Information Skills Survey
TECHNICAL MANUAL

First Edition
July 2005

Ralph Catts PhD

C·A·U·L
Council of
Australian University Librarians

COUNCIL OF AUSTRALIAN UNIVERSITY LIBRARIANS

Information Skills Survey

TECHNICAL MANUAL
First Edition

Ralph Catts PhD
July 2005

Publication Details:
Published by Council of Australian University Librarians (CAUL)
LPO Box 8169 (Licensed Post Office), Australian National University
Canberra ACT 2601 AUSTRALIA

About CAUL
CAUL is the Council of Australian University Librarians. Funding from CAUL supported the data collection on which this publication is based. A full list of the publications available through CAUL can be located at the following web site. http://www.caul.edu.au/

ANZIIL
The Australia and New Zealand Institute for Information Literacy supports institutions and individuals in the promotion of Information Literacy and in particular, embedding Information Literacy within educational programs. Funding from ANZIIL supported the data collection on which this publication is based.
A list of publications and activities supported by ANZIIL is available from the website at: http://www.anziil.org/

ISBN 1-876008-04-0

Acknowledgements

This report is the result of work undertaken by staff in seven Australian Universities with very limited financial support. It has been achieved because many dedicated people contributed time and energy to collect data. The members of the CAUL management committee for this project are listed in an appendix. They had carriage of the project at the various sites, and were supported by many colleagues.

Although funding was limited, the project could not have been completed without the generous support of both the Council of Australian University Librarians (CAUL), and of the Australian and New Zealand Institute for Information Literacy (ANZIIL). The support of these organisations is acknowledged with gratitude.

Special mention is made of Ms Margaret Appleton, former Librarian, Central Queensland University, who chaired the management committee. She was the driving force that enabled support across the universities involved. Ms Debbie Orr from CQU library continued this role throughout 2004 to assist in the completion of this project.

TABLE OF CONTENTS

1. SUMMARY

The Council of Australian University Librarians Information Skills Surveys (CAUL ISS) is for use to measure the Information Literacy levels of groups of students in specified academic disciplines. There are two versions of the survey available, one designed for use in the Social Sciences and related disciplines, and the other specifically for the Law discipline. This manual describes the technical characteristics of the ISS survey forms, and is published to support the use of the survey instruments in Australian and New Zealand Higher Education (HE) institutions.

The analysis has confirmed that Information Literacy as defined by the ISS is a coherent concept. The content validity of all items is confirmed which means that academics and librarians agree that the items measure the relevant Standards for which they are written. In addition, all items from all standards are positively correlated. This means that they measure one concept called Information Literacy. Each survey has been shown to have sufficient statistical coherence for use to benchmark group achievement of overall Information Literacy outcomes as specified in the Information Literacy Framework, 2nd. Edition. With the exception of Standard One for the general form of the survey, it is possible also to describe the performance of groups of students in relation to each standard with moderate levels of reliability.

The concurrent validity of both forms of the ISS has been confirmed by comparing the estimate of Information Literacy obtained from student self-reports with Librarian assessments. This means that student self-reports of their skills are consistent with the assessments made by librarians. There were some limitations with the comparability of assessments between librarians, mainly due to limited funds and hence lack of adequate training and support. However, for the Law form of the ISS, four coders out of five achieved assessments that were consistent with each other and with the self-report estimates by students. For the Education students, overall correlations between librarian assessments and student self-report estimates were satisfactory, and a majority of librarians agreed with student self ratings on all Standards.

Evidence of the predictive validity of the scales was sought by considering the correlation between Information Literacy scores and both students' entry scores and results. This assumes that the course assessment relates to the ability to use information, as opposed to simply recalling information. Zero order correlations were obtained. There is little research into the possible relationship, and further research is needed into how information literacy affects student performance.

In terms of risks of bias, there was no differentiation between results for males and females. Further research is needed to establish in particular whether first nation peoples or overseas students from NESB may be disadvantaged by the language used in the generic self-report scale. However, on the Law scale it was concluded that the ISS is not biased in relation to people from a non-English speaking background.

It is recommended that the generic form of the ISS be used with students enrolled in Social Science disciplines and that further data be collected to investigate the validity and reliability of the scale in these and other disciplines. For the Law form of the ISS, it is recommended that use be restricted to students who have completed their first semester of Law studies.

2. PURPOSE OF THE CAUL INFORMATION SKILLS SURVEY

The Council of Australian University Librarians Information Skills Surveys (CAUL ISS) have been designed for use by librarians and academics who have signed an authority to use the surveys. The purpose of the surveys is to support investigation of the Information Literacy levels of groups of students in specified academic disciplines. Information Literacy is defined by the *Australian and New Zealand Information Literacy Framework (CAUL*, 2001; Bundy, 2003). There are two versions of the survey available, one designed for use in a range of social sciences and related disciplines, and the other specifically for the Law discipline.

This manual describes the technical characteristics of the ISS survey forms, and is published to support the use of the survey instruments in Australian and New Zealand Higher Education (HE) institutions. The instruments have been released in response to the demand for assessment tools for use at Faculty and Institutional level, but prior to sufficient trialling to meet normal criteria for a standardized assessment tool. The analyses reported for each version of the survey in this first edition of the technical manual are based on one cohort of Law undergraduate respondents for the Law form, and a separate cohort of undergraduate Education students for the generic form. Further data collection and analysis is required to address the characteristics of the surveys across disciplines and cohorts of students. Hence, initial use requires caution and careful interpretation of data. As further analysis and information becomes available it will be reported on the CAUL web site[1]. Users of the surveys interested in collaborating in further research should contact either the author or CAUL.

The general version of the survey is intended for use in investigations of Information Literacy skills of large groups of students within and across programs and disciplines, especially in the social sciences and related professions. The Law version includes references to information databases and information use that are specific to the practice of Law. It is intended for benchmarking information skills and as a tool for researching the effectiveness of strategies to develop Information Literacy in the Law.

At the level of an individual learner, Information Literacy is acquired within the context of the discipline(s) in which they are engaged. It follows that assessment of individual skill in Information Literacy should be within the context of the discipline, and assessed as part of the overall discipline specific achievement of students. Data collected about the achievement of individuals within the context of a discipline has been described by Lally and Myhill (1994, 22) as ideographic assessment. Such context-based assessment precludes direct comparison of groups of individuals in different disciplines, and hence makes it difficult to assess overall improvements in Information Literacy across an institution. The CAUL Information Skills Survey is designed to provide data on information skills at either a program or an institutional level, and hence can complement the ideographic program based assessment of attainment of Information Literacy.

It is a deliberate policy decision to design these surveys in a form that is not appropriate for the assessment of individual student achievement for grading purposes. The form of student assessment can influence student perceptions of the

[1] www.CAUL.edu.au/caul-doc/publications.html

curriculum and hence influence where students focus their learning (Ramsden, (1979). An advantage of a self-report assessment tool is that it is not appropriate for reporting grades and hence cannot be used for the assessment of individuals. Given that information literacy is widely recognised as context bound, assessment of learning outcomes for individual students should be situated in the context in which they are expected to use their information skills. If a generic tool that assessed the information literacy of individuals were adopted as a global measure, and especially if government tried to hold institutions accountable for outcomes on such a measure, there is a risk that the curriculum could be narrowed to the limited range of outcomes contained in a generic assessment. This risk is avoided by the adoption of a self-report form for evaluation of programme or institutional initiatives to enhance information literacy.

These limitations do not apply when considering the overall performance of groups of students using self report. The ISS is designed for use when an evaluation of policy implementation is needed at institutional level. The ISS is designed to complement program level assessments of student information skills within their discipline context. It allows for a comparison of changes over time across programs.

2.1 Need to Measure Information Literacy Skills at Institutional Level.

The importance of achieving certain generic skills as an outcome of undergraduate education has become widely accepted in Australia and elsewhere (MacNair, 1990; Weil, 1992; Higher Education Council, 1992; Candy et al, 1994). Information Literacy has been accepted as one of the generic outcomes of undergraduate education at most universities in Australia. It is normally specified as one of the graduate attributes to be achieved through undergraduate education. Assessment of Information Literacy attainments of graduates is therefore essential both for confirming information skills at the level of the discipline, and also for confirming at an institutional level, whether policies to enhance information skills are effective. Without both monitoring implementation and the use of assessment, students, academics and librarians will be entitled to consider the institutional policies not to be deemed important by managers. Likewise other stakeholders including employers and external quality assurance agencies will not be able to observe evidence of the commitment of the institution.

The generic form of the CAUL ISS has been designed for application across a range of disciplines to complement discipline specific measures of performance across social science programs. It is intended for use in particular before and after initiatives to improve Information Literacy. It could also be applied to compare the performance of similar students in comparable courses in cooperating institutions

Further research is needed to identify to what extent the use of information in professional practice is consistent across disciplines. Preliminary evidence developed in an unpublished pilot study, indicated that in some science and technology areas including engineering, undergraduate students access and use information differently to social science students. In technology and the sciences at undergraduate level, some information may commonly be seen as absolute, whereas in the social sciences interpretation is more theory dependent, and knowledge is commonly contested. Therefore it is recommended that the utility of the generic instrument be investigated before it is relied upon as part of institutional evaluation of policy implementation,

especially in Science-based disciplines. Institutions or individuals interested in participating in further research are invited to contact the author or CAUL.

The Law version of the ISS has been designed as a discipline specific form. The use of this form will allow Law Faculties to benchmark Information Literacy outcomes within their programs for the purpose of monitoring the effects of interventions. It will also allow collaboration between Law Faculties in benchmarking activities because the form involves general questions about Law Information Skills, rather than assessing specific curriculum content that may vary between Faculties.

2.2 Uses and Misuses of Benchmarking

In Australia, benchmarking performance between universities, and across or within programs in institutions, has been widely adopted, especially in relation to the use of the Course Experience Questionnaire (CEQ) (Wilson, Lizzio and Ramsden, 1997). Benchmarks enable an institution to monitor performance against specified criteria across the institution as a whole, or across whole program areas. In the context of an increased emphasis on quality assurance, institutions utilise indicators to report the impact of policies. Where a university adopts a policy of developing graduate attributes, including Information Literacy, it follows that there will be a need to benchmark and demonstrate enhancements from policy initiatives.

It has been noted by Wilson et al (1997, 47) that the Australian Vice-Chancellor's Committee (AVCC) has expressed concern about the publishing of claims of superior performance by some universities based on CEQ scores. Selective publishing of results on individual scales has been used to promote a perception of excellence. Some institutions make such claims based on results for selected sub-scales, or even individual items drawn selectively from the CEQ. Such selective reporting of outcomes can be misleading if it is used to convey an impression of overall superior performance. The consequences could be to confuse audiences and potentially to devalue the credibility of the survey.

Unlike the CEQ, the Information Skills Survey (ISS) is designed for administration by staff in HE Institutions. In this situation, it might be relatively easy to induce false reporting by students by using exhortations about the implications for the credibility of the institution, and by implication the respondents. Therefore, a condition of use of the CAUL ISS is that results not be used for marketing and promotion. CAUL has agreed that permission to use these copyright materials will be withdrawn from any institution that uses results from the Information Skills Survey for marketing or promotion. This condition of use is essential because, as a self-report inventory, validity depends on truthful answers by respondents.

2.3 Use of Self-report

The use of self-report to verify group outcomes is widely accepted in the evaluation of higher education programs. It is the method employed in the Course Experience Questionnaire (Wilson et al, 1997) and has been supported as a valid means of gaining evidence from students by Lally and Myhill (1994, 19). Some have expressed concern that students, and especially those with limited knowledge, may rate themselves inaccurately. In the present case, this has been addressed by investigating the

correlation between self-reported Information Literacy skill, and performance as assessed in a one-hour performance-based assessment by skilled librarians. Making allowance for limitations in the available data, it was concluded that librarian assessments of information literacy are consistent with student self-reports. The analysis is reported under the heading, 'construct validity'.

Effective student assessment helps students to develop the skills of self-assessment (Ramsden, 1994), and hence a self-report form encourages students to identify their skills and areas where they may benefit from further training. However, concern about accurate rating of some items was confirmed during the trial of the Law version of the scale with first year undergraduate students. They had not encountered some of the specialist Law databases, about which questions were asked. This has been addressed by recommending that the Law scale not be used until these databases have been introduced, and students have had the opportunity to make some use of them. In the generic social science version of the survey, some items trialled but not included in the final survey form, asked questions about advanced information literacy practices that were beyond the experience of most first year students. These items were excluded, although they may have been valid for post-graduate students.

2.4 Choice of Scale

A four choice Likert Scale was adopted and all items were written to accord with a choice of responses that included absolute and qualified outcomes, namely 'never', 'sometimes', 'often', and 'always'. 'Sometimes' was defined as less than 50 percent of the time, and often as more than 50 percent of the time. Four choices forces the respondent to differentiate between their common and their less frequent practices. It avoids the risk that responses to a three or five choice scale may cluster around the central position. Hence it encourages each respondent to make a differentiation between common and less common practice. This allows more opportunity to identify actual differences in practice by each respondent. A five choice Likert scale does not provide this benefit, and a three choice scale does not accommodate either zero or constant practice from intermittent actions. A six choice scale was rejected because it was neither feasible nor necessary to make finer distinctions. The more categories that one provides, the more difficulty students have in differentiating their responses. They can tell things they do most of the time, from things they do less than half the time, but may not easily distinguish things they do more than three quarters of the time from those they do more than half the time but less than seventy-five per cent of the time. Kunkel et al (1996) showed that fifty per cent of students are seldom or never required to use library information sources to complete their assignments, and that only those required to often or always use information sources improve their information skills during their college studies. The four point scale can be used to investigate these claims in other HE contexts.

