion that scores can be added across all standards to give a total score for quality assessment against the HACC National Service Standards. Across the whole Instrument, no one standard stands out as unrelated to the other standards and relationships between objectives are almost as high as the relationships of standards within objectives. This suggests that there is one basic theme to the 27 standards: they are cohesively related to quality in a diversity of HACC agencies.

3.3.4 Finding the Instrument Score

Using the revised scoring system (which reduced the total number of standards to be scored to 19), on average, agencies in the pilot scored 28.7 out of a maximum possible

score of 38. The range of total agency scores was large; some agencies scored the maximum possible, some scoring as low as 5 and 6 out of 38. The standard deviation of total agency scores was 7.6.

Summed scores for the Instrument and individual objectives were calculated using assessor scores. Although assessor scores were available for 74 agencies, summed scores were calculated only for 60 of these. This was because agencies with any unscored standards were excluded from the sample. Exclusion of cases with missing ratings is necessary at this stage since comparison of agency's added scores is dependent on each being rated by an equivalent number of standards.



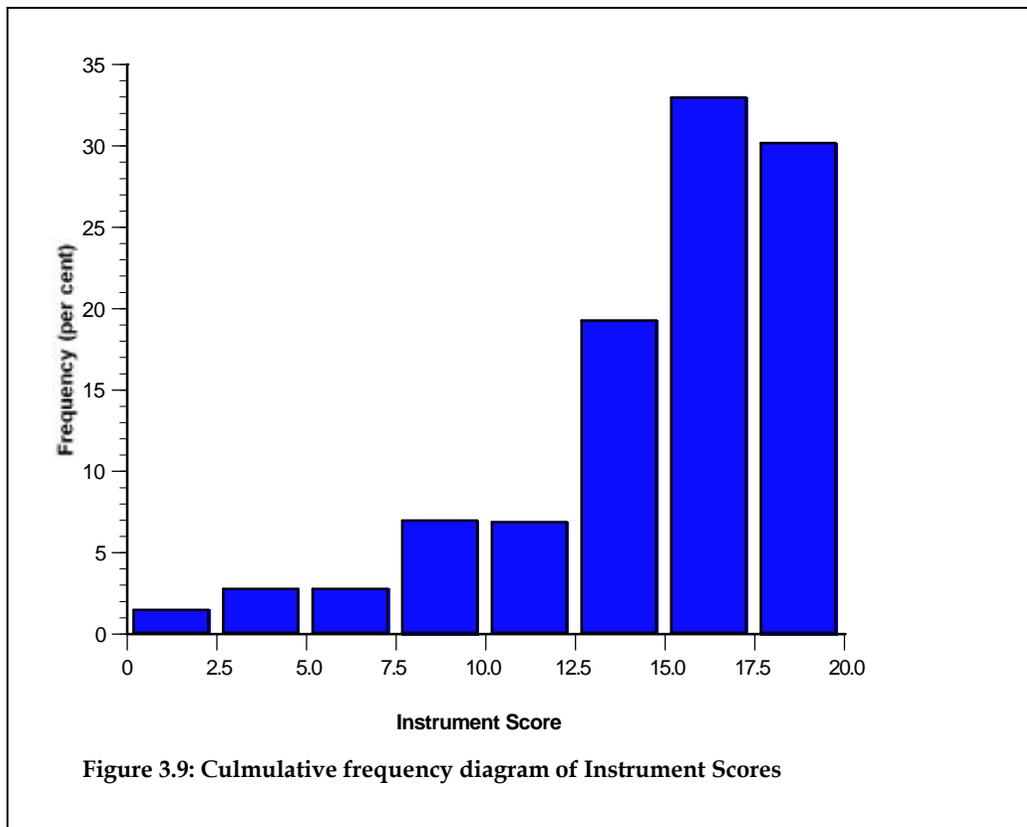
It was recognised that in some cases (e.g. Standard 4.4 in the pilot test) it may be necessary for an agency to have no score against a standard. This causes a problem if scores are simply added, because agencies which are legitimately excluded from assessment against a particular standard lose the value of a score for that standard.

To eliminate this problem, and to facilitate comparison of performance between agencies, an agency's scores for each standard are added together. This total is then divided by the number of applicable standards to arrive at an average score for each agency. Multiplying this score by 10 produces an Instrument Score with a range from 0 to 20. This Instrument Score is a valid way of representing an agency's performance

against the Instrument, assuming that each standard is of equal weighting. As discussed above, the recommended changes to the Instrument result in the removal of redundant performance information.

The mean Instrument Score for all agencies in the pilot test with assessor ratings is 15.0. The standard deviation is 4.0. Figure 3.8 groups agencies according to their Instrument Scores and shows the proportion within each group. Sixty three per cent of agencies obtained an Instrument Score greater than 15.0. These agencies have scored a 'met' rating for the majority of standards.

Figure 3.9 shows that the vast majority of agencies have Instrument Scores between 10.0 and 20.0—or an overall rating of between 'partly met' and 'met'. A relatively small proportion of agencies (11%) had Instrument Scores between 0.0 and 10.0 or an overall rating of between 'not met' and 'partly met'. Targeting the relatively few poor performers that are identified using this method may be one response of program managers to the outcomes of an appraisal process in their jurisdiction.