2.5 Relationship to CEQ

The Course Experience Questionnaire (CEQ) used in Australia as a survey of outcomes of university undergraduate education is composed of a series of sub-scales measuring various aspects of the experience of university students. Each sub-scale of the CEQ contains five or six items that measure a single construct. The CEQ was expanded to include a six-item sub-scale termed 'generic skills', which grouped

together a wide range of outcomes. The sub-scale includes one item for each of the following components; problem solving, analytic skills, teamwork, confidence in tackling unfamiliar situations (sometimes referred to in cognitive studies as far transfer), planning, and communication skills.

In an analysis of the factor structure of the CEQ, Wilson et al (1997, 42) reported that the generic skills sub-scale was not a coherent factor and did not improve the statistical power of the CEQ. Hence, the generic skills sub-scale does not represent a single factor, and does not provide a valid assessment approach in the multiple domains of the generic skills. Wilson et al (1997) have recommended that the generic skills subscale be discarded for research purposes. It is concluded that measures of each generic skill should be developed separately and, unless some form a coherent factor structure is identified, each should be reported separately.

The CEQ generic skills scale was developed prior to the influential report of Candy et al (1994) recommended that information literacy be included as a generic outcome of Higher Education. This may explain why Information Literacy was not included in the CEQ generic skills sub-scale. A six item Information Literacy sub-scale with a coherent factor structure could be constructed from the CAUL ISS for inclusion in the CEQ, using one item per standard. This is one of many possible developments that could arise from the research reported in this manual. In the interim, the generic form of the CAUL ISS can be used by institutions to complement the use of the CEQ, at least in social sciences and related disciplines. It allows an efficient, valid and reliable assessment of the effectiveness of institutional and/or program level interventions to enhance Information Literacy outcomes across a large group of students. In contrast, the Law version of the ISS is designed for application specifically in Law.

3. VALIDITY

An essential criterion for confidence in a survey is demonstrated validity. Four aspects of validity were identified by Lally and Myhill, (1994, 17) as necessary to establish the quality of a survey. These were:

- content validity of the survey,
- construct validity,
- concurrent validity, and
- predictive validity.

The evidence of validity obtained during the development and trial of the instruments is described in this report. The content validity was demonstrated using expert judges. Evidence of construct validity is demonstrated in the statistical analysis of the Education and Law Scales. The concurrent validity of the ISS was established by comparing student self-report on the survey with the assessments by librarians, using observation and interview. Evidence of predictive validity was sought by comparing students' level of information literacy to their subject results, but these were not found to be related. Evidence of predictive validity requires further research into how information literacy affects student performance. Assignments requiring searching for and using information may be better predictors than examinations. These investigations need to be made in a range of settings.

Marsh (1984) has suggested that an additional consideration in establishing the credibility of a survey is the "face validity" of the instrument. This means the credibility of the procedure with interested parties. In this context, he argues that ratings by students have face validity in public policy contexts because students are the primary clients. The CAUL ISS (Information Skills Survey) is designed primarily for program wide and institutional level evaluation or research into the implementation of teaching and learning strategies aimed at improved Information Literacy outcomes among students. Student ratings across programs provide a method of comparing student perceptions about the initiatives, especially when gain scores are used to determine the effects of policy implementation. The advantage of using gains is that if there is a systematic effect in students' self report that might inflate the mean scores for a program, this will be a constant factor that will not affect the difference in their estimations before and after receiving information literacy programs. Only if various information literacy programs lead to a differential reappraisal by students of their information literacy skills would the difference scores lack validity.

3.1 Content Validity

The content validity of the benchmark surveys was established against the first edition of the CAUL Information Literacy Standards (CAUL, 2001). Because the items were validated against specific elements in each Standard, it has been possible to match the survey items to the Outcomes and Standards of the second edition of the Information Literacy Standards (CAUL, 2003).

For each of the two disciplines, Education and Law, six teams of librarians and academics drawn from three universities were convened and provided with a training program to introduce or consolidate understanding of the first edition of the CAUL Information Literacy Standards. In the first edition there were seven Standards, and each Standard had embedded between two (Standard Seven) and five (Standards Three and Five) outcomes. The training consisted of material to be read before a one-hour familiarisation workshop. During the workshop participants were introduced to the seven Standards and given a framework in which to interpret the Standards. They were then given examples of student behaviour, and asked to identify which Standard was reflected in each example. The examples were selected to familiarise the participants with the seven Standards, and to highlight issues of interpretation.

Pairs of librarians and academics then drafted items that addressed each Outcome within each Standard. In the two hour drafting session, each pair of participants focused on three of the Standards and developed items that addressed each outcome for the selected standards. This intensive approach was intended to foster a richer and wider range of examples than would be generated by a superficial attempt by each pair of writers to draft items across all seven Standards in the available time.

Items were collected and edited to achieve consistent form and style. Negatively worded items were converted to a positive form where possible, and some items were deleted. For instance items that asked whether students would steal books from a library, or tear pages from a journal were not included. There were two reasons for this decision. The first was an observation from use of previous data sets that there was a systematic difference in responses to negatively framed items that could be differentiated using non-orthogonal factor analysis. This suggested that negatively

worded items might introduce a small systematic difference in responses. The second reason was the observation that students in a previous study, who had completed a similar survey and were interviewed, said that they had learnt about information skills from responding to the survey. It was decided to avoid the risk that participants might learn unacceptable practices from the survey.

The edited items were presented to the same groups of academics and librarians in subsequent workshops conducted a month or so after the initial sessions. The participants were re-familiarised with the standards, and then presented with items written by other item writing teams. They were asked to validate each item by identifying the standard and outcome that each item referenced. This procedure is a recommended method of demonstrating content validity in the published standards for Educational and Psychological Assessment (APA, 1999, part 1). Each item was assessed independently by at least two teams including one academic and one reference librarian from the relevant discipline. Only those items for which there was agreement on the Standard and its Outcomes were accepted as content valid. The criterion applied was that at least two of either two or three teams of coders agreed on the assignment of the item to a Standard and Outcome. Examples of accepted items are presented in Table One. As can be seen, most Law items were explicitly referenced to the discipline, while the Education items were deliberately written to serve as generic items across a wide range of Social Science disciplines.

TABLE ONE: Examples of Validated Items.

Standard	Discipline	Item
1.2	Law	When my course materials are not sufficient to answer an assignment, I search for more information.
2.1	Law	Because the law is constantly changing, I use the library to access up- to- date legal information.
3.2	Education	I can pick out the main ideas in an article.
4.1	Education	I keep accurate details of everything I read.
5.1	Law	When presented with a lead issue, I can determine the relationships between key elements.
6.3	Education	I only list resources I have actually referred to in my assignment bibliography.
7.2	Law	I assess the impact of new law on current civil rights and democratic institutions.

Items were prepared for all outcomes across all Standards. For Education, a total of 258 items were submitted for content validation and 124 were confirmed. For the Law discipline, a total of 255 items were submitted for validation of which 110 were confirmed. Education discipline experts, validated no items for four of the twenty-six outcomes across the seven Standards, and insufficient items (defined as less than 20%) for a further four Outcomes. Among Law discipline experts, no items were validated for three of the Outcomes, while less than 20% of items were validated for a further four items. The results of this content validation are summarised in Table Two. Only items that were content validated were accepted for inclusion in the subsequent analysis.

TABLE TWO: Item Validation Findings for Education and Law

Standard and Outcome	**Education**			**Law**		
	Number of Items Written	Number of Items Validated	Proportion Validated	Number of Items Written	Number of Items Validated	Proportion Validated
1. The information literate person recognises the need for information and determines the nature and extent of the information needed. The information literate person:						
1.1 defines and articulates the need for information	6	5	.83	9	3	.33
1.2 understands the purpose, scope and appropriateness of a variety of information sources	14	8	.57	9	5	.56
1.3 consciously considers the costs and benefits of acquiring the needed information	7	6	.86	9	2	.22
1.4 re-evaluates the nature and extent of the information need	6	0	**0***	7	2	.29
2. The information literate person accesses needed information effectively and efficiently. The information literate person:						
2.1 selects the most appropriate investigative methods or information access tools for finding the needed information.	8	3	.38	8	7	.88
2.2 constructs and implements effectively designed search strategies	14	7	.50	14	7	.50
2.3 retrieves information using a variety of sources	8	4	.50	10	10	.99
3. The information literate person evaluates information and its sources critically and incorporates selected information into their knowledge base and value system. The information literate person:						
3.1 assesses the utility of the information accessed.	13	0	**0**	4	3	.75

Standard and Outcome	**Education**			**Law**		
	Number of Items Written	Number of Items Validated	Proportion Validated	Number of Items Written	Number of Items Validated	Proportion Validated
3.2 summarises the main ideas extracted from the information gathered.	10	3	.30	5	2	.40
3.3 articulates and applies initial criteria for evaluating both the information and its sources.	12	17	**1.42**[*]	8	8	.99
3.4 validates understanding and interpretation of the information through discourse with other individuals, subject area expertise and / or practitioners.	14	10	.71	13	4	.31
3.5 determines whether the initial query should be revised	9	4	.44	6	0	**0**

Continued next page

[*] In the first edition, Standard One, Outcome Four referred to the process of "re-evaluating the nature and extent of the information needed". For Education, all items that were written were validated consistently as belonging to Standard Three, Outcome Three, which stated "applies initial criteria for evaluating both the information and its sources".

4. The information literate person classifies, stores, manipulates and redrafts information collected or generated, The information literate person:						
4.1 extracts, records, and manages the information and its sources.	16	8	.50	8	4	.50
4.2 preserves the integrity of information resources, equipment, systems and facilities.	12	1	**.08**	5	0	**0**
4.3 legally obtains, stores and disseminates text, data, images, or sounds.	8	1	**.13**	11	4	.36
5. The information literate person expands, reframes or creates new knowledge by integrating prior knowledge and new understandings individually or as a member of a group. The information literate person:						
5.1 applies prior and new information to the planning and creation of a particular product	6	0	**0**	9	1	**.11**
5.2 synthesises main ideas to construct new concepts	6	1	**.17**	10	6	.60
5.3 compares new understandings with prior knowledge to determine the value added, contradictions, or other unique characteristics of the information	10	3	.30	11	2	**.18**
5.4 revises the development process for the product	3	-	**0**	8	0	**0**
5.5 communicates the product effectively to others	11	7	.64	9	4	.44
6. The information literate person understands cultural, economic, legal, and social issues surrounding the use of information and accesses and uses information ethically, legally and respectfully. The information literate person:						

6.1 understands cultural, ethical, legal and socio-economic issues surrounding information and information technology	6	5	.83	16	5	.34
6.2 follows laws, regulations, institutional policies, and etiquette related to the access and use of information resources	13	8	.62	15	6	.40
6.3 acknowledges the use of information sources in communicating the product	9	5	.56	8	11	.73
7. The information literate person recognises that lifelong learning and participative citizenship requires information literacy. The information literate person:						
7.1 appreciates that information literacy requires an ongoing involvement with learning and information technologies so that independent lifelong learning is possible	17	15	.88	21	10	.48
7.2 determines whether new information has implications for democratic institutions and the individual's value system and takes steps to reconcile differences	20	3	**.15**	22	4	**.18**
TOTALS	258	124		255	110	

As has been discussed elsewhere, (Catts, 2002) the difficulties in confirming the content validity of items intended for some elements of the Standards is likely to reflect areas of ambiguity in the first edition of the Standards. In the revision of the Standards, these ambiguities were identified and addressed**.**

4. EXPLORATORY DATA ANALYSIS

The statistical model adopted for validating each of the six sub-scales (representing the six standards) is congeneric factor analysis (Holmes-Smith and Rowe, 1994). This is a form of structural equation modelling. This technique allows an exploration of the extent to which items share common variance. Those with the most common (or shared) variance provide a cluster of measures of the latent variable which each item is designed to measure.

Six Standards are defined for the Information Literacy Framework (Bundy,2003). Congeneric factor analysis uses the statistical characteristics of a set of content valid items to measure the extent to which each item shares common variance with all other content valid items. The technique therefore complements the evidence of content validity derived from expert judgement, by adding evidence of construct validity. This adds statistical evidence that the items form a coherent construct of each Standard, and hence adds evidence of the quality of the items. Congeneric factor analysis is powerful because, for the population on which the data is collected, it allows selection of the content valid items that provide the best statistical evidence for each Standard.

Before undertaking statistical analysis it is normal and proper practice to consider the quality of the available data. One of the most important considerations is the effect of non-responses by students. Missing values require careful consideration in structural equation modelling. Byrne (2001, 287) has pointed out that missing values can provide useful information about the nature of a construct and how it is defined by the observed variables, which in this case are the content valid items. This is the case especially when missing values are considered in the context of reports by those responsible for the collection of the data. Important issues in deciding how to treat missing values include the number of missing values, the distribution of missing values, and the likely reasons for the missing values.

For some items in each data set collected there were more than fifteen percent of responses missing, which was more than was expected. The explanation for this situation is as follows. This project developed through a series of discussions. An early decision was to collect data from first year undergraduates, partly to set a benchmark about what people knew about Information Literacy when they started higher education studies, and partly because the largest and perhaps most amenable groups of students are found in first year classes. A later decision was to write items to address skills across the range of experiences of undergraduate students from their initial year of HE through to their final year. This was considered necessary to provide a tool for monitoring progress in Information Literacy during undergraduate studies. Had we had the time, resources and wit to trial the items with a small number of students, we might well have realized that some items were beyond the experience of virtually any first year undergraduate. In retrospect, this decision should have led to changes in the measurement model and the data collection. The model used was

criterion-referenced measurement, and this would have been appropriate for use with students who were expected to have achieved an agreed criterion for undergraduate Information Literacy. In retrospect, it is concluded that a latent scale model is more appropriate for measurement of change during undergraduate studies. However, working on a miniscule budget, and with staff who contributed in addition to their normal duties, this reconsideration did not occur until after the data was collected. Nonetheless, exploratory data analysis allowed the missing values issues to be addressed in a way that enabled scales to be generated.