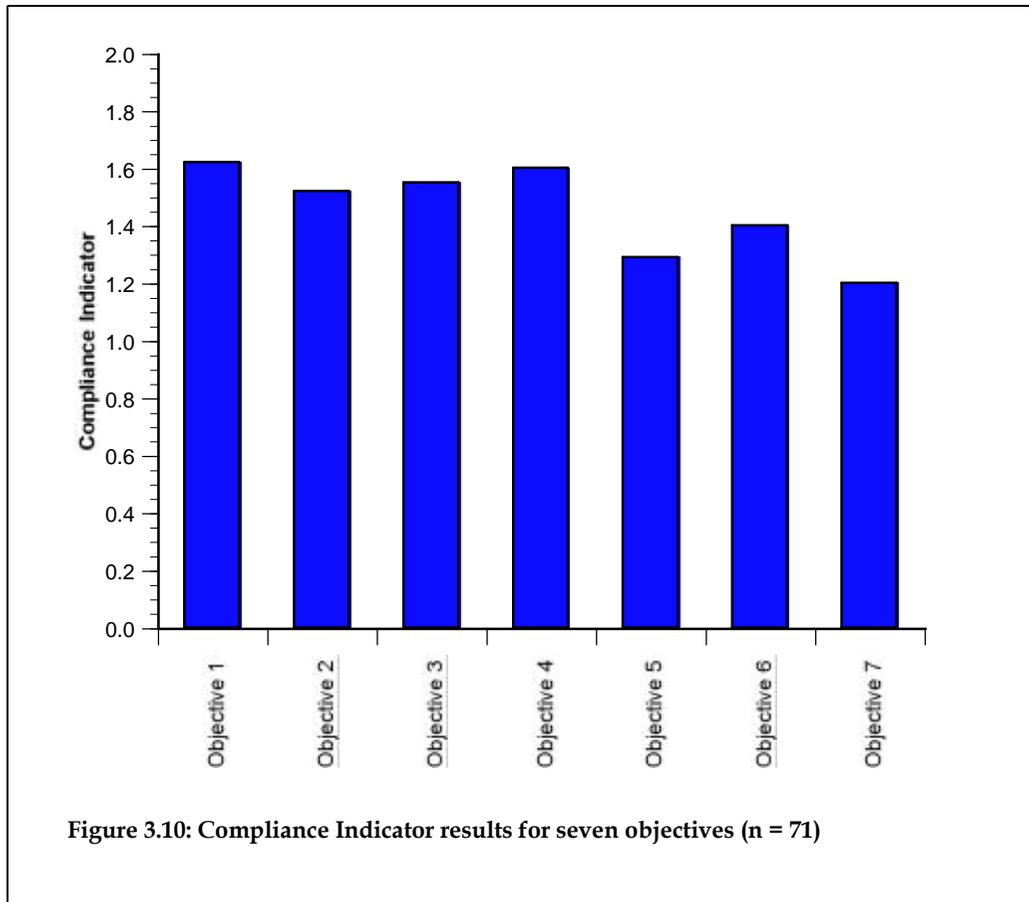


3.3.5 The Compliance Indicator

To make a comparison of performance across objectives it is necessary to calculate a Compliance Indicator. If scores were simply added across standards within an objective it would not be possible to compare performance against one objective with another.

This is because the number of standards in each objective varies. However, the following method is proposed to develop an indicator based on agency scores which allows for comparison across objectives.

To arrive at the Compliance Indicator the mean total score per objective is first calculated. For each agency, the scores against standards within an objective are added. The maximum possible score for an objective is determined by the number of standards within it. By dividing the total objective by the number of standards rated within the objective, we arrive at the Compliance Indicator, which is always expressed as a score out of 2.



For example, the Compliance Indicator of performance against Objective 1 is calculated in the following way. Using the revised version of the Instrument, three standards are used to assess Objective 1 within the Instrument. Scores for these 3 standards are added for each agency to achieve a total score for the objective. Total scores for each agency are then divided by the number of scores under the objective to obtain the Compliance Indicator. This calculation for Objective 1 produces an average Compliance Indicator of 1.63. This indicator can be calculated at the individual agency level as well as at the

aggregate level for all agencies and can be useful for comparison between objectives and agencies.

A breakdown of Compliance Indicator results against each objective, averaged over all agencies, is presented in Figure 3.10.

The objective that agencies performed most poorly on in the pilot test was Objective 7 (Advocacy). The average Compliance Index for this objective was 1.20. Agencies also performed considerably less well against Objective 5 (Compliance Indicator = 1.29). Agencies generally performed well against Objective 1 and Objective 4 (Compliance Indicator = 1.62 and Compliance Indicator = 1.61, respectively). Performance against Objectives 2, 3 and 6 were comparably high (1.52, 1.56 and 1.40, respectively).

3.3.6 Some possible uses of scores

As outlined in previous sections, with the recommended changes to the Instrument (see 3.3.3) there are various ways that scoring of the Instrument could be used. A few of the possible questions that can be answered by the scoring methods outlined in previous sections are listed below.

Question: How are HACC agencies performing against the National Service Standards?

This question could be asked in the context of Business Reporting requirements at both the national or State/Territory levels. The Instrument Scores could be used to derive an overall performance assessment for agencies. For example, the overall performance against the standards of agencies with an Instrument Score less than 10.0 could be described as poor. Scores between 10.0 and 14.9 could be described as basic, scores between 15.0 and 17.5 could be described as good, and between 17.5 and 20.0 could be described as high.

One way of measuring the aggregate performance of HACC agencies in a jurisdiction is to identify the proportion of HACC agencies that fall into these groupings associated with overall Instrument Scores. Alternatively, business reports could adopt an exception-reporting approach and ask only for the proportion of agencies that fall below an agreed level of performance and/or above a certain level of performance (e.g. below 10.0 or above 17.5 on the Instrument Score scale, respectively).

Question: Which agencies should be the subject of targeted action to assist them improve their performance against the National Service Standards?

The agency Instrument Score and the resources available for service development in a jurisdiction are factors to be considered in the choice of agencies to be reviewed. The level of resources and the number of agencies falling below an agreed point between 0.0 and 20.0 on the Instrument Score scale could be used as one basis for selection of agencies for targeted action. By looking at the distribution of agency Instrument Scores along the Instrument Score scale of 0.0 to 20.0 (see Section 3.3.4), a score below which agencies are considered appropriate for targeting could be established. Different approaches to risk management may mean that jurisdictions opt for different points along the Instrument Score scale below which an agency is considered appropriate for targeted action. The selection of this score may well be influenced by the resource implications of targeting all agencies falling below the chosen point.

Question: Where should available program development funds (at national or State/Territory levels) be directed?

An indication of where training and service development resources for improving service quality in the HACC program can be obtained by the use of the Compliance Indicator (see Section 3.3.5). This indicator relates to agency performance against each objective and, at the aggregate level, indicates the relative performance of all agencies for each objective. For example, in the pilot test, agency performance against Objective 7 (Advocacy) is the poorest of all, suggesting that this objective would be an appropriate focus for training and development activities. Decisions about program development activities would, of course, also take into account other factors, including the relative priority placed on particular objectives by the relevant jurisdiction. The Compliance Indicator does, however, provide a useful summary indicator of relative performance against particular objectives.

Where overall Compliance Indicators for each objective show that problems with one objective are common across all jurisdictions, the need for national training or development activities may be indicated. Where problems with objectives are specific to jurisdictions, they may need to address these separately through State- or Territory-specific quality improvement strategies.

3.4 Concurrent validity

Two tests of the concurrent validity of the Instrument are made in this section. The first tests the concurrent validity of the individual standards ratings. Ratings given to an agency by assessors are compared to agency self-ratings. The second test focuses on the concurrent validity of the Instrument Score. The Instrument Scores derived in the previous section are compared with the overall appraisals of agency performance requested of agencies and assessors at the end of the Instrument. Individual standards ratings are also compared with the Instrument Score to examine the relative contribution of each standard to the assessment of quality.

3.4.1 Agency and assessor agreement on ratings against individual standards

To determine the extent to which agencies and assessors were in agreement, the ratings that agencies gave themselves against individual standards in the Instrument were compared with the ratings given to them by the assessor on the Ratings Summary Form. From this information a measure of agreement was calculated. Agreement means that if the assessor rated their agency 'met', the agency also rated it as 'met'; or if the assessor rating was 'partly met', the agency rating was also 'partly met'; or if the assessor rating was 'not met' the agency rating was also 'not met'. Figure 3.11 shows these agreement measures for each standard.

There are several factors at work influencing the level of agreement between agency and assessor ratings:

- the bias of the individual rater and their tendency to be harsh or lenient in deciding on ratings;

- the consistency between the Instrument’s instructions and the assessor guidelines in terms of placing equivalent emphasis on factors associated with the standard; and
- the nature of the data used to appraise quality of service against the standards— for some standards, the evidence required to support the performance information is more difficult to obtain and assessors may be left to rely on the agency’s word.

The average percentage of agreement between agencies and assessors was 76% over all standards. The lowest level of agreement was 66% for Standard 6.1 (Consumers are aware of the complaints process). This represents a high level of concurrence given that agencies received less information than assessors regarding how to rate standards.

The highest levels of agreement were for Standards 1.1 (Assessment occurs for each consumer; 90%), 2.2 (Consumers are aware of services available; 100%), 4.2 (Each consumer has a service delivery/care plan which is tailored to individual need and outlines the service he or she can expect to receive; 86%), and 6.3 (Services are modified as a result of ‘upheld’ complaints; 84%). A common feature of Standards 1.1 and 4.2 is that they were specific about the evidence required of agencies to receive a ‘met’ rating. Respectively, these standards required proportions of consumers who had been assessed or who had a service delivery/care plan. As mentioned under Section 3.1, ratings against these standards may have been affected by the lack of a clear definition of ‘assessment’ and ‘care plan’ for the purposes of the Instrument. In addition, Standard 1.1 was largely rated as ‘met’, resulting in less variance in scores to facilitate comparison.

Standard 2.2 required that agencies describe how they provide information about other services to consumers. In addition to the consistency between the Instrument and the guidelines regarding what was required of agencies to achieve a ‘met’ rating, this standard was one able to be applied consistently to a range of agencies without difficulty. Service providers found this standard to be appropriate to their agencies in all but one case.

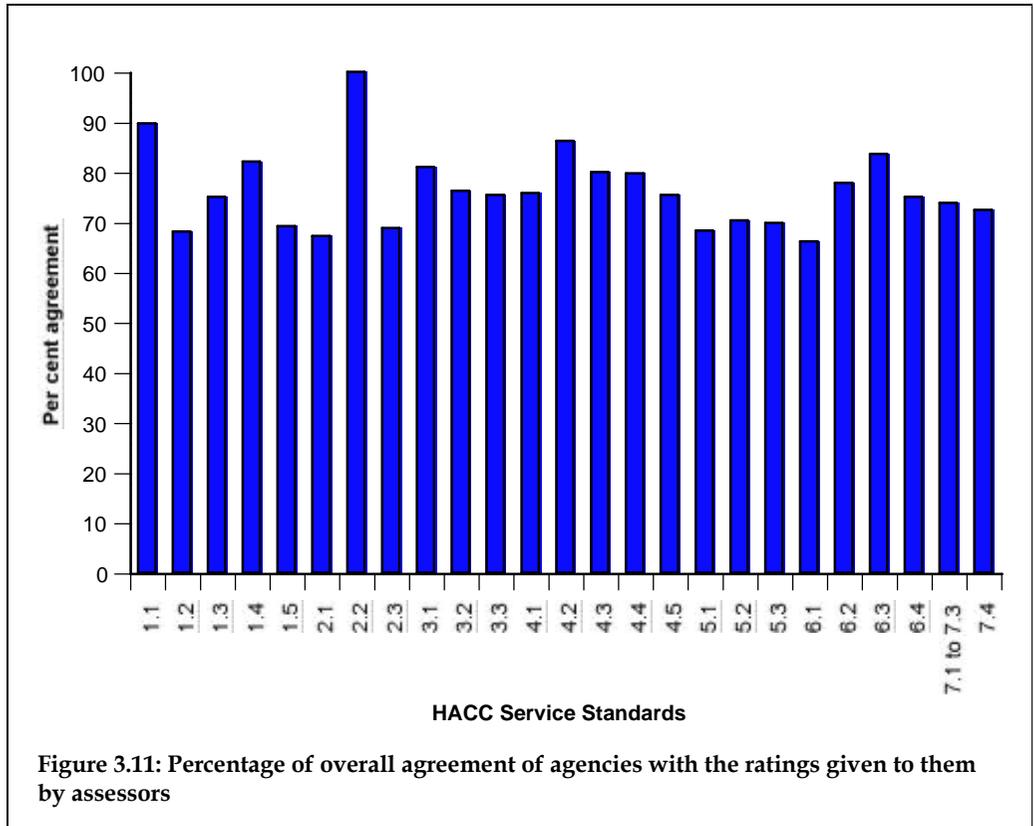
Standard 6.3 required that agencies give an example of service modifications following an upheld complaint. When agencies are able to provide evidence of such an event, with accompanying documentation, there is little assessors can do to disprove the agency’s claims regarding the standard. Unless provided with evidence that upheld complaints do not result in service changes, the assessor must take the agency’s word on the matter. In addition, assessors noted that it was difficult for them to decide on a rating when the agency reported no recorded complaints. Standard 6.2 was reported by assessors to be similarly difficult to assess when no complaints were recorded.

Sources of ratings disagreement

While in general there was a reasonably high degree of consensus between assessor and agency ratings, it is useful to examine individual standards for sources of ratings disagreement in order to further refine and improve the Instrument and the quality appraisal process.

Differences in emphasis and in the amount of information provided between the assessor guidelines and the Instrument contributed to ratings disagreement in some cases. Standards 1.5 (Consumers who reapply for services are assessed with needs

being prioritised), 2.1 (Consumers are aware of their rights and responsibilities) and 5.1 (Consumers are informed of the privacy and confidentiality procedures and understand their rights in relation to these procedures) were particularly affected by this. The percentage agreement between agency and assessor ratings for these standards was 69%, 67% and 68% respectively.



The disagreement on Standard 2.1 (Consumers are aware of their rights and responsibilities) was the result of overly prescriptive ratings directions in the assessor guidelines. A ‘not met’ rating against this standard was directed to be given if the agency made no provision to allow relevant special needs groups to understand their rights and responsibilities, including people of a non-English-speaking background and Aboriginal and Torres Strait Islander peoples. In attempting to rate themselves against this standard, agencies indicated that they believed this to be too harsh. It was also not clear to assessors how often ‘periodic reminders’ should be.