As there were a significant number of missing values on some items, an item by item exploratory review was undertaken. This included statistical analysis of the responses for each item, reports by data collectors, and an analysis of the assumed knowledge on which each item relied. This review suggested a combination of reasons for missing values. For a few items, data collectors reported that students said that they had no experience of the issue referred to in the question. One such case was an item that asked if they included a copyright restriction when they published original work. It is hardly surprising in retrospect that some students chose not to respond to this item. Therefore, for some items, data is missing for reasons that are independent of responses on other items, but not independent of other responses on the specific item. The 'missingness' in these cases may raise a question as to the trustworthiness of the responses given by other students for the specific item. It is, therefore, legitimate on the available evidence to not include this item in the statistical analysis.

One student, having signed the consent form for data collection, changed her mind and asked for her responses to be returned. She had misgivings about revealing her practices that contravened university rules, notwithstanding guarantees of confidentiality. It is possible that a few students addressed this issue by either omitting such items, or by not answering further items after encountering one about which they had reservations. Where a student omitted a large number of items, the missing values were attributed to a case-specific response to the survey. It was decided that any individual case with more than ten missing responses out of forty-two was to be deleted from the analysis. This arbitrary figure was chosen because it differentiated between the majority of cases in which students had omitted between zero and four items, and the exceptions[2] where more than twenty responses were missing.

Commonly, a student would have one or two missing responses across the full set of items. As a result, across more than two hundred cases, any of the forty-two items could accumulate up to four missing values. In these cases, it was assumed that the most likely explanation was that the missing values occurred at random for the item. It is possible that the missing values could be predicted by reference to the performance of the student in comparison to others on other items. However, it could not be determined whether or not these missing cases were missing completely at random (MCAR), or simply missing at random (MAR). Byrne (2001, 288) defines MCAR as a situation where the missingness is independent of both the missing and the recorded values of all other observed variables in the data set. This would be the situation if the student simply missed a question unintentionally. She distinguishes

[2] For the Education students, one had many missing responses. There were six Law students in this category.

this situation from the missing at random (MAR) condition where the missingness is not independent of the observed values for the individual on other variables, in which case it is linked to the responses given to other variables in the data set. For instance, this would be the situation that prevailed when students chose not to answer questions that they felt did not apply to their studies to date.

Uncertainty about the nature of the isolated missing values is why it was decided not to use regression or maximum likelihood predictions of isolated cases of missing values. This is a judgement based on previous experience with missing cases in data sets, that led to the conclusion that the score predicted by regression, is not necessarily valid for determining the unobserved score for an individual case. This judgement implies that missing cases caused by chance events (MCAR), are more likely than cases where the MAR model is the explanation.

Rather than predict the missing value from regression scores, the median value was selected to replace up to four missing values for an item. This has the advantage that it is the most likely score if the data element is MCAR. The median was used instead of the mean, because the data is integer, and a few cases in decimal form would confound the asymptotic solution of the Structural Equation Model (SEM), which depends on polychoric correlations. As Byrne (2001, 291) points out, the use of a measure of central tendency will shrink the variance of the distribution for the item. As a consequence, the correlation of the variable with others in the model will be reduced. This is a conservative solution, because it does not inflate the correlations. Only when the central tendency estimate is applied to a substantial number of cases would this problem be of practical concern, and this was not the case in this analysis. However, if regression is used to predict scores when this is not appropriate, it could have an unpredictable effect on correlations and variance.

The initial stage of the exploratory data analysis was the application of exploratory factor analysis to the set of items. A primary information literacy factor was identified in each of the Law and Education sets by the items that load positively on the primary factor. The procedure also allowed a consideration of the extent to which secondary factors were related to the separate Standards, as intended by the design of the survey.

Liberal criteria were established for initial inclusion of items for the second phase of the analysis. The criteria were that the item loaded more than 0.4 on the factor, had no cross-loadings greater than 0.3, and that the primary loading was at least 0.2 larger than any cross-loadings. The aim was to try to obtain information about patterns of data that could inform subsequent decisions about the items to be accepted to specify the sub-scale for each of the six Standards.

A liberal interpretation of the criteria for inclusion of items in congeneric factor analysis is justified because all items considered had demonstrated content validity (i.e. were verified as measuring the relevant Standard), and the assessment tool is intended to reflect the scope of the Standards. It is justified also because the research is exploratory and intended to provide insights for more detailed studies to follow.

5. STATISTICAL ANALYSIS OF THE EDUCATION SCALE

A set of forty-two items from the full set of content valid generic items were administered to students enrolled in a first year core unit in the Faculty of Education at Central Queensland University. Two hundred and twenty-five students signed the consent form and completed the survey. After preliminary data analysis one case was deleted because of substantial missing data, leaving two hundred and twenty- four cases for analysis.

Exploratory factor analysis was performed using principal moments, and produced fourteen factors with Eigen vectors greater than one. The seven largest factors were discernable from the remaining factors by examination of the scree plot. Thirty-two items had loadings on the first factor greater than 0.3. None of the other factors were clearly defined because there were numerous cross-loadings.

To address the parameters of the revised Information Literacy Framework (ILF) (Bundy, 2003), the five items that were content validated against the deleted Standard Seven were excluded. This left thirty-seven items for analysis. To address the six standards in the ILF, a six factor solution was explored, using principal components to determine the extent to which the data conformed to the six components in the ILF. The six factor solution produced an acceptable level of resolution with a KMO test of 0.78, using both replacement of missing values with the mean, and using pair wise reduction. Six factors could be distinguished, both visually from the scree plot and from the Eigen vector values. The Eigen vectors for the six factors ranged from 6.4 to 1.5, and explained forty-one per cent of the variance. Twenty-four of the items had component loadings greater than 0.3 on the first factor, and all items loaded positively on the first factor. This is consistent with the view that the six Standards in the ILF describe various aspects of a single construct called 'Information Literacy'. It supports the theoretical stance developed by Bruce (1997) and known as the relational model. In this model it is postulated that the various elements of information literacy are used in an iterative rather than a sequenced or linear manner. This same position was adopted in the development of the ILF (Bundy, 2003).

Varimax and oblim rotations were performed with the factor analysis to explore the correlated nature of the constructs in the Standards. The results indicated that both Standard Three and Six were clearly identified. For Standard Three, four items loaded on one factor with a fifth item also loading but not within the criteria for inclusion specified above. For Standard Six, six items loaded on one factor. For the other four Standards, the analysis produced less clearly defined factors. The six items for Standard One were divided between two factors. Two of three items met the criteria for inclusion for each of these factors. There were four items from Standard Five out of a total of nine items that loaded on the first factor, with the remaining items for this factor drawn from Standards Two and Three[3]. Hence the problem of multiple cross loadings in the data affected primarily Standards One, Two, and Five. This was similar to the outcome of the analysis for Law items as reported in the next section, where Standards One, Three and Five were confounded across factors.

[3] In the content validation exercise, items written for Standard One, Outcome Four from the first edition of the Standards were consistently validated as belonging to Standard Three, Outcome One (see Table Two). Hence, additional items were available for Standard Three. In the second edition of the Information Literacy Framework, Outcome Four was deleted from Standard One.

To address the cross loadings, a five factor solution was attempted. Forcing items into five factors helped clarify the patterns in the data. For the five factor solution, Standards Three and Six were again clearly distinguished. The four items with strongest loadings on the second factor were all from Standard Three, with loadings using the oblim analysis ranging from 0.75 to 0.48. All however had substantial (>0.2) cross loadings on other factors. For Standard Six, five items loaded on one factor, of which two had no substantial cross loadings, and a further three items[4] from Standard Six loaded with less satisfactory primary and cross loadings. Items drawn from Standard Five were once again dominant in the first factor. Of the eleven items that loaded at greater than 0.3 on this factor, five including three of the top four were from Standard Five. One of the Standard Five items had no substantial cross loadings. Other items loading on this factor were drawn from across Standards One, Two and Three. For the fourth factor, three of four items were drawn from Standard One, although most loadings were weak. For the fifth factor two items representing Standard Four, Outcome One, met criteria, with two other items from Standard Four showing weaker primary loadings and substantial cross loadings. Thus the five factor solution helped to distinguish all Standards except Standard Two. In a five factor solution, one of the six Standards is necessarily forced to load onto other factors. In summary, Standards Three and Six were clearly distinguished, Standards Five and Four were discernable, and Standard One was identifiable, although poorly specified. On the basis of these results, the items identified above from the exploratory analysis were used in the confirmatory analysis for Standards Three, Four, Five and Six. For Standards One and Two, all content valid items were included.

5.1 Analysis for Standard One

Given the poor specification of Standard One in the exploratory factor analysis, limited prospects were held for identifying a satisfactory model, with an acceptable level of reliability. It was found that the seven items generated a solution for congeneric factor analysis with a sound overall factor structure. The Kronbach Alpha reliability coefficient was 0.46, accounting for about twenty per cent of the variability in the data. One item dominated the factor loadings with values between two and three times the other six items. When this item was removed, the reliability estimate for the remaining six items fell to 0.35, which indicated that a four or six item solution would need to be built around the strongest item. This item addressed awareness of the utility of journal articles. By examining the alpha estimates when each item is deleted, solutions with 4 items were examined which included the strongest item. The best two four item solutions had reliability coefficients of 0.36 and 0.35. These two four item solutions were compared for goodness of fit. While both models met the criteria specified for a good fit by Hu and Bentler (1995) and Holmes-Smith (1999), the second option had better results on all selected indicators, with the exception of the AIC, for which the difference was relatively small. The goodness of fit data for the second option are consistent also with the goodness of fit data for the seven item model. The congeneric factor analysis confirm that items that were content validated for Standard One produce a good fitting model. These results are provided in Table Three.

[4] The items included two from what had been Standard four, Outcome three in the first version of the Information Literacy Standards. This added supporting evidence for the decision to relocate this Outcome to Standard six in the second edition of the Information Literacy Framework (Bundy, 2003).

TABLE THREE Goodness of Fit Statistics, 4 and 7 Item Models, Standard One

Fit Statistic	Critical Value	7 item model	4 item model chosen option	4 item model 2nd Option
Chi Square	p>0.05	0.729	0.249	0.823
Normed Chi Square	1<NCI<2	0.746	1.39	0.19
RMSEA	<0.05	0.000	0.042	0.00
NFI	>0.95	0998	0.999	1.00
TLI	>0.95	1.001	0.999	1.003
CFI	>0.95	1.000	1.000	1.00
AIC	Lowest	52.41	26.78	24.40

It was evident that the content validity of these items was confirmed. However, the lack of a coherent loading of these items on a single factor in the exploratory data analysis remained a problem, and the reliability coefficient was not adequate for the purpose of identifying changes in performance for Standard One, because with a value of 0.35 it accounts for less than one eighth of the variance.

5.2 Analysis for Standard Two

For Standard Two, three of the five items loaded with the five items from Standard Five on the first factor in the exploratory factor analysis. This indicated that there was substantial co-variance between items related to information search strategies (Standard Two) and information use (Standard Five). There is a post-hoc rationale for this relationship if it is accepted that better use of information is inter-dependent with better searching for information. Hence there were prospects of identifying a coherent sub-scale with a higher reliability than emerged for Standard One.

There were only two cases of missing data involving one item. The median value was substituted for these two cases. The distributional properties of the five items were considered. Two items had a range of three, and in one of these cases, over sixty per cent of respondents selected the option, ‘always’. The other three items each had the maximum range of responses. Of these three, two had an adequate spread of scores, but the other item had just three responses for the option ‘never’.

The initial analysis of the five item set resulted in a good fit with all selected indicators. Examination of the factor weight scores indicated that the item with two missing values, provided a very low factor weight (0.03) compared to the other four items. Three of these items loaded between 0.093 and 0.15, while one item dominated the weightings with a factor weight of 0.386. To produce a four item model to estimate Standard Two, the item with the lowest factor weight loading was deleted. Further evidence that this item was not making a contribution to the model were the low inter-item correlations, and the finding that the Kronbach alpha estimate of reliability rose from 0.52 for all five items to 0.54 for the four item solution excluding this item. Thus the four item model accounts for approximately thirty per cent of the variance. The results for the four item model and for the asymptotic solution are reported in Table Four.

TABLE FOUR Goodness of Fit Statistics for Standard Two

Fit Statistic	Critical Value	5 item model	4 item model	4 item asymptotic
Chi Square	p>0.05	0.387	0.777	0.793
Normed Chi Square	1<NCI<2	1.048	0.252	0.232
RMSEA	<0.05	0.015	0.000	0.000
NFI	>0.95	0.998	1.000	0.991
TLI	>0.95	1.000	1.003	1.106
CFI	>0.95	1.000	1.000	1.000
AIC	Lowest	35.239	24.504	16.464

These results indicate that the four item solution satisfies all selected goodness of fit criteria. The result is confirmed in the asymptotic model, with all selected indicators satisfying the criteria for good fit, with the qualifications that the NFI is below the optimum criterion and the TLI, an incremental fit index, is exceptionally high. These results are likely to be due to the non-normal item distributions. The factor loading weights for the four item model were uneven, with the dominant item weight at 0.403 and the other three items loading at 0.143, 0.122, and 0.090. This is an indication that a better prediction for Standard Two can be obtained by using the factor weightings to compute a composite score and a composite score reliability estimate. This has not been undertaken for this report.