Standards where some degree of ratings disagreement occurred as a result of the difficulty of the information to be assessed included the following: 1.2 (Consumers are allocated available resources according to prioritised need); 1.3 (Access to services by consumers with special needs is decided on a non-discriminatory basis); 5.2 (Consumers have signed confidentiality release forms); 6.4 (Each consumer receives

assistance, if requested, to help with the resolution of conflict about a service that arises between the frail elderly person or younger person with a disability and his or her carer); 7.1 to 7.3 regarding knowledge of and access to advocacy services (74% agreement); and 7.4 (The agency involves advocates in respect to representing the interests of the consumer) with 72% agreement.

Standard 1.2 (Consumers are allocated available resources according to prioritised need) required an assessment of the manner in which agencies prioritised need and allocated resources. Appropriate performance information for this standard varied greatly across service types, target group, and demands on the agency. As a result, the assessor guidelines were fairly open in their description of what would constitute a 'met', 'partly met', or 'not met' rating. In some cases, agencies did not prioritise resource allocation at all because they were able to meet all demand. These factors were reported by assessors to have made rating difficult and may have contributed to the discrepancy between agency and assessor ratings. Notwithstanding these difficulties, agreement on ratings occurred 68% of the time.

Standard 1.3 (Access to service by consumers with special needs is decided on a non-discriminatory basis) was difficult to assess because of the presence of specific services for people from non-English-speaking backgrounds and Indigenous peoples in the same geographic area. Agreement on the ratings for this standard occurred 75% of the time.

For Standard 6.4, some assessors were uncertain as to whether all services should have developed protocols for dealing with conflict between carers and the frail elderly or younger person with a disability, or just for issues of abuse of older people. This standard may also have suffered from lack of clarity about the type of conflict situation to which the standard refers. That is, the standard refers to conflict between a carer and a care recipient about a service – not conflict in general, which may be more closely connected to issues of elder abuse. Agreement about ratings for this standards occurred 75% of the time.

The best practice for Standard 5.2 was also called into question by assessors. Consumers may not have wanted to sign or give blanket release. They may have preferred to deal with each case or event when it occurred. Agreement about ratings for this standard occurred 70% of the time.

More generally, standards under Objectives 2, 6 and 7 were noted to be difficult to assess without consumer feedback. Assessors reported that agencies often gave only verbal coverage to issues such as advocacy, making assessment of these standards difficult and leading to disagreement. This was also the case with complaints, particularly with regard to Standard 6.3 concerning upheld complaints. Clearer guidelines on the appropriate role of agencies acting as advocates at their clients' request were also required to resolve some of the disagreement between ratings of assessors and agencies. Agreement about the ratings for Standards 7.1 to 7.3 occurred 74% of the time and agreement about the ratings for Standard 7.4 occurred 72% of the time.

3.4.2 Validating the Instrument Score against the global assessment of quality

After assessors and agencies had completed their ratings of the agency's performance against the standards, they were asked to give the agency an overall appraisal of service

quality with respect to the standards. In particular, they were asked to answer the question, 'On the basis of information gathered in this quality appraisal, I would rate this agency against the HACC National Service Standards as:...'. Four categories: 1 (*fails to meet the HACC National Service Standards*); 2 (*meets to a minimal or basic standard*); 3 (*meets to a good standard*); and 4 (*exemplary*) were available in response. The instructions given to assessors in the training session regarding this question were that they were to give a global assessment based on their own judgment of overall service quality.

If the correlation between the global assessment and the Instrument Score was too low there would be some serious concerns about the validity of the summed Instrument Score. It would also indicate that assessors and service providers may have some doubts about the meaningfulness of the Instrument as a tool for assessing quality of service in HACC-funded agencies. If the correlation is very high, it indicates that the same assessment results could be obtained by simply asking one question from assessors and agencies, without need to individually rate standards. It may suggest that individual standards ratings are unduly subject to the same subjective judgments as the overall appraisal.

Most commonly, both agencies and assessors indicated that agency service quality fell into the third category on the global assessment measure: 'meets to a good standard'. Sixty-eight per cent of agencies were rated in this category according to both assessors and the agencies themselves. The correlation between the global assessment and the calculated Instrument Score, where both of these data items were obtained from assessors, was $r = 0.74$.³ When both of these data items were obtained from agencies the correlation was $r = 0.64$.⁴ Both of these correlations indicate a satisfactory level of concurrent validity. The lower correlation for agency assessments may reflect the fact that agencies were not provided with detailed guidelines for the rating of individual standards. Nevertheless, both agencies and assessors clearly draw some congruence between standards ratings and overall agency service quality.

3.4.3 Validating the standard ratings against the global assessment of quality

Table 3.10 shows the correlations between individual standards ratings, according to assessors, and the assessor's global assessment of the agency. Each standard shows a positive correlation with the global appraisal. This confirms the notion that each standard measures some aspect of service quality. The lowest correlations occur between the global appraisal and Standards 5.2 (Consumers have signed confidentiality release forms, $r = 0.26$) and 5.3 (Consumers are able to gain access to their personal information, $r = 0.20$). This suggests that, according to assessors, these standards are of lowest validity with respect to the assessment of overall service quality in HACC funded agencies. Standard 5.2 was considered one of the least practical and appropriate by service providers.

3. Number of cases with available data = 59

4. Number of cases with available data = 79

Standard 1.1 (Assessment occurs for each consumer) and 6.3 (Services are modified as a result of 'upheld' complaints) were also among the least strongly correlated with the global assessment. These standards were noted earlier to have the highest frequency of 'met' ratings. With almost 90% of agencies scoring a 'met' rating for Standard 1.1, there was very little variability in ratings available to produce a high correlation with a global assessment of quality. It was also noted earlier that it was difficult for assessors to adequately verify the information provided to them by agencies regarding Standard 6.3.

Standards with the highest correlations with the global assessment of quality included both 1.3 (Access to services by consumers with special needs is decided on a non-discriminatory basis) and 4.3 (Consumers' cultural needs are addressed), standards that were concerned with adequate service provision to special needs groups. Clearly, agencies that are able to provide equity in service provision are perceived by assessors to be performing well in terms of the National Service Standards. Accountable management practices were also of apparent influence in the assessor's perception of performance against the standards: ratings against Standard 3.2 correlated highly with global assessments ($r = 0.58$). Similarly, the managerial competence demonstrated under Standard 3.1 (Consumers receive appropriate services provided through the processes of ongoing planning, monitoring and evaluation of services) was strongly associated with the assessor's overall appraisal of agency performance ($r = 0.56$). Regarding the more practical aspects of service delivery, ratings against Standard 4.2 (Each consumer has a service delivery/care plan which is tailored to individual need and outlines the service he or she can expect to receive) were a strong indicator of the assessors' overall perception of agency performance against the standards.