5.3 Analysis of Standard Three

There were seven items available for Standard Three. Only one missing value was recorded on an item for which the mean was 2.98 and the median 3. The missing value was replaced by the median for the item, and a maximum likelihood congeneric factor analysis was undertaken on the full seven item set. The initial solution was not a good fit with an unsatisfactory chi square (p=0.002). Both the normed chi square (2.411) and the RMSEA value of 0.08 were in the range for adequate as opposed to good fit as defined by Hu and Bentler (1995) and Holmes-Smith (1999).

Recommended modification indices were provided implicating covariance of error estimates for some items. After three iterations to include covariance of error estimates for three pairs of items, a model was generated that satisfied the good fit criteria for all the selected indices. With each iteration, all indices improved and the AIC decreased, indicating that allowing for each co-varied pair of error terms enhanced the model. The seven item solution with three co-varied pairs of error terms is reported in Table Five.

Examination of the factor loadings for the seven item model indicated a dominant item, and three that were weak. The highest loading was 0.263 and the weakest items had loadings of 0.017, 0.053 and 0.064. Hence the highest loading item had a weight five to ten times greater than the weakest items. The items with low weightings were also involved in the first two of the model modifications needed to produce a good fit. Therefore these three items were deleted to generate a four item model.

The four items selected were those identified as loading on a single factor in the exploratory factor analysis. For this model some goodness of fit indices were not satisfactory for the initial maximum likelihood solution. In particular, the RMSEA of 0.089 was outside the recommended boundaries for adequate fit, and the normed chi square of 2.776 was in the adequate, as opposed to good fit range. However, co-varying the error term for one pair of items produced a tight fit with the criteria. This was confirmed by the asymptotic estimation, which is appropriate with ordinal data. The results for the co-varied four item solution are also reported in Table Five.

TABLE FIVE Goodness of Fit Statistics, for Standard Three

Fit Statistic	Critical Value	7 item, error co-varied	4 item co-varied	4 item co-varied asymptotic
Chi Square	p>0.05	0.296	0.918	0.923
Normed Chi Square	1<NCI<2	1.178	0.011	0.009
RMSEA	<0.05	0.028	0.000	0.000
NFI	>0.95	0.997	1.000	1.000
TLI	>0.95	0.999	1.004	1.071
CFI	>0.95	1.000	1.000	1.000
AIC	Lowest	60.958	26.011	18.009

The factor loading weights were 0.308, 0.271, 0.123 and 0.122. While not equal, the differential between the strongest pair and the weaker pair of items was between 2 and 3, which is a substantial improvement on the case for the seven item model. The Kronbach alpha reliability for the four item model was 0.78. This value is nearly at the level acceptable for estimating a single factor, which is generally set at 0.80. The Kronbach Alpha estimate represents the lower boundary of the reliability of the scale, (Holmes-Smith and Rowe, 1994). The value reported indicates that more than sixty per cent of the variance in the estimation is accounted for in this sub-scale. Given the differential factor loadings, a weighted composite for this Standard is expected to have a higher composite reliability estimate, but has not been reported. It is concluded that Standard Three is well specified by the four selected items, and therefore the scale can be used to estimate performance of individuals and groups for this Standard.

5.4 Analysis of Standard Four

There were six items available for Standard Four of which four items had no missing values. These four items loaded on one factor in the exploratory factor analysis. The other two items (5 and 12) did not load substantively on this factor. There were five missing values on one of these items, and nineteen on the other item. Two cases had missing values on both items and these two cases were deleted. This left three missing values on one item, and these were replaced by the median value. There were still seventeen missing values on the remaining item, and consequently two data sets were established. In one data set, all seventeen cases were deleted, making a data set with 203 cases. A separate data set was saved in which the median value was substituted for all seventeen missing values, providing a data set with 220 cases. The analysis proceeded using the two data sets in parallel to monitor the consistency of results. In structural equation modelling, sample size is crucial in the stability of goodness of fit

parameters (Hu and Bentler, 1995, p95), but replacement by median values reduces the variance, with consequences for estimates. This issue was discussed in the introduction to this section. Further details are provided by Byrne (2001, 287 cf).

The initial six item solutions provided satisfactory goodness of fit statistics. All selected indices met the good fit criteria, with the exception of RMSEA. For this index of fit the values were in the acceptable range but larger than the 0.05 best fit criterion. For the 203 item model with cases deleted, the RMSEA index was 0.06, while for the 220 item model, a value of 0.052 was obtained. The only modification identified in each case was the covariance of error scores on the two items with highest factor weights. The results with this modification for both methods of estimating the six item model are reported in Table Six. All fit criteria were similar in the two models and, with the inclusion of covariance for error scores on two items, both satisfy all good fit criteria. The confirmation of the six item model under both approaches to treatment of missing values provides assurance that the solution is stable. The Kuder-Richardson Alpha Coefficient of reliability provides a minimum estimate of the true reliability for the weighted composite estimation of reliability. For the six item model, the estimation was 0.54, for both the model with 17 missing values on item 5 replaced by the median, and for the model with these cases deleted.

TABLE SIX: **Fit Statistics, six item models for Standard Four.**

Fit Statistic	Critical Value	6 items, with error co-varied (203 cases)[5]	6 items, with error co-varied (220 cases)[6]
Chi Square	p>0.05	P = .401	P=.489
Normed Chi Square	1<NCI<2	1.042	0.932
RMSEA	<0.05	0.014	0.000
NFI	>0.95	0.997	0.998
TLI	>0.95	1.000	1.000
CFI	>0.95	1.000	1.000
AIC	Lowest	46.335	45.453

A four item model was also analysed. This was established by deleting the two items with missing cases. Therefore, it was not necessary to replace missing values. The analysis used the 220 cases data set[7]. A summary of the results for goodness of fit criteria for the 220 cases model is reported in Table Seven. The result for the asymptotic solution is also reported. The results confirm that the four item model satisfies all fit criteria, with the RMSEA falling into the good fit level with the asymptotic solution (0.041), and into the satisfactory criteria level for the maximum likelihood estimation. As the asymptotic solution is appropriate with polychoric correlations, the enhancement of the fit with this solution supports the acceptance of the 4 item model. However for this model the results for the NFI and the TLI indices were within the adequate range, rather than satisfying the best fit criteria.

[5] Cases with missing values deleted.
[6] Missing values replaced by median value.
[7] It is noted that the full 222 cases could have been used as there were no missing values. It makes no practical difference to the results.

TABLE SEVEN: Fit Statistics, four item model for Standard Four

Fit Statistic	Critical Value	4 item solution	4 item asymptotic
Chi Square	p>0.05	P=.183	P=.258
Normed Chi Square	1<NCI<2	1.696	1.373
RMSEA	<0.05	0.056	0.041
NFI	>0.95	0.999	0.936
TLI	>0.95	0.997	0.939
CFI	>0.95	0.999	0.998
AIC	Lowest	27.392	18.746

The factor loadings for the four items using the asymptotic estimation are in the range 0.095 to 0.189 and thus vary by a factor of two. Hence, while the weighted solution will provide an advantage, an unweighted solution is a satisfactory approximation of the model. For the four item model, the estimate of reliability was 0.58, which is higher than that obtained for the six item model. Hence, the four item model provides a basis for identifying performance on Standard Four, with more than thirty-five percent of the variability across the group explained by the four item model.

5.5 Analysis of Standard Five

For Standard Five, all five items written and content validated against the Standard loaded on a single factor in the exploratory data analysis, but only one item had no substantive cross loadings. Of the other four items, two met the criterion with a primary loading greater than 0.5 and a minimum difference of 0.2 with a cross loading on a second factor. The remaining items had primary factor loadings of 0.478 and 0.430 and cross loadings that exceeded the 0.2 gap criteria. Therefore all items were included in the initial congeneric factor analysis, in order to determine whether a four item solution with satisfactory statistical characteristics could be identified.

The initial solution for the five item model provided goodness of fit statistics that were adequate according to Hu and Bentler (1995) and Holmes–Smith (1999), but the RMSEA of 0.064 was above the 0.05 criterion for an indicator of good fit, and Chi Square and normed chi square estimates were near the respective good fit boundaries. Factor loading weights for all five items were of the same order, ranging from 0.107 to 0.146, indicating that the solution approximated an additive model.

The modification indices identified covariance between the error terms for two items. As reported in Table Eight, with this adjustment, the fit statistics satisfied all criteria with the TLI exceeding 1, and the AIC reduced from 39.5 to 33.4, indicating an improved solution. The factor score weights for the two items with co-varied error declined to 0.074 and 0.081, while weightings on two of the other items increased to 0.178 and 0.164, a ratio of more than two to one.

The pair of items for which error was co-varied in the five item solution, both addressed the same outcome in the first edition of the Information Literacy Standard (Outcome Five). The solutions were compared for four item models generated by the deletion of each of these items in turn. Having regard first for goodness of fit indices, the objective was then to choose the solution for which the range of factor score

weights was least, thus providing the better unweighted solution. The range of the factor weightings for the two, four-item models were 0.90 and 0.72. The solution with the smaller range of factor weightings also had a better AIC estimate (difference of 0.85) indicating a marginally better solution. There were no substantive differences in the fit indices for the two 4 item models. Each exceeded the good fit criteria for all indices. The results for the preferred model are also reported in Table Eight.

TABLE EIGHT: **Goodness of Fit Statistics, for Standard Five**

Fit Statistic	Critical Value	5 Items, co-varied error	Preferred 4 Item solution	Preferred 4 Item asymptotic
Chi Square	p>0.05	P=0.851	P=0.939	P=0.961
Normed Chi Square	1<NCI<2	0.34	0.063	0.04
RMSEA	<0.05	0.000	0.000	0.000
NFI	>0.95	1.000	1.000	0.998
TLI	>0.95	1.003	1.003	1.128
CFI	>0.95	1.000	1.000	1.000
AIC	Lowest	33.359	24.172	16.080

The alpha reliability coefficient for the five item model was 0.64. This was better than either of the four item solutions, which produced estimates of 0.60 for the preferred model, and 0.59 for the alternate four item solution. For the preferred four item solution thirty-six per cent of the variance is accounted by the model.

5.6 Analysis of Standard Six

For Standard Six, five of the six items written and content validated against the standard loaded on a single factor in the exploratory data analysis. The remaining item did not have any factor loadings greater than 0.2. In addition, one item originally validated against Standard Four Outcome Three in the first edition of the Standards, was included[8] because it loaded with satisfactory criteria on the factor identified with Standard Six in the exploratory analysis. This provided six items for consideration. Four items had zero, one or two missing values, one had five missing values, and one had eleven values missing in the two hundred and twenty-two usable cases. Two cases were deleted that had two missing values for two out of the six items. This reduced from five to four the missing values on one item, and to zero or one the number of missing cases for a further four items. For these items, an estimate of central tendency was used to replace the missing values. In all but one case, the median was the value selected. An exception was made for one item where the median was also the maximum score of 4, but the mean was 3.31. Given that the mean value was closer to 3, this value was used to replace missing values for this item.

Surprisingly, the remaining item still had eleven missing values. Therefore two data sets were created to check the stability of the six item solution. One replaced all eleven missing values with the median value for the item, and in the other all cases with missing values on this item were deleted. There were 220 cases where all

[8] In the Information Literacy Framework, 2nd Edn., this Outcome was included in Standard Six.

missing values were replaced by an estimate of central tendency, and 209 cases when cases with a missing value were deleted.

The congeneric factor analysis for the six items, using the data set with missing values for one item deleted (209 cases), provided a model with good performance against all the goodness of fit statistics, without any model modification. For the larger data set with median replacement (220 cases), the result was replicated with marginally better goodness of fit indices. This is consistent with the exploratory factor analysis, for which all six items loaded on one factor. The selected goodness of fit indices for the six item solution, using the reduced data set, are provided in Table Nine.

The six items divided into two groups of three when factor score weights were considered. One group of three items had weights in the range 0.078 to 0.095, while the other three had weights between 0.269 and 0.274. Similar results, but with a wider range within the two groups of factor weight loadings, were obtained for the data set with missing values replaced. There was little to differentiate the three items with lower factor weight loadings, and consequently the impact of deleting each pair of these items on the Kuder-Richardson Alpha reliability coefficient was examined. By deleting one pair of items, the value slightly exceeded the value for the six item solution, whereas for the other options it declined. The four items selected provided at least one item for each of the three outcomes for Standard Six, thus maintaining the content validity of the scale. The goodness of fit data for the 209 case, four-item solutions are also presented in Table Nine. The results for the full set were consistent, but the factor weight loading for the item with missing values increased with the full set. Therefore the factor loadings derived for the 209 cases solution is the preferred option if estimating factor loadings for a composite score.

TABLE NINE: **Goodness of Fit Statistics, for Standard Six**

Fit Statistic	Critical Value	6 Items, 209 cases	4 Item solution	4 Item asymptotic
Chi Square	p>0.05	P=0.178	P=0.152	P=0.351
Normed Chi Square	1<NCI<2	1.408	1.882	1.046
RMSEA	<0.05	0.044	0.065	0.015
NFI	>0.95	0.997	0.999	0.923
TLI	>0.95	0.998	0.997	0.987
CFI	>0.95	0.999	0.999	0.996
AIC	Lowest	48.672	27.763	18.092

The Kuder Richardson Alpha reliability estimate for the preferred four item solution was 0.554, while for the six item solution it was 0.547. Thus more than thirty per cent of variance is accounted for by both versions of the sub-scale for Standard Six.