Significantly, some standards that were considered by service providers to be of questionable clarity, desirability, appropriateness and practicality were among those that received strong validation in terms of their correlation with assessors' overall appraisals. Most notably, Standard 6.4 (Each consumer receives assistance, if requested, to help with the resolution of conflict about a service that arises between the frail elderly person or younger person with a disability and his or her carer) had been questioned by service providers on all of the above criteria but appears to be strongly associated with the global assessment ($r = 0.46$). Standard 3.2 was among the most frequent to be considered impractical and unclear by service providers, yet it is shown in Table 3.10 to be most strongly associated with a global assessment of quality ($r = 0.58$).

Table 3.10: Correlation of assessor standard ratings with global assessment of quality

	Standard	Correlation
1.1	Assessment occurs for each consumer.	0.37
1.2	Consumers are allocated available resources according to prioritised need.	0.46
1.3	Access to services by consumers with special needs is decided on a non-discriminatory basis.	0.55
1.4	Consumers in receipt of other services are not discriminated against in receiving additional services.	0.49
1.5	Consumers who reapply for services are assessed with needs being prioritised.	0.47
2.1	Consumers are aware of their rights and responsibilities.	0.51
2.2	Consumers are aware of services available.	0.47
2.3	Consumers are informed of the basis of service provision, including changes that may occur.	0.41
3.1	Consumers receive appropriate services provided through the processes of ongoing planning, monitoring and evaluation of services.	0.56
3.2	Consumers receive services from agencies that adhere to accountable management practices.	0.58
3.3	Consumers receive services from appropriately skilled staff.	0.50
4.1	Each consumer receives ongoing assessment (formal and informal) that takes all support needs into account.	0.44
4.2	Each consumer has a service delivery/care plan which is tailored to individual need and outlines the service he or she can expect to receive.	0.57
4.3	Consumers' cultural needs are addressed.	0.55
4.4	The needs of consumers with dementia, memory loss and similar disorders are addressed.	0.43
4.5	Consumers receive services which include appropriate coordination and referral processes.	0.51
5.1	Consumers are informed of the privacy and confidentiality procedures and understand their rights in relation to these procedures.	0.51
5.2	Consumers have signed confidentiality release forms.	0.26
5.3	Consumers are able to gain access to their personal information.	0.20
6.1	Consumers are aware of the complaints process.	0.54
6.2	Each consumer's complaint about a service, or access to a service, is dealt with fairly, promptly, confidentially and without retribution.	0.43
6.3	Services are modified as a result of 'upheld' complaints.	0.37
6.4	Each consumer receives assistance, if requested, to help with the resolution of conflict about a service that arises between the frail elderly person or younger person with a disability and his or her carer.	0.47
7.1	Each consumer has access to an advocate of his or her choice.	0.42
7.2	Consumers know of their rights to use an advocate.	0.42
7.3	Consumers know about advocacy services—where they are and how to use them.	0.42
7.4	The agency involves advocates in respect to representing the interests of the consumer.	0.45

Note: Correlations between the global assessment of quality and standards 7.1, 7.2 and 7.3 are the same because these three standards receive one common assessment rating.

3.5 Rater reliability

An assessment tool is of little use and can have little credibility where different assessors cannot produce the same result using it to assess the same agency. The credibility of the Instrument as a reliable tool is examined first from the perspective of the assessors who used it. Both quantitative and qualitative data on the difficulty of rating standards are discussed. The results of a reliability study are then presented to empirically test the inter-rater reliability of the Instrument.

3.5.1 Qualitative feedback from assessors

Flexibility and the 80/20 rule

In order to allow assessors flexibility in applying the Instrument to the wide variety of HACC agencies, an assessment rule was devised as a general scoring guide. Assessors were explicitly instructed during training to use an '80/20 rule' in rating standards as 'met', 'partly met' and 'not met'. This was described in more detail in Section 2.2.2.

The 80/20 rule worked well for some assessors but not for others. It was noted that the 80/20 rule may have been applied somewhat inconsistently by assessors as, although it had been part of the assessor training, it had not been included in the assessor guidelines. It was further suggested that the ratings decisions that occurred as a result of this rule may have been partly dependent on the emphasis the assessor placed on practice versus policy. This problem may have been eliminated or reduced by clearer, more specific guidelines. There was strong support for guidelines which clearly outline the essential elements in meeting a standard and provide a checklist for assessors to use in determining ratings. Open-ended questions were noted to be particularly difficult to score accurately.

As was noted earlier in this report, however, there exists a conflict between the aim to clearly specify how agencies should be scored against the standards and the aim to apply the same Instrument to all agencies. By clearly specifying how to determine ratings for an agency the Instrument's reliability is increased. However, the more detail that is used in specifying certain criteria, the less applicable the Instrument becomes across agency types. In other words, as specificity regarding requirements to meet the standards increases, the generalisability of the Instrument decreases. An Instrument that does not generalise across agencies is one that is not valid for different agency types. For these reasons a checklist approach is still not recommended.

Comments on the rating categories

'Met', 'partly met', 'not met'

The three-point rating scale used to assess agencies against each standard was seen to be unfair by some assessors. A wide range of service quality was noted to have been subsumed into the category of 'partly met'. It was seen as discouraging for those agencies who were attempting to improve the quality of their agency to receive the same score as one doing nothing. It was believed that more rating categories would make the scoring clearer and would also identify progress and give recognition to those agencies where something had been done.

It was also noted that there was no provision within the existing three-point rating scale to recognise agencies which exceeded the requirements of the standards. A number of assessors felt that questions and answers required in the Instrument and assessor guidelines did not relate to, or measure, best practice. The addition of an 'excellent' category (or equivalent) would, however, add to the complexity of scoring agencies against the standards and would raise further issues regarding the consistency between raters when deciding on how to classify an agency as 'excellent'.

Assessors noted that some agencies tended not to give themselves 'met' ratings because they could see ways their service could be improved against each standard. These agencies tended to interpret 'met' as being perfect, implying no improvement could be made. As discussed in 2.5.1, under the 80/20 rule, assessors were instructed that this 'met' category was not intended to indicate that no further improvement could be made. However, this was not clear to agencies who did not have the benefit of the assessor guidelines. Clearer instructions to agencies in relation to scoring will assist with this problem.

Assessors also noted that where there were many criteria to meet against a single standard the assessment of what rating was appropriate was more complicated, and that the assessor guidelines required clearer and more consistent direction in determining ratings.