5.7 Summary of Analysis for Standards One to Six

In summary, it has been possible to select four items from the content valid item pool for each of the six Standards, and hence to construct an Information Literacy Scale referenced to the Information Literacy Framework (Bundy, 2003). However, as

reported in Table Ten, the reliability coefficients for each sub-scale vary from a substantial 0.78 for Standard Three, to a relatively low 0.35 for Standard One.

The result for Standard Three is exceptional. It accounts of over sixty percent of the variance and hence this sub-scale can be used for both group estimates and for diagnosis of individual students. For Standards Two, Four, Five, and Six, the sub-scales each account for between thirty and thirty-five percent of the variance. This indicates that the sub-scale estimates are useful, especially for describing groups of students. This is the primary purpose of the instrument. The results indicate also that the sub-scales can be used for diagnostic purposes with individual students.

In contrast, the reliability estimate for Standard One accounts for just twelve percent of the variance, which is not sufficient for either group for individual use. It was found in addition that the estimate of reliability of the full twenty-four item scale was less than the reliability when the four items for Standard One were deleted.

TABLE TEN: **Summary of the Model Fit and Subscale Reliability Coefficients for the General Version of the CAUL Information Skills Survey.**

Standard	Number of items	Kronbach Alpha	Model – variation from Good Fit Criteria	Comment
1	4	0.35	Nil	Alternate 4 item model has Alpha=0.36
2	4	0.54	Error terms co-varied for two items	
3	4	0.78	Error terms co-varied for one item	
4	4	0.58	Nil	
5	4	0.60	Nil	5 item co-varied model achieves alpha=0.64
6	4	0.55	Nil	6 item alpha=0.55

The items for Standard One were set aside and the twenty items that addressed Standards Two to Six inclusive were selected to establish a generic information literacy self-report survey. A split half reliability coefficient was computed for the twenty item form. Items were assigned to each half of the scale at random within standards. This was done by consulting a table of random numbers and assigning items to an even (form A) or an odd (form B) half of the scale, with the proviso that once half the items for a Standard were assigned to one form, the remaining items representing that Standard were assigned to the other half of the form. Hence two items were assigned to each half of the scale for each Standard. The split half reliability coefficient was 0.77 (Ferguson 1976, 428), which equates to a full form reliability coefficient of 0.87. This means that for the twenty item survey seventy-six per cent of the variability in the data is explained.

6. STATISTICAL ANALYSIS OF LAW SCALE

A set of forty-two items from the full set of content valid items were administered to students enrolled in a first year core unit in the Faculty of Law at University of Melbourne. Two hundred and eighty students signed the consent form and completed the survey. After preliminary data analysis, six of these cases were deleted because of substantial missing data, leaving two hundred and seventy-four cases for analysis.

An exploratory factor analysis was performed using principal moments, and it produced fourteen factors with Eigen vectors greater than one. The analysis was repeated to force seven factors, to see to what extent the data conformed to the seven Standards in the first edition of the CAUL Information Literacy standard. The seven-factor solution produced an acceptable level of resolution with a KMO test of 0.80, using replacement of missing values with the mean. Seven factors could be distinguished, both visually from the scree plot and numerically. The Eigen vectors for the seven factors ranged from 7.11 to 1.41 and explained 41% of the variance. All items loaded positively on the first factor, although some loadings were below 0.3. The Standards are described as components of Information Literacy that occur as part of one process, and are performed in an iterative manner. Therefore, positive loadings for all items on the largest factor, confirmed that the Information Literacy Standards as a whole describe the various aspects of the construct called Information Literacy.

A varimax rotation was performed to explore the correlated nature of the constructs in the seven standards. An examination of the data revealed that Standards Two, Four, Six and Seven were identifiable. There were between three and six items loading on a separate factor for each standard. However, for the items representing Standards One, Three and Five there were multiple cross loadings on the remaining factors.

Standard Seven was elevated to the status of a global construct dealing with lifelong learning awareness in the second edition of the standards. Therefore, items that had been validated against this standard were deleted, and a six factor solution was explored. The six factor solution, using the Varimax estimation, produced separate factors related to Standards Two, Four and Six. For Standard Six there were four content valid items, all of which had factor loadings that exceeded the selection criterion, with no significant cross loadings. For Standard Two there were three items with no significant cross loadings, and one additional item with cross loadings between 0.2 and 0.3. For Standard Four there were also three items with no significant cross loadings, and an additional item with a cross loading of 0.22 on Standard Six. For Standards One, Three and Five there were two composite factors. One factor had three items from Standard Three, Outcomes Three and Four, plus two items from Standard Five Outcome Two. Three of these items had no significant cross loadings, and for the other two items the cross loadings on Standard Six were 0.21 and 0.23. The other composite factor contained four items, with the dominant loadings in terms of size (>0.5) on Standard Five outcome Five. There were in addition, two items from Standard One, Outcome Two. The final factor was defined with weak loadings of either zero or one items from each of the six Standards.

Standards One, Three, and Five address aspects of information use, including awareness of need, evaluating information and sources, and applying information. The occasional cross loadings with Standard Six, which addresses ethical and legal

aspects of information use, were consistent with the possibility that there was common variance for items measuring information use. Standard Two, which addresses searching techniques, Standard Four which, in the second edition of the CAUL Standards addresses saving, storing and recalling information, and Standard Six were identified as distinct factors.

As the sixth factor had weak loadings and could not be attributed to a single standard, consideration was given to forcing a four or five factor solution to resolve the combination of information awareness, evaluation and application. The five factor solution proved to be the better model. It generated higher factor loadings and fewer substantial cross loadings. In addition, it resulted in a clear isolation of Standard Five. The remaining factor consisted of a combination of items representing Standards One and Three and this was the least well specified. Using Varimax estimation, with the five factor solution, there were six items with no significant cross-loadings for Standard Six[9], and an additional item with a cross loading of -.233 on Standard One. For Standard Five, there were four items with no cross loadings, and primary loadings ranging from loadings of 0.81 to 0.41. For Standard Two there were three items with loadings from 0.68 to 0.41, with no substantial cross loadings. There were also an additional two items for Standard Two with cross loadings less than 0.3 and a difference of at least 0.2 in favour of the primary factor. For Standard Four, there were three items with no significant cross loadings and an additional item with a cross loading on Standard Six, but a difference of 0.2 between the primary and the cross-loading. The final factor contained three items from Standard One and two from Standard Three. All but the weakest loading item had cross loadings, which in three cases exceeded 0.3. It was concluded that there was limited prospects of this factor leading to a well-defined sub-scale for Standards One and Three. However it was noted that in a five factor solution, two standards had to load on one factor, and therefore it was left to the congeneric factor analysis to determine the viability of the sub-scales.

6.1 Analysis of Standard One

Given that in the exploratory factor analysis Standards One and Three were associated, the options were to construct a sub-scale for each standard from the content valid items, or to seek to derive a composite scale for Standards One and Three. Given that the Standards were published as discrete components, the items that had been included in each sub-scale on the grounds of content validity were analysed separately.

For Standard One, five items had satisfactory distributions of responses, but the sixth item was effectively a dichotomy and very positively skewed. A further item had five missing values, and the likely cause was the respondents' lack of opportunity to have experienced the specific information process. The item asked about checking relevant legislation, which some students would not have encountered early in the first semester of their studies. Deletion of two cases each with missing values on two of the subset of items, reduced the issue to missing values on two items, one with three missing values, and a mean and median of 2, and the other with two missing values,

[9] Items written for Standard Four, Outcome Three of the first edition of the Information Literacy Standards were included with Standard Six, in accordance with the relocation of this outcome to Standard Six in the second edition of the Information Literacy Framework.

and a median of 3 and mean of 2.8. The missing values were replaced by the median in each case and the resulting 272 cases were analysed.

TABLE ELEVEN: **Goodness of Fit Statistics, Standard One, Law Items**

Fit Statistic	Critical Value	6 item Model	5 item model	4 item Model
Chi Square	p>0.05	0.67	0.621	0.624
Normed Chi Square	1<NCI<2	0.742	0.703	0.472
RMSEA	<0.05	0.000	0.000	0.000
NFI	>0.95	0.998	0.999	1.000
TLI	>0.95	1.001	1.001	1.002
CFI	>0.95	1.000	1.000	1.000
AIC	See below	42.68	33.513	24.943

The results, as reported in Table Eleven, indicate that the model satisfies the criteria for a good fit with six, five or four items. The range of the factor loadings for the six item model was from 0.228 to 1.265. This indicates a six fold difference in weights, which would lead to a substantial difference between the unweighted and weighted solutions. For the five item model the range is from 0.608 to 1.223, indicting that the item with the lowest index has a value about half that of the strongest loading items. All other items had factor weights greater than or equal to 1. For the four item model the range is from the fixed value of 1 to a maximum of 1.450, indicating that an unweighted solution would be an adequate additive model, which would be an advantage for use of the scale without resorting to weighting responses.

The Kronbach Alpha reliability coefficients for the three models were computed. The six, five and four item models had values of 0.44, 0.49, and 0.45 respectively. While the differences are small, they indicate a minor advantage for the five item option, but this is not sufficient to warrant a deviation from the pattern of four items per standard achieved for the other five standards. The reliability coefficient for the four item model accounts for twenty per cent of the variance in the results. Given the limitations in the separation of factors identified by cross-loadings in the primary factor analysis, this level of reliability is a satisfactory outcome and suggests that the survey items allow for a description of group performance on Standard One.

6.2 Analysis – Standard Two

Four items from the six that had been demonstrated as content valid for Standard Two were found to load on a single factor in the exploratory factor analysis. Of these, three had loadings ranging from 0.68 to 0.41, with no substantive cross loadings. The fourth item (item 39) was more problematic. It had a lower loading (0.29) and a cross loading of 0.23 on the factor defined as Standard Four (see below). In addition there were 14 missing values and the distribution of responses was strongly negatively skewed. The estimate of skewness was large (2.12) as was the Kurtosis (3.75). Although it was overall an unsatisfactory item from a statistical perspective, it was included in the initial analysis to allow sufficient degrees of freedom to provide an insight into the performance of the other items.

Exploratory data editing identified two cases that had missing values on two of the four items, and these were deleted. The median value was used to replace a single remaining missing value in each of the other three items. For item 39, the median of 1 was used for all remaining 12 missing values. This value was assigned because it was the response of seventy-four percent of students. This item asked about use of two law specific data bases. Data collectors had reported that some students had not yet been introduced to these specialist packages, or if aware of them, had yet to learn how to use them. This would help to explain the frequency with which the option 'never' was selected, as well as the number of missing values. The congeneric factor analysis of the 4 item model provided an adequate fit. All fit indices had results within the tolerable range, but outside the margins for good fit set by Hu and Bentler, (1995), and Holmes-Smith (1999). In particular, the Normed Chi Square was above 2, and the RMSEA was greater than 0.05. The results are reported in Table Twelve. Allowing for covariance of error terms between the two weakest items, as suggested by the modification indices, a tightly specified model was obtained. The Tucker Lewis Index (TLI) exceeded 1 indicating the non-normed character of the data. The Akaine Information Criterion (AIC) is an indicator of the parsimony of the model, and a decrease in the value indicates an improved model. This index was developed to allow a comparison between models with different numbers of parameters or items (Hu and Bentler, 1995, 86). The small reduction in the AIC, from 28 to 26, indicated that adjusting for the covariance of errors improved the model. The asymptotic solution, which does not assume normally distributed integer data, was calculated using the co-varied error terms. This confirmed that the model was a satisfactory specification. The results are reported in Table Twelve.

TABLE TWELVE: Goodness of Fit Statistics, Standard Two, Law Form.

Fit Statistic	Critical Value	4 item Model	Error co-varied model	Asymptotic 4 item soln	Asymptotic 3 item solution
Chi Square	p>0.05	P~ 0.13	P=0.39	P = 0.52	P=0.755
Normed Chi Square	1<NCI<2	2.06	0.39	0.42	0.098
RMSEA	<0.05	0.063	0.000	0.000	0.000
NFI	>0.95	0.998	1.000	0.994	0.998
TLI	>0.95	0.996	1.003	1.058	1.052
CFI	>0.95	0.999	1.000	1.000	1.000
AIC	See below	28.119	26.390	18.421	10.098

A three item fully specified solution was estimated by deleting item 39, the item with a large number of missing values and extreme skew. The degrees of freedom available for the three item model make it necessary to fix the regression weights for the relationships between two of the items with the latent variable, Standard Two. Selecting the two items with the largest factor loadings in the four item solution, allowed the level at which the other item loaded to be determined. This proved a sounder solution than the other option of leaving free the factor weight of the strongest item. The result for the three item asymptotic solution is also reported in Table Twelve. The fit statistics indicate a good fit for the model. The further reduction in the value for the AIC for the asymptotic solution supports the perception that item

39 should only be used with Law students who have had some further exposure to Law Studies.

The Kronbach Alpha reliability coefficients for the three and the four item solutions were estimated. For the three item model the result was 0.62, and for the four item model the reliability coefficient was 0.58. When missing values were excluded for the four item solution, the result was 0.60. This result supported the view that there is a marginal advantage in the three item solution, but that the four item solution gives comparable results. The reliability estimate indicates that thirty-six per cent of the variance in responses is explained, which is a satisfactory level for interpreting group scores. The scale could also be used with caution to diagnose information literacy skills for individuals. Use of this sub-scale with Law students after their first semester is recommended to reduce the risk of missing responses related to the Law specific content. This restriction is expected to result in higher estimates of reliability.

6.3 Analysis of Standard Three

The finding in the exploratory factor analysis that items written for Standard Three co-varied with items written for other standards can be explained by the central function of this standard in the Information Literacy process. This standard addresses skill in evaluating information. To do so one should have regard for the purpose of the search, and for the intended use. This may explain why items that address Standard One (awareness of information needs) and Standard Five (information use) loaded in the exploratory factor analysis with items for Standard Three.

There were five items available for analysis for standard three. All had acceptable distribution characteristics, and there was only one missing value in the data set. This was replaced by the median value for the item. The result for the five-item solution was a set of goodness of fit indices that met the criteria for a good fit specified by Hu and Bentler (1995) and by Holmes-Smith (1999). No modifications were necessary to achieve this result. It was noted however that one item made a substantially smaller contribution to the solution than the other four items.

The asymptotic solution revealed some effects on the solution due to the non-normal and ordinal nature of the data. In particular, the TLI score exceeded 1 but the AIC was lower. The results are reported in Table Thirteen. The deletion of the item with low factor weighting was considered both because it might lead to a set of four items that had more equal factor loadings, and in order to conform to the number of items selected for the other sub-scales. The advantage of items that load more or less equally is that there will be little difference between the additive model, and the model derived from weighting scores using factor weights.

As reported in Table Thirteen, the four-item model generated a solution with good fit indices, and with no recommended modifications. Again, the asymptotic solution generated indices that signified a good fit of the data. In fact, the four item solutions may be preferable in that it generated TLI scores less than one and lower AIC values, compared to the five item model. The calculation of the Kronbach Alpha reliability coefficient produced a value of 0.52 for five items, and 0.54 for four items. This confirmed that there was no deficit in predictive power as a result of using the four

item solution. More than twenty-nine per cent of the variance is explained by the four item model.

TABLE THIRTEEN: Fit Statistics, Standard Three, Law Form.

Fit Statistic	Critical Value	5 Item Model	Asymptotic 5 Item	4 Item Model	Asymptotic 4 Item
Chi Square	p>0.05	0.468	0.595	0.209	0.303
Normed Chi Square	1<NCI<2	0.917	0.738	1.564	1.195
RMSEA	<0.05	0.000	0.000	0.045	0.027
NFI	>0.95	0.999	0.941	0.999	0.96
TLI	>0.95	1.000	1.049	0.995	0.978
CFI	>0.95	1.000	1.000	1.000	0.993
AIC	Lowest	34.568	23.692	27.128	18.391

6.4 Analysis of Standard Four

Four of the six items that had been validated for Standard Four were identified from the exploratory factor analysis as clustered in one factor. Of these items, one was written for Standard Four, Outcome Three in the first edition of the Standards It made reference to the use of copyrighting when publishing. This was likely to be beyond the experience of most first year students. There were 15 missing values on this item, which indicated that some students responded in this manner because they had no experience of publishing. The median value was one, indicating no experience, and skewness and kurtosis were high. However, infrequent or no practice was evident also for two other items that were related to the use of bibliographic software. Again, this was likely to indicate lack of experience at this early stage of their academic career. Therefore these items may be more appropriate for students beyond the first semester of their Law studies. Given the likely cause of missing values was non-random, deleting all such cases was justifiable. It was found in this case that doing so provided a better solution than was obtained by substituting median values.

The congeneric factor analysis of the 4 item model using the reduced data set resulted in a good fit with all selected indicators within the margins for good fit set by Hu and Bentler (1995) and by Holmes-Smith (1999). In particular, the Normed Chi Square was 1.35, and the RMSEA was less than 0.05. The results are reported in Table Fourteen. No modifications to the model were recommended. The asymptotic solution confirmed that the model was a satisfactory specification. The TLI was less than 1 and the normed chi-square was 1.07. The asymptotic solution resulted in a reduction in the value for the AIC. This added to the evidence for a good model fit for this non-normal ordinal data.

Given the various limitations identified with one item, a three factor solution was attempted. Due to the limit on the number of degrees of freedom, it was necessary to fix the regression weights for the relationships between two of the items and the latent variable, Standard Four. Selecting the two stronger items allowed the weaker item to find the level at which it could contribute. This proved a sounder solution than the other option of leaving free the regression weight for the strongest item, although in

both cases the result conformed to the criteria for a good fit. The results are reported in Table Fourteen. The fit statistics indicate a good fit for the model, although the TLI is exceptionally high, highlighting the non-normal, and skewed nature of the ordinal data.

TABLE FOURTEEN: Fit Statistics, Standard Four, for LAW

Fit Statistic	Critical Value	4 Item Model	Asymptotic 4 Item	3 Item Model	Asymptotic 3 Item
Chi Square	p>0.05	P=0.259	P=0.344	P=0.835	P=0.825
Normed Chi Square	1<NCI<2	1.351	1.066	0.043	0.049
RMSEA	<0.05	0.037	0.016	0.000	0.000
NFI	>0.95	0.998	0.917	1.000	0.997
TLI	>0.95	.998	0.980	1.004	1.196
CFI	>0.95	1.000	0.993	1.000	1.000
AIC	lowest	26.70	18.132	16.043	10.049

An analysis of the Kronbach Alpha reliability estimates, and the distribution characteristics for the three and four item models indicated that there are limitations with the specification of this Standard with the present data. For the four item model the Kronbach Alpha reliability coefficient was 0.41, and for the three item form the estimate was 0.39. Neither is a satisfactory level for an independent assessment of performance on this sub-scale. The four item solution accounts for less than seventeen per cent of the variance. This will provide limited information of use in considering the performance of groups of students. It was concluded that as the four item model generated good conformity with the model, by specifying Standard Four with four items, the sources of evidence were increased and hence the reliance on any one item was reduced. The weight of each individual item in the final model allowed for an additive solution.

6.5 Analysis of Standard Five

For Standard Five, four of the six content validated items were identified in the exploratory factor analysis as having acceptable factor loadings and no significant cross loadings. The factor loadings ranged from 0.81 to 0.42. The other two items had primary factor loadings above the 0.50 level, with cross factor loadings between 0.2 and 0.3. Thus all six items met the liberal criteria established for inclusion in the congeneric factor analysis. There was only one missing value in the data set for this sub-scale, and it was replaced by the median value for that item. For this case, all responses to the other five items were at the respective median value, so assigning the median for the missing value, fitted with the other responses for this case.

As indicated in Table Fifteen, the initial six item solution was not an adequate model fit. The values for both the normed chi square (4.71) and the RMSEA (0.12) exceeded established fit criteria. The other goodness of fit indices were acceptable. Based on the modification indices provided in the analysis, three pairs of error variances were co-varied and this produced indices that met the criteria for a satisfactory model fit. In addition, no further modifications were suggested for this analysis. For the co-varied

model, the normed chi squared figure met the good fit criteria, and the root mean square estimate of 0.067 was in the range that is identified as acceptable by Holmes-Smith (1999). The results for the TLI and normed chi squared indicate that the model with co-varied error terms is better than the model without co-varied error, from which it is derived. The TLI is an incremental fit index, and the improvement from 0.972 to 0.996 indicates a better fit. This is confirmed by the decline in the estimate for the AIC index from 44.26 to 28.21.

TABLE FIFTEEN: Goodness of Fit Statistics, Standard Five, 6 item model, Law.

Fit Statistic	Critical Value	6 Item Model	Error Co-varied Model	Asymptotic 6 item solution
Chi Square	p>0.05	P<0.00	P=0.517	P=0.129
Normed Chi Square	1<NCI<2	4.706	0.87	1.612
RMSEA	<0.05	0.117	0.000	0.047
NFI	>0.95	.991	0.999	0.998
TLI	>0.95	.984	1.001	0.997
CFI	>0.95	.993	1.000	0.999
AIC	lowest	78.354	47.208	51.216

The asymptotic solution for the model with three pairs of co-varied error variances, confirmed that the model is adequate. The normed chi-squared, the RMSEA and the TLI all showed improvement. The AIC estimate of 51.216 was slightly higher than for the co-varied model. However, overall this result is satisfactory because all other indicators were acceptable against the good fit criteria (Homes-Smith 1995).

Given that all other standards were to be estimated by four items, achieving a four item solution for Standard Five simplifies the procedures for achieving an overall estimate of Information Literacy. By deleting two items with low factor weights, a four item solution was generated and compared to the solution including all six of the items that had been shown to have content validity for Standard Five. As for the six item solution, the initial result provided unsatisfactory goodness of fit statistics. However, after allowing for covariance between one pair of error terms, the model satisfied goodness of fit criteria at the adequate or good level. The asymptotic solution for the model with the pair of co-varied error variances, confirmed that the model is adequate. The normed chi-squared, the RMSEA and the AIC all showed further improvement. The TLI estimate declined to 0.917. This result is classified as adequate as opposed to a best-fit model (Homes-Smith 1995).

The fit statistics for the modified four item model are not uniformly as good as for the six item model. The six item co-varied model has fit statistics that indicate an over-fit, while for the four item model, the normed chi square value of 2.209 satisfied the adequate, rather than good criterion. The results are reported in Table Sixteen.

Given that all other standards were to be estimated by four items, achieving a four item solution for Standard Five allowed for an assessment tool with four items for each standard. For an expanded version of the assessment tool, all six items could be used. The Kronbach Alpha reliability index was estimated for the four and six item

forms and the results indicated that the four item model was marginally better than the six item model (0.65 compared to 0.63). Therefore the four item model accounts for more than forty-two per cent of the variance. Taken together, all the evidence indicates that the four item model is acceptable. It requires less modification to achieve an acceptable fit, provides a marginally better reliability estimate, and has demonstrated factor structure coherence.

TABLE SIXTEEN: **Fit Statistics, Standard Five, 4 item model, Law.**

Fit Statistic	Critical Value	4 Item Model	4 Item Co-varied Model	4 Item asymptotic solution
Chi Square	p>0.05	P<0.00	P=0.137	P=0.142
Normed Chi Square	1<NCI<2	10.128	2.209	2.157
RMSEA	<0.05	0.183	0.067	0.065
NFI	>0.95	0.994	0.999	0.976
TLI	>0.95	0.972	0.996	0.917
CFI	>0.95	0.994	1.000	0.986
AIC	See below	44.256	28.209	20.157

6.6 Analysis for Standard Six

A total of seven items were included in the full list from which to construct the estimate for Standard Six. Of these items, only one had any substantial (>0.20) cross loadings in the five factor exploratory analysis. This item loaded at 0.37 on the factor identified as addressing Standard Six, but cross-loaded on the ill-defined factor that was identified as a composite of Standards One and Three, with a value of 0.23. Although this item was compromised and made a limited contribution to the primary factor, it was included for the initial analysis because it was statistically the best of the available items reflecting Standard Six, Outcome Three. However, in the revision for the second edition of the Information Literacy Framework, the original Standard Four, Outcome Three was incorporated into Standard Six. Two items that were initially designed for this element of the standard also had moderate loadings on the factor defining Standard Six, and hence were also included in the initial analysis.

Two items had twelve and eleven missing values respectively. One of these items had been content validated against Standard Six, Outcome One, and had the highest factor loading for items addressing this outcome. It was therefore desirable to consider this item in the analysis. Seven of the twelve cases with a missing value for this item, also had a missing value for the other case with a large number of missing values. These seven cases were therefore dropped, and the median value substituted for the remaining missing values in the data set. This provided a data set with 267 cases.

The congeneric factor analysis produced a result that indicated an adequate fit for the data. All the indices satisfied the critical values specified by Holmes-Smith (1999) and by Hu and Bentler (1995). An analysis of the data indicated that the contribution of the item with low factor loading and high cross factor loading was, at best,

marginal. The only modification index suggested from the full analysis was to co-vary error terms for this and one other item. This item had a squared multiple correlation of 0.052, which was much lower than for any other item. The item had highly skewed responses, with nearly seventy per cent of respondents selecting the highest value of 4. The item had zero order correlations with three of the other items, and correlations less than 0.2 with the remaining items. The Kronbach Alpha for the seven item scale was 0.57, and the value obtained for the six item scale was 0.58. Given this statistical evidence, the item was deleted from the analysis. The result for the analysis for the six item data set is provided in Table Sixteen.

TABLE SIXTEEN: Fit Statistics, Standard Six, Law Scale.

Fit Statistic	Critical Value	Seven Item Model	Six Item Model	Six Item Asymptotic
Chi Square	p>0.05	P=0.348	P=0.604	P=0.484
Normed Chi Square	1<NCI<2	1.103	0.813	0.945
RMSEA	<0.05	0.020	0.000	0.000
NFI	>0.95	0.996	0.998	0.863
TLI	>0.95	0.999	1.001	1.017
CFI	>0.95	0.999	1.000	1.000
AIC	Lowest	57.4	43.3	32.5

The analysis indicated a good fit with the model. The results for the Normed Chi Square, and TLI indicate a possibility that the model is over-fitted, however there was an improvement in the AIC compared to the seven item solution, which is supportive of the decision to delete the seventh item. The AIC is an indicator of the parsimony of the model, and a decrease in the value indicates an improved model.

The final step in analysis of the six item model was to generate the asymptotic solution. This takes into account the ordinal nature of the data and is calculated using polychoric correlations, which are more appropriate with variables specified on a four point scale. The results for this analysis are also reported in Table Sixteen. The Normed Chi Square value is close to the preferred minimum value of 1 and the value of the AIC is again lower than the alternative model. Both these results are encouraging. However, the estimate for NFI dropped below the desired level of 0.95 and below the minimum acceptable level of 0.90 (Hu and Bentler, 1995). The NFI ‘represents the proportion of total variance explained by a target model when using the null model as a base line’ (Hu and Bentler, 1995, 83). The decline in this value when the asymptotic model is applied is a consequence of generally low co- variances among the contributing variables. It is taken as indicative of the anticipated error of measurement in the data. This may reflect the timing of the data collection. Standard Six refers to ethical and legal practices in the use of information, about which students might not be sure in their first semester.