Some assessors indicated greater difficulty in differentiating between 'partly met' and 'not met' ratings than between 'met' and 'partly met' ratings. Splitting the 'partly met' rating into two separate categories could be used to help to overcome this problem. However, this option is not recommended because of the increased complexity involved with providing guidelines sufficient to help assessors differentiate between two 'partly met' ratings. To some extent, grades within the 'partly met' category can be reflected in assessor comments contained in the Ratings Summary Form which are then transferred to the Action Plan section of the Instrument.

'Not appropriate' category

For some agencies, all the standards within the Instrument can be applied to their service only with considerable flexibility of interpretation. It was suggested that a 'not applicable' category be added to the Instrument to take into account agency diversity and to ensure that agencies were not penalised by receiving 'not met' scores on standards that were not applicable to them.

Under four of the five assessment methods trialled, however, assessors were available to agencies to interpret the standards for their service. Where assessors were involved in interpreting the standards to agencies, standards were generally found to be appropriate. Indeed, in joint assessments, all standards (except 4.4 – see next page) were found to be appropriate to all agencies. This result concurred with pre-pilot testing results which indicated that, provided assessors were flexible in their approach to understanding agency service delivery, an assessment could be made for all of the standards in the Instrument for all agency types included in the pilot.⁵

5. Agencies providing information and advocacy services were excluded from the pilot. Pre-testing of the Instrument revealed that service quality in these agencies would be more appropriately assessed by a service-specific Instrument.

One exception to this, however, is Standard 4.4 (The needs of consumers with dementia, memory loss and similar disorders are addressed). This standard is clearly not appropriate to HACC services provided to the young disabled. In this case, a 'not applicable' category may be appropriate.

The more general inclusion of a 'not applicable' category is not, however, recommended. If many agencies took the opportunity provided by a 'not applicable' category to exclude their agency from measurement against standards, the level of non-response to standards would be likely to increase sharply, thus threatening the validity and reliability of the Instrument. The pre-pilot testing and the results of the pilot test indicate that inclusion of this category is not generally warranted. However, it is noted that the assistance of the assessor may be crucial to an agency's ability to interpret and understand just how each standard applies to their particular service. This is one of the reasons that inclusion of a visit from an assessor to the agency during the assessment process is highly recommended, particularly in the initial stages of implementation when agencies have less familiarity with the Instrument and the quality appraisal process.

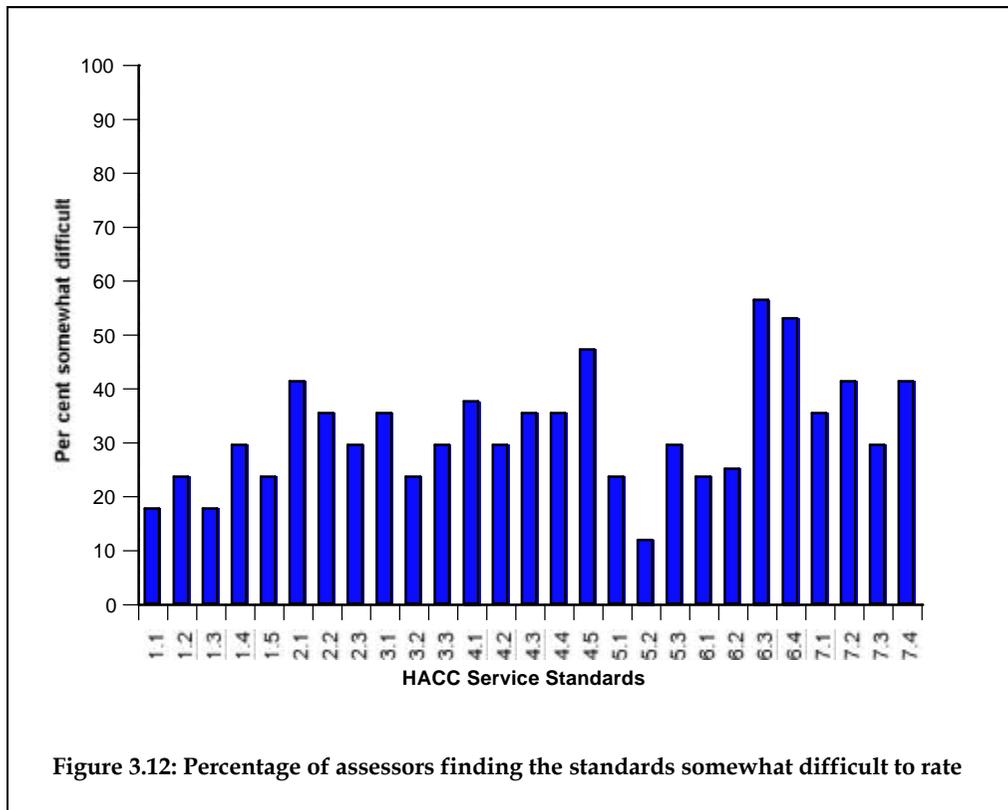
3.5.2 Assessor views on the difficulty of rating standards

Difficulty of rating the standards

An Assessment of Review Process survey was distributed to assessors after they had conducted their assessment interviews, with 17 returning completed Instruments. Assessors were asked to indicate how difficult they found each of the standards to rate using a three-point scale: 1 (*not at all difficult*); 2 (*somewhat difficult*); or 3 (*very difficult*). In general, assessors did not find the standards difficult to rate. Averaging over standards, 64% of assessors found the standards not at all difficult to rate, and only 4% found them very difficult. Given that this was their first encounter with appraising agency performance against the standards, these results are very encouraging.

Where problems occurred for assessors in deciding on how to score agencies against particular standards these were more likely to be described as making the rating process somewhat difficult rather than very difficult. The standards most frequently considered very difficult to rate were Standard 3.2 (Consumers receive services from agencies that adhere to accountable management practices), Standard 7.2 (Consumers know of their rights to use an advocate) and Standard 7.3 (Consumers know about advocacy services— where they are and how to use them). However, these standards were considered very difficult to rate by only three of the 17 assessors who responded to the survey. Standard 3.2 covered a range of issues relevant to accountable management practices without clearly specifying evidence that assessors should sight to determine the rating. The breadth of the standard left assessors unsure that they had adequately covered all of the relevant factors to arrive at a correct rating decision. The standards under Objective 7 (Advocacy) were found to be difficult to rate without knowledge of consumer experience in this regard. A further complicating factor in the task of rating against this standard was the lack of consistent agency practice in recording the involvement of advocates, particularly informal advocates. It was unclear to assessors what role and responsibilities an agency should undertake when requested by their consumers to act as advocates.

Figure 3.12 displays the proportion of assessors who considered the standards to be somewhat difficult to rate. The standard most frequently considered somewhat difficult to rate was Standard 6.3 (Services are modified as a result of ‘upheld’ complaints). Just over half the assessors considered this standard somewhat difficult to rate. As noted earlier, it was difficult for assessors to find evidence to contradict the statements made by agencies against this standard. Complaints not upheld were not necessarily recorded and reported by the agency. Consumer input reporting the outcome of complaints would be required to adequately assess this standard.



Standard 6.4 (Each consumer receives assistance, if requested, to help with the resolution of conflict about a service that arises between the frail elderly person or younger person with a disability and his or her carer) was considered somewhat difficult to rate by around half of the assessors. One assessor considered this standard very difficult to rate. Assessors questioned whether it was appropriate to apply this standard in the same way to all agencies.

Standard 4.5 (Consumers receive services which include appropriate coordination and referral processes) was also considered somewhat difficult to rate by around half the assessors. Like Standard 3.2 (Consumers receive services from agencies that adhere to accountable management practices), this standard contained two questions relating to

the standard. Assessors expressed some difficulty assigning one rating when the agency may have satisfied one question but not the other.

As noted earlier, where assessors experienced difficulty deciding on ratings, it was also more likely to be the case that agencies would disagree with the assessors' rating. Agencies were not given the assessor guidelines, however. The reliability of the Instrument is tested when assessor ratings are compared with the ratings of another individual who has been similarly trained and given the same information about the assessment process. The rater reliability study presented in the next section addresses this issue.

3.5.3 Rater reliability study

The project team members conducted reliability assessments for 25 agencies. Fifteen of these involved a visit to agencies that had undergone either a joint assessment, self-assessment with verification or peer review assessment, and 10 were done as desk audits at the Institute from Instruments and documentation sent in by agencies who had undergone self-assessment but who had also been part of a sample selected later for verification by a government officer. The examination of reliability undertaken here addresses the overall consistency between ratings given by assessors and those given by a reliability rater. This consistency is compared with consistency between ratings that agencies gave themselves and ratings given to them by assessors, and between ratings given by agencies and those given by a reliability rater.

Reliability assessed by desk audit

One member of the project team conducted desk audits for 10 agencies. These agencies had completed the Instrument as a self-assessment, with a later verification interview with a government project officer (assessor). The auditor read through the agency's Instrument and the documentation supplied and determined ratings for the agency against the 27 standards, without further consultation with the agency. The desk auditor did not view the ratings given for the agency by the assessor.

The inter-rater reliability coefficient was 0.56 (the correlation between the assessor Instrument Scores and the reliability rater's Instrument Scores). The third column of Table 3.11 lists the proportion of agreement on ratings for each standard between the assessor and the desk auditor. On average, the assessor and desk auditor agreed 47% of the time. At worst, exact agreement between the assessor and desk auditor occurred for only 22% of agencies in this reliability sample.

The first column of Table 3.11 shows the proportion of agreement between the agency and the assessor and the second column of this table shows the agreement between the agency and the desk auditor. The highest level of agreement occurred between assessors and agencies' own ratings (72% agreement on average, with complete agreement on a standard across agencies occurring more than 50% of the time, except for Standard 1.5 where agreement occurred in only 38% of cases). Agreement between the agency and the desk auditor occurred on average 52% of the time, with the lowest level of agreement on a standard being 10% of the sample (i.e. agreeing with only one agency out of the 10). The lowest level of agreement occurred between the assessor and the desk auditor, (on average 47% of the time).

Table 3.11: Per cent of agreement between ratings given by the agency, the assessor and the desk auditor (n = 10)

Standard	Agency and assessor	Agency and auditor	Assessor and auditor
1.1	78	80	78
1.2	56	60	67
1.3	67	10	22
1.4	56	50	33
1.5	38	33	44
2.1	63	50	38
2.2	67	50	33
2.3	63	56	67
3.1	67	80	44
3.2	100	67	50
3.3	63	67	56
4.1	78	30	33
4.2	100	90	100
4.3	100	44	33
4.4	88	33	33
4.5	75	63	67
5.1	63	33	22
5.2	63	56	67
5.3	88	44	44
6.1	50	44	44
6.2	86	38	22
6.3	75	33	22
6.4	100	63	38
7.1, 7.2, 7.3	63	44	50
7.4	67	75	57
<i>Mean</i>	72	52	47

These results suggest that desk audits of agency Instruments have poor reliability when compared with the ratings of an assessor who has conducted a visit. Agreement between a desk auditor and either an agency or a visiting assessor will, on average, occur only half of the time. The probability of exact agreement between ratings occurring by chance is one in three, or 33%.

Reliability assessed by visit

A member of the Institute conducted reliability visits to 15 agencies. Five of these agencies had undergone a joint assessment, five had undergone a self-assessment with verification, and five had undergone a peer review assessment. Prior to the visit, the reliability rater read through the agency's Instrument (including their self-ratings) and the documentation supplied. Additional information required to score agencies was obtained during interviews with the agencies. This method was applied to all 15

agencies visited, regardless of the type of assessment they had undergone initially. The reliability rater did not view the ratings given to the agency by the assessor.

Table 3.12: Per cent of agreement between ratings given by the agency, the assessor and the visiting reliability rater (n = 15)

Standard	Agency and assessor	Agency and rater	Assessor and rater
1.1	92	93	92
1.2	77	73	62
1.3	85	67	62
1.4	92	93	85
1.5	77	67	69
2.1	62	60	69
2.2	62	67	62
2.3	85	93	75
3.1	85	64	62
3.2	69	80	85
3.3	69	80	62
4.1	77	93	77
4.2	92	80	77
4.3	75	71	77
4.4	82	71	82
4.5	69	80	77
5.1	67	64	54
5.2	85	67	77
5.3	77	80	62
6.1	62	60	69
6.2	92	83	92
6.3	100	80	75
6.4	77	73	46
7.1, 7.2, 7.3	85	53	62
7.4	83	79	69
Mean	79	75	71

The inter-rater reliability coefficient was 0.85 (the correlation between the assessor's ratings against the standards and the reliability rater's ratings). The third column of Table 3.12 lists the proportion of exact agreement on a rating between the assessor and the reliability rater. On average, the assessor and visiting reliability rater agreed 71% of the time. At worst, exact agreement between the assessor and visiting reliability rater occurred for only 46% of agencies in this reliability sample. For only two standards (5.1 and 6.4) did the level of agreement fall below 60%.

Column one of Table 3.12 shows the proportion of agreement between the agency and the assessor, and column two shows the agreement between the agency and the reliability rater. The highest level of agreement occurred between assessors' and agencies' own ratings (79% agreement on average, with complete agreement on a

standard across agencies occurring at least 62% of the time). Agreement between the agency and the visiting reliability rater occurred on average 75% of the time, with agreement on a standard across agencies occurring at least 53% of the time. The lowest levels of agreement occurred between the assessor and visiting reliability rater.