Given that four item sub-scales had been identified to describe the other standards, consideration was given to the feasibility of establishing a four item estimate for Standard Six. Two approaches were compared. In the first, items were selected to ensure that there was at least one item for each of the three outcomes for Standard Six. The fourth item was then selected on the basis of more equal regression weights

and superior factor score. This resulted in two items selected for Outcome One and one item for each of the other two Outcomes. In the second approach, items were selected in terms of the size of the regression weights and factor loadings. This resulted in two items each for Outcomes One and Three, and none for Outcome Two. A comparison of the fit statistics for these two four item models, compared to the six item model is reported in Table Seventeen.

TABLE SEVENTEEN: **Comparison of Fit Statistics for Four Item Models for Standard Six.**

Fit Statistic	Critical Value	Coverage 4 item model	Weights 4 item Model	Comparison with 6 Item model
Chi Square	p>0.05	P=0.248	P=0.272	P=0.604
Normed Chi Square	1<NCI<2	1.393	1.303	0.813
RMSEA	<0.05	0.038	0.034	0.000
NFI	>0.95	0.999	0.938	0.998
TLI	>0.95	0.998	0.950	1.001
CFI	>0.95	1.000	0.983	1.000
AIC	lowest	26.786	18.605	43.3

The solution that provided coverage of all three outcomes of the Standard had fit statistics that met the good fit criteria established by Hu and Bentler (1995) and by Holmes-Smith (1999). In contrast, the alternative solution using items with the largest factor weights, had an estimate for NFI that was in the adequate as opposed to the good range, and a TLI estimate that exceeded 1. However, the alternative model had a more parsimonious AIC estimate.

The Kronbach Alpha Reliability coefficients were computed for the seven, six, and for the two four item models. There was an advantage for the 6 item model over either four item options. The value for the six item model was 0.58. For the four item coverage solution in which items were selected to include all three outcomes, the estimate was 0.47. For the alternative model based on the best factor score loadings, the reliability was 0.51. These results confirm that six items are suitable for estimating Standard Six. As it was desirable to construct a scale with four items per standard, the four item model with best coverage of the outcomes was adopted, even though it has a reliability for the validation population that is slightly lower than the alternative four item model. This is based on the benefits of demonstrating content validity rather than relying solely on statistical criteria. This is justified in part by the restricted quality of the overall statistical evidence which may be associated with the time that the items were administered, before substantial exposure to Law related information literacy tasks. The preferred four item model accounts for over twenty-two per cent of the variance.

6.7 Summary of Analysis for Standards One to Six

In developing assessment tools for Information Literacy the inter-related nature of the six components of the Standards has to be considered. The goals were to obtain an overall estimate of Information Literacy from the self-report survey, and in addition to gain an indication of the performance of students on each sub-scale. This requires

content valid items for each of the six Standards, as well as items that are internally consistent when used as a single measure of Information Literacy. This analysis has demonstrated that as a single construct, Information Literacy can be specified as a graduate attribute by combining items drawn from across the six Standards, while maintaining the content validity of the tool.

Two forms of the Law version of the Information Skills Survey were created. One form included four items per Standard to allow each Standard the potential to contribute an equal number of items to the estimate of Information Literacy. This provided a twenty-four item scale. The second form included two extra items for standard five, and two more for Standard Six. While it was opportunistic that extra items were found to contribute statistically for these two Standards, there is a post-hoc rationale in considering an expanded form. It can be argued that the application of information (Standard Five) and the ethical and legal use of information (Standard Six) are the culminating and hence the most important outcomes. Allowing additional items for each of these Standards may enhance the face validity of the tool. Including the extra items generated a twenty-eight item form.

A split half reliability coefficient was computed for the twenty-four and twenty-eight item forms. Items were assigned to each half of the scale at random within standards. This was done by consulting a table of random numbers and assigning items to an even (form A) or an odd (form B) half of the scale, with the proviso that once half the items for a Standard were assigned to one form, the remaining items representing that Standard were assigned to the other half of the form.

The twenty-four item form had four items for each of the six Standards, and two items were assigned to each half of the scale for each standard. The split half reliability coefficient was 0.622, which equates to a full form reliability coefficient of 0.77. This means that the twenty-four item form accounts for fifty-nine per cent of the variability in the data. The alternate version had twenty-eight items. This included six items for Standard Five and six for Standard Six. This form had a split half reliability coefficient of 0.65, which equates to a full form reliability coefficient of 0.84. This means that the twenty-eight item form accounts for over seventy percent of the variance in the data.

With first year students there appeared to be a limited grasp of aspects of Standard Four, and hence the inclusion of items for this standard might well be restricted to learners with some substantive experience of higher education. The four item solution was adopted with the proviso that the performance of these items should be reviewed with a more experienced group of students. It is possible that with students who have progressed further in their studies, there will be few students unable to respond due to lack of experience, and hence reliabilities may be more satisfactory. This suggestion can only be investigated with further data collection using students who have progressed further in their studies.

In summary, it has been possible to select four items from the content valid item pool for each of the six Standards. The amount of the observed variance within each sub-scale accounted for by the models range from just seventeen percent for Standard Four, to forty-two percent for Standard Five. These results are sufficient to allow

predictions of group performance based on the sub-scale scores. The results for all six sub-scales are reported in Table Eighteen.

TABLE EIGHTEEN:
Summary of the Model fit and Subscale Reliability Coefficients for the Law Versions of the Information Skills Survey

Standard	Number of items	Kronbach Alpha	Model – variation from Fit Criteria	Comment
1	4	0.46	Nil	
2	4	0.58	4 item model, TLI = 1.06	3 item model, R=0.62 TLI = 1.05
3	4	0.52	Nil	
4	4	0.41	Nil	
5	4	0.65	4 item model, Normed Chi Square exceeds 2, and RMSEA= 0.065	6 item model has better fit statistics, but R= 0.63
6	4	0.47	Nil	6 item model has R=0.58

7. CONSTRUCT VALIDITY

The direct method of assessing information skills is to observe student behaviour. This is too expensive to apply to large cohorts of students, but can be used with a small number of students to gather data that can be compared with performance on the self-report instrument. Fifty interviews were planned with Law students who had completed the Information Skills Survey at University of Melbourne, and a further fifty were planned with Education students at Central Queensland University. All students were interviewed within a fortnight of having completed the self report survey.

To make an assessment of information skills against the Standards by observation of student behaviour, an interview protocol was developed. The protocol was trialled by librarians at the participating institutions and the interview form was modified on the basis of feedback, to enhance the clarity and effectiveness of the process. The plan was to make an assessment for each student on each Standard, and for each Outcome within each Standard. As Standard Seven had been deleted from the second edition of the Information Literacy Framework, the observed performance for this Standard was not included in the subsequent analysis.

The revised protocol was used in hour long interviews conducted by librarians. The focus of each interview was an observation of a student's behaviour when presented with a realistic but hypothetical assignment topic. The hypothetical exercise was similar in format to ones required as part of the student's current program. Thus the interview and the self-report by students were both directed at the assessment of their current practice. The interview protocol required the librarian to ask a question directed at each outcome in the Information Literacy Standards, and to probe the

responses to confirm the level of information skills of the student. This included asking the student to demonstrate procedures. Two benefits of participation were offered to students. Each person was provided with feedback on their performance at the close of the interview. In addition, each participant received a twenty dollar photocopy card for use in the Library.

The correlation between the level of Information Literacy observed by a librarian and the estimate obtained from the relevant self-report tool was examined using three indicators of correlation, namely Person's product moment correlation, Kendall's Tau and Spearman's rank order correlation. The Pearson estimate assumes integer data. This means that the data is considered to be a scale that allows for the difference between values to be assumed to be consistent. Given the ordinal nature of the judgements on which the scales were constructed, this assumption needs to be treated with caution. Therefore two non-parametric measures of correlation that rely on ordinal data were estimated to provide further evidence of the accuracy of the estimated relationships. The findings are reported separately for the Education and the Law versions of the ISS.

7.1 Education Form

Fifty-seven Education students were interviewed by librarians who had been trained in the use of the protocol. Three students who were interviewed had not completed a survey consent form for their self-report data. Their data was not used, even though they had consented to the interview, because they had not granted permission for the use of their self-report data with which the Librarian assessment was to be compared. Seven Librarians undertook the interviews, spending one hour per student. As indicated in Table Nineteen, they interviewed between three and sixteen students. In the discussion below the Interviewers are referred to as Coders One to Seven.

TABLE NINETEEN: Number of Education Student Interviews per Coder

Coder	Frequency	Percent
1	16	29.6
2	3	5.6
3	6	11.1
4	4	7.4
5	9	16.7
6	12	22.2
7	4	7.4
Total	54	100.0

An inspection of the interview coding forms demonstrated that all interviews had been conducted in a systematic and comprehensive manner. Nonetheless there were many variations in the manner in which data was recorded. Assessing students on each of the twenty-five Outcomes listed in the first edition of the Standards proved to be a demanding task. Not all coders were able to do so for each student. Comments indicated the difficulties experienced by Coders. In a few cases, a statement indicated that there had been insufficient time to check each specific Outcome.

It had been intended to compare the average of the assessments across the outcomes within each standard with the overall assessment awarded for each standard. This was considered a means of checking the internal reliability of the assessments made by each coder. The concern about the internal reliability of the assessments made by each coder was in part ameliorated by the evidence in the form of notes that showed that the assessments had been undertaken with care. With one exception, each coder appeared to have adopted a consistent approach to the assessments. However, the variable methods adopted indicated that the between coder reliability was in question because different methods had been applied by some coders to report their assessments of information skills.

One coder wrote copious notes on each survey form, and either placed ticks beside boxes or left them blank. The ticks seemed to indicate that the coder had made an assessment, but not the level of the assessment. In some cases comments were provided instead of a number such as ‘no’ or ‘not really’. These comments could be interpreted as implying negligible skills, but other comments were less easily interpreted in terms of an ordinal scale. Consequently it was not possible to transform the comments into a scale for use in the analysis. Another coder provided an overall numerical rating for each Standard for all students, but did not use the four point ordinal scale proposed. The assessments ranged from 2, to 3.5 in increments of 0.5 plus an added scale point called ‘3 to 3.5’. This meant that there were five points on the scale adopted by this coder compared with four points on the standard scale. Before making a decision as to how to transform this scale, notes made by the coder beside the numerical estimates were reviewed. These confirmed that the full range from 1 to 4 was appropriate. For instance, a value of 2 was rescaled as 1, because the notes indicated that the student went about finding library materials by asking a librarian, implying negligible searching skills. This coder was responsible for twelve of the fifty-four estimates of information skills, and hence it was important to transform this data so that it could be included. It was decided to convert this to the standard four point scale by equating the ‘3 to 3.5’ range estimate to the higher value of 3.5, and then applying the standard four point scale of 1 to 4. Scale transformation is a common preliminary data analysis procedure and does not alter the rank order of the relationships identified by the coder. It makes it possible to combine the scale with that used by other coders.

Two coders provided a complete data set including an assessment against each outcome for all students. In the other cases, the number of missing assessments varied from three (4%), to a total of one hundred and forty-six (68%). In this case, some students had been assessed against most of the Outcomes, and others had very few assessments by Outcomes. This indicated that the coder may have been inconsistent in the way assessments were made of different students. The written comments provided by the coder for each student were compared with the ratings and found not to be consistent. Students for whom there were markedly different written comments about performance were given the same overall rating. This coder did not differentiate levels of performance between students and either assigned the same numeric value to each student, or used just two points in the range to assess all students on each standard. These factors raised concern about the consistency of the assessments and hence issues for validity and reliability.

Table Twenty: Range of Missing Data for 18 Outcome Estimates for Standards Two to Six by Coder; Education Data

Coder	Number of Students Assessed	Minimum Missing per Outcome	Maximum Missing per Outcome	Total Missing Estimates	
				Number	Per cent
1	16	0	0	0	0
2	3	0	2	19	6.6
3	6	0	3	11^	10.2
4	4	0	1	3	4.2
5	9	0	0	0	0
6	12	6	9	146	67.6
7	4	0	4	18	25.0

^seven of these missing estimates occurred across three of the five Outcomes in Standard Five.

The extent of the missing data for individual outcomes meant that in order to consider assessments made by all coders, the correlations with student self reports had to be estimated using the overall coder assessments of information literacy for each standard. The extent of the missing data for coder assessments by Outcomes is summarised in Table Twenty.

Given the range of difficulties in analysing the assessments made by some coders it is reiterated that all coders undertook their task with diligence and care. Inconsistencies noted are therefore a consequence of the design and in particular the lack of feedback and guidance that was available. This in turn was caused by the very limited funding available to undertake the content validation. It meant that the researcher was not able to advise coders, and they were not always able to compare their understandings and resolve difficulties in processing the data.

A comparison was made of overall coder assessments for each Standard, with the assessment of each Standard calculated from the average of the Outcome assessments. There was substantial consistency for those coders who assessed each Standard overall, and by Outcome, as is evidenced in Table Twenty-one.

Table Twenty-one: Comparison between Coder Assessments by Outcomes and Overall for each Standard, for Education Cohort

Standard	Pearson r	Spearman r	Number
2	.92	.90	42
3	.85	.87	41
4	.74	.72	39
5	.88	.88	34
6	.83	.83	42

The Kendall Tau estimate is not included in this Table because it was entirely consistent with the other two estimates. The strength of the relationships indicated

that coders were consistent in rating performance on each Outcome and on the Standard overall. With the exception of Standard Four, all correlations exceeded 0.80. In the case of Standard Four, the coders were making overall judgements against the Standard as defined in the first Edition of the Information Literacy Standards. Outcome Three which had addressed copyright issues was transferred to Standard Six in the second edition of the Standards, and hence the estimate based on the Outcomes was not directly comparable to the assessment by Outcomes.

In contrast with the missing data for assessments of individual outcomes within each of the Standards, all coders had provided an overall single numeric assessment for each standard for all fifty-four students. Consequently it was possible to determine an overall assessment of information literacy for each student by summing the overall estimates for the six standards. This could then be compared with the overall estimate of information literacy from the sum of the self-report responses. The correlations were 0.42 for the Pearson product moment correlation, 0.44 for the Spearman estimate, and 0.31 for the Kendall Tau estimate. All three satisfied the 1% significance criterion. Given that the reliability of the twenty-four item scale was 0.87, and assuming that the inter-coder reliability for coders was about 0.80, the estimated maximum correlation is 0.70. Given the substantial evidence of variability in numeric assessments between coders, a lower inter-rater reliability coefficient of 0.60 would mean that the maximum correlation that could be achieved would be 0.52. Hence the product moment correlation of 0.42 indicates that a substantial amount of the variability[10] that could be expected to be accounted for in the ideal situation where there was full agreement has been demonstrated. For the non-parametric estimates of correlation, the maximum correlation is also limited to a value less than one both by the inter-rater reliability issues and by the restrictions on the maximum value of estimates imposed by ordinal data. This is especially the case with the Kendall estimate. One can consider these two non-parametric estimates of correlation to represent an upper and lower boundary. Across all coders the non-parametric estimate is between 0.31 and 0.44. The Pearson estimate falls within the range of the estimates from non-parametric sources, and hence the three sources of evidence indicate that the coder assessments are consistent overall with the student self-report survey results.

Correlations between student self reports and the coders' assessments of Information Literacy were also calculated for the five standards included in the self-report survey, using the overall assessment by each coder.

Table Twenty-Two: Correlations for Standards Two to Six between overall coder assessments of Information Literacy and Education student self-report estimates.

Standard	Pearson	Spearman	Kendall
2	0.37**	0.37**	0.30**
3	0.33**	0.31*	0.26*
4	0.31*	0.21	0.17
5	0.20	0.37**	0.30**
6	0.03	0.13	0.12

[10] A correlation of 0.42 accounts for over 17.6%; a correlation of 0.52 accounts for 27%.

** Significant at 0.01 level.
*significant at 0.05 level
As this was limited to a maximum range of 1 to 4, the non-parametric correlation coefficients were expected to be more robust. There were fifty-four cases. The results are reported in Table Twenty-two.

For all but Standard Six, the results indicate that there is a statistically significant relationship between the assessment made by the coders and the self-report estimates from students. The result for Standard Six is not only in contrast with the correlations for the other standards but, as will be seen below, is also in contrast with the results obtained for the Law students. A reason for this finding was the limited differentiation that some coders made between Education students on Standard Six. This is illustrated in Table Twenty-three. The most extreme case was Coder Five who assessed nine students. This coder rated all nine students at the same value (3) for both the overall rating on Standard Six, and on all three outcomes. This lack of differentiation meant that the self-reported levels for these nine students had, by definition, a zero correlation with the coder assessment. Two other coders who assessed sixteen and twelve students respectively also favoured an assessment of three. This explains the low correlations on Standard Six. It demonstrates how reliance on a single overall assessment of information literacy for each standard is an inadequate basis for a quantitative analysis. This is why the combination of assessments across the Outcomes was the intended method of analysis.

Table Twenty-three: Range of Overall Assessments for Standard Six by Coder for Education Cohort

Coder	**Ratings**				**Total**
	1	**2**	**3**	**4**	
1	0	1	13	2	16
2	0	1	0	2	3
3	0	2	2	2	6
4	0	2	2	0	4
5	0	0	9	0	9
6	1	0	11	0	12
7	0	1	2	1	4
Totals	**1**	**7**	**39**	**7**	54

As reported previously in Table Twenty, complete data was available for the overall assessment of each student for each Outcome for Coders One and Five, who rated a total of twenty-five students. In addition, Coder Three had few missing estimates once those in Standard Five were identified, and Coder Two had missing values for Standards Three and Six. This meant that a further nine estimates were available for some Standards. Coder estimates derived from assessments of each Outcome could be calculated from three coders on Standard Six, for four coders on Standards Three and Five, and from five coders on Standards Two and Four. These findings are summarised in Table Twenty-four.

Given that some coders did not differentiate between students in their assessments the performance of each coder was analysed for each standard to see if this could help to

explain the results overall. Cases were found where no correlations were available due to missing values, rather than a lack of differentiation between student assessments.

An exceptional result was identified with Standard Four. Three coders provided low positive correlations for students on Standard Four, but two who accounted for eight cases provided large negative correlations. As noted above, the coders rated students against the overall requirements in the first edition of the Standards. This included an additional Outcome that referred to copyright. This Outcome was transferred to Standard Six in the second edition of the Information Literacy framework, and hence the student self-report estimates were based on only storage and retrieval components.

For Standard Six, only three coders provided sufficient assessments by Outcomes. All three coders provided positive correlations with student self-report estimates. Hence this was judged to be a three out of three agreement, notwithstanding the low overall correlation reported. These and other variations among the results are recorded in comments in Table Twenty-four.

TABLE Twenty-four: Correlations for each Standard based on Coder Assessments by Outcomes for Education Cohort

Standard	Pearson	Spearman	Agreement	Comment
2	.35*	.35*	5 /5	Two had positive but low results
3	.32*	.32*	3 /4	Coder 4 had negative Pearson and Zero Spearman correlations (4 cases)
4	.09	.01	3 /5	Coder 4 (n=4) and Coder 3 (n=4) produced large negative correlations.
5	.26	.20	3 /4	Coder 1 (n=16) had a small negative correlation, Coder 5 (n=9) had a significant positive correlation.
6	.08	.07	3 /3	Coder 3 (n=5) had a significant positive correlation; coder 1 (n=16) had a 0.24 correlation, and coder 4 had a 0.21 Spearman, but zero Pearson correlation.

These results show that three coders achieved assessments that were consistent both with each other and with the self-report estimates by students. Consequently it is concluded that the results provide evidence of the concurrent validity of the Education form of the scale. The Education form of the ISS provides estimates of Information Literacy which are consistent with those made by a majority of the coders for whom assessments by outcomes were available. The overall correlations between these coders' assessments and student self-report estimates for each standard are reported in Table Twenty-five, and are compared with the estimates provided by all seven coders using their overall assessments for each standard, as reported above in Table Twenty-One. The results for all five coders are included, because it was not considered appropriate to make post-hoc adjustments to selectively delete assessments for particular standards, even where a post-hoc rationale was offered.

Table Twenty-five: Correlations between coder assessments of Information Literacy and Education student self-report estimates, by Standard

Standard	all coders – overall assessments by standard			coders – assessments by outcomes within standards		
	Number	Pearson's r	Spearman's r	Number	Pearson's r	Spearman's r
2	54	0.37**	0.37**	35	.35*	.35*
3	54	0.33**	0.31*	36	.32*	.32*
4	54	0.31*	0.21	35	.09	.01
5	54	0.20	0.37**	31	.26	.20
6	54	0.03	0.13	34	.08	.07

** beyond 0.01 level of significance
* beyond 0.05 level of significance

In addition, the overall estimates of the correlation between student self-reports and coder assessments based on outcomes within Standards were calculated and compared with the estimates provided by the coder assessments overall for each standard. As reported in Table Twenty-six, the results were consistent, but lower for the Outcomes based assessments. It was noted that the use of the individual outcomes to calculate an assessment reduced the numbers by about forty per cent.

Table-Twenty-six: Correlations between different sources of Coder Assessments with Education Student Self-report of Information Literacy.

Coder Assessment	Pearson	Spearman
Overall (n=54)	.42*	.44*
By Outcomes (n=35)	.35*	.35*

As noted previously, there are two factors that impose an upper limit on the possible correlations. For the Pearson product moment correlations, the reliability of the twenty-four item self-report form is 0.87. As there will be some degree of variation in the assessments across the coders who provided assessments for individual Outcomes, the reliability of their estimates will also be less than one. If the reliability across the coders is of the same order as the reliability for the student scale then the maximum Pearson Product Moment correlation is 0.70, which is the product of the estimated reliabilities. If the inter-coder reliability is 0.60, the maximum predicted correlation is 0.52. The correlation for all coders was 0.42, and for the five coders who provided assessments by outcome data was 0.35. These results are therefore satisfactory and indicate substantial agreement between the self-report data collected from students and the assessments of coders.

Another aspect is the comparison between the estimates of reliability for each standard for the ISS and the correlation between coders and students. This comparison is reported in Table Twenty-seven. The correlations are high for Standards Two, Three and Four, especially when the level of the reliability of the student self-report estimates is taken into account. Given the reported inconsistencies in coding, an inter-coder reliability estimate of 0.8 is generous. If the inter-coder reliability is estimated

to be 0.6, then the expected estimates are much lower as indicated in the final column of Table Twenty-seven. This illustrates why inconsistent coding can have a marked impact on the correlation estimates.

Table Twenty-seven: Comparison of observed Product Moment Correlations and Estimated Maximum Correlations for Education Cohort

Standard	Number of items	Kronbach Alpha	Pearson's r	Estimated maximum correlations	
				Coder reliability 0.8	Coder reliability 0.6
2	4	0.54	0.37**	.40	0.32
3	4	0.78	0.33**	.62	0.47
4	4	0.58	0.31*	.46	0.35
5	4	0.60	0.20	.48	0.36
6	4	0.55	0.03	.44	0.33

*beyond 0.05% confidence interval
** beyond 0.01% confidence interval

Notwithstanding the limitations in the data caused by the design of the content validity assessment, in a majority of cases the overall correlation between the judgement by coders of information literacy and the self assessment of Education students of their information skills was both substantial and statistically significant. For the two Standards where correlations are low, the lack of numerical differentiation by coders between students is a contributing factor.

7.2 Law Form

Fifty Law students who consented to be interviewed were also assessed in one hour interviews by librarians who had been provided with an assessment protocol and been trained in its use by a local member of the project management committee. Unfortunately, the student identification (ID) numbers on the survey (self report) forms were not recorded for nearly half the students, and therefore had to be cross referenced by checking names or signature for twenty-seven of the students who had been interviewed. All but three were identified and hence data was matched for forty-seven students. Five coders conducted between eight and fourteen interviews. The distribution of interviews among coders is reported in Table Twenty-eight.

Table Twenty-Eight: Number of Law Student Interviews per Coder

Coder	Frequency	Percent
1	8	17.0
2	8	34.0
3	9	19.2
4	14	29.8
5	8	17.0
Total	47	100.0

The coders assessed each student on each of the Outcomes for each of the Standards. For each Standard, the average of the assessments for each Outcome was calculated as an estimate for that standard. An overall estimate of information literacy for each student was computed by summing the estimates for each Standard. Thus each Standard contributed equally to the overall assessment for each student. Separately, an overall estimate of each student's self-reported information literacy was computed by summing their scores on the twenty-eight item Law scale. The use of this scale gives the contribution of Standards Five and Six by fifty percent in the overall estimate of Information Literacy. However, this is the recommended version of the Law ISS, and has a higher estimated reliability than the alternative twenty-four item form.

Correlations between student self reports and the assessment across all five coders were calculated, followed by estimates for each coder taken separately. For all three estimates of association, the combined results across the five coders were statistically significant. There were 36 cases with no missing values. These provided estimates of association of 0.39 for the product moment correlation, and 0.41 for the Spearman estimate. Both satisfied the 1% significance criterion. The Kendall estimate was 0.27 with a probability of 0.012, or just over 1%.

Each coder was trained and asked to rate students against a set of criteria. However, as there was no mechanism to review of rater reliability, it was decided to compare coder's results by calculating the correlation with student self-reports for each coder. This was in part to investigate the consistency of the findings across the coders. The number of assessments per coder for which there were no missing values in both the self-report data provided by students, and the assessments made by the coders ranged from 5 to 13. With such small numbers, the size of the observed correlations is more important than statistical significance. This is because statistical significance depends on the number of cases. The results are reported in Table Twenty-nine. They indicate that four of the five coders obtained moderate to high correlations with student self-report estimates, while the fifth coder obtained zero order estimates.

Table Twenty-nine: Correlations between Coder assessments of Information Literacy and Law student self-report estimates.

Coder	Number	Pearson	Spearman	Kendall
1	8	0.60[11]	0.37	0.26
2	7	0.28	0.59	0.41
3	7	0.70*	0.79*	0.62*
4	13	0.03	0.18	0.09
5	5	0.38	0.41	0.32

The overall correlations were recalculated after deleting the results for the coder with zero order correlations with student estimates. The results were 0.46 for the Pearson estimate, 0.55 for the Spearman estimate, and 0.38 for the Kendall estimate. These estimates are all significant beyond the 1% level of probability.

[11] P=0.059
*P < 0.05

There are two factors that impose an upper limit on the possible correlations. For the Pearson product moment correlations, the reliability of the twenty-eight item self-report form is 0.84. As there will be some degree of variation in the assessments across the five coders, the reliability of their estimates will also be less than one. If the coder reliability is of the same order as the reliability for the student scale then the maximum Pearson product moment correlation is 0.71, which is the product of the estimated reliabilities. The correlations for five coders of 0.39 and for four coders of 0.46 are therefore satisfactory and indicate substantial agreement between the self-report data collected from students and the assessments of four coders out of five. For the non-parametric estimates of correlation, the maximum correlation is also limited by the form of the estimate to a value less than one. The Kendall estimate and the Spearman estimate can be considered as representing an upper and lower