These results suggest that for assessments involving visits, the Instrument has an acceptable level of inter-rater reliability. Desk audits did not provide an acceptable level of inter-rater reliability. It should be noted that the reliability assessments were undertaken by independent raters without industry experience. From this perspective, our method provides an extremely conservative measure of rater reliability because it relies upon the Instrument providing a guide to raters without assuming prior experience. With this in mind, it appears that the consensus that is reached during discussion between an agency and someone rating them is more influential in gaining consistent ratings than simply using the Instrument as a guideline to scoring. That the agency and assessor ratings are in agreement most of the time is not surprising, however, given that many assessors indicated that they had shared the information in their guidelines with agencies during the assessment.

3.6 Summary

3.6.1 Findings

- Most frequently, agencies received ‘met’ ratings against standards, although there were some standards that could be identified as having substantially more ‘not met’ and ‘partly met’ ratings than others. Standards under Objective 7 (Advocacy) and Standard 6.4 fell into this latter category.
- There was a very high level of agreement among service providers that all of the standards and their performance information were clear. On average, standards were considered clear by 95% of service providers.
- There was a very high level of agreement among service providers that all of the standards and their performance information were desirable. On average, standards were considered desirable by 99% of service providers.
- Service providers were somewhat less inclined to agree that the standards were practical, although the overall level of support remained high. On average, standards were considered practical by 91% of service providers.
- Service providers were also somewhat less inclined to agree that the standards were appropriate. Nonetheless, on average, standards were considered appropriate by 94% of service providers. The appropriateness of standards was affected by the type of service, the type of client, and the location of the service in terms of rural/urban/remote.
- Agencies did not find the standards or performance information inappropriate when the assessment method involved agencies completing the Instrument with an assessor to assist in interpreting the standards.
- Examination of the patterns of ratings against standards revealed that some standards provided information that was measured with equivalent effectiveness

by other standards. In these cases, the redundant performance information could be dropped without adverse effect on the coverage of the Instrument.

- In general, statistical tests suggested that the standards were satisfactorily related to their relevant objectives and that ratings against individual standards could be added to compare the performance of agencies across the seven objectives.
- The performance of agencies against Standard 6.4 was considered to be insufficiently related to performance against other standards within Objective 6 to retain it as an indicator of performance under this objective. It was, however, sufficiently related to performance against other standards within the Instrument to be included in the overall Instrument Score.
- The objectives were found to be sufficiently distinct to indicate that they measured different aspects of service quality but were sufficiently related to justify adding the ratings for standards to form an overall measure of service quality in HACC funded agencies (the Instrument Score).
- Differences in rating between agencies and assessors were noted to come from three sources: the harshness or leniency of the individual rater; consistency between information supplied to agencies in the Instruments and the information supplied to assessors in the assessor guidelines; and the ability of assessors to verify the information supplied by agencies against standards. Exact agreement on ratings between agencies and assessors was moderate, being on average 66%.
- The concurrent validity of the Instrument was supported by the correlation between the global appraisal of quality and the summed standards ratings, represented by the Instrument Score. From the appraisal of assessors, the correlation between Instrument Scores and an overall appraisal of performance was 0.74. From the appraisal of agencies, the correlation between Instrument Scores and an overall appraisal of performance was 0.64.
- Standards with the weakest validation against the global assessment of quality service were those concerned with matters of release of personal information and access to personal information under Objective 5. Standards with strong validation with this measure included those related to accountable and competent management (Standards 3.1 and 3.2), those related to the provision of services to special needs groups (Standards 1.3 and 4.3) and a standard related to service delivery/ care plans tailored to consumer need (Standard 4.2). Those standards with strong validation with this measure also included those that service providers had indicated to be impractical, inappropriate or unclear (in particular, Standard 3.2 and Standard 6.4).
- Flexibility was requested of assessors in rating agencies against the standards by asking them to apply an 80/20 rule. This was suggested to have led to some inconsistency in ratings. Nonetheless, it remains important to avoid overly prescriptive directions for appraising agencies in order to maintain the applicability of the Instrument to the variety of HACC agencies. Further clarity on the weight of certain issues and the circumstances in which exceptions should be considered should ameliorate this concern.
- The three-point rating scale was criticised by some assessors as unfair from two perspectives: (1) the 'partly met' category covered a broad middle ground that

may not have discriminated between 'poor' and 'satisfactory'; and (2) there was no provision to identify those who exceed the requirements of standards.

- Some standards were not applicable to all agencies. Both assessors and agencies found it difficult to decide on ratings against these standards. Particular types of agencies were more affected by this than others. A 'not applicable' category was proposed by some assessors.
- The inter-rater reliability of the Instrument was acceptable; a coefficient of 0.85 was obtained for reliability assessments that included a visit to the agency. Exact agreement between the reliability rater and the assessor occurred, on average, 71 % of the time.
- Desk audits of agency Instruments produced the lowest reliability between raters. Agreement on ratings occurred, on average, less than half of the time.

3.6.2 Recommendations

- Guidelines for scoring should continue to be improved and address the issues that variously affect HACC agencies. This should include both assessor training and instructions for scoring tailored to specific issues such as service type and client target groups.
- The consistency between the Instrument and the assessor guidelines should be addressed by combining these two documents into one.
- The ability to verify information supplied by agencies should be improved by the changes made to the revised Instrument, and by incorporating assessor knowledge of consumer feedback into the assessment process (see Chapter 5).
- It is recommended that an Instrument be developed which specifically addresses the characteristics of Aboriginal and Torres Strait Islander people, particularly in more remote regions where indigenous culture more strongly affects service provision.
- Agencies should be clearly advised that there is no inconsistency if their service is still able to improve performance in an area where the standard is considered to be met.
- It is recommended that the three-point rating scale be retained.
- It is not recommended that a 'not applicable' category be written in to the Instrument since, for many agencies, the standards only apply with flexible interpretation. It may occur that agencies would overuse a 'not applicable' category.
- The revised Instrument (found in Appendix A) is recommended on the grounds that it minimises the amount of repetition within the instrument while still obtaining all the relevant information required to assess agencies adequately against the standards.
- Ratings against standards should be added to obtain an Instrument Score that would provide a summary measure of agency performance against the standards. The calculation of the Instrument Score is further detailed in Appendix C.
- A Compliance Indicator of agency performance against individual objectives, calculated from the relevant standards ratings, could be used to examine how

agencies perform against one objective compared to another. Objectives characterised by poorer levels of performance for a number of HACCC agencies could be identified and training and development activities appropriately targeted. The calculation of the Compliance Indicator is further detailed in Appendix C.

- The Instrument is most reliably rated by assessors who undertake a visit to the agency. The assessment of agencies should be undertaken, where possible, by verification visits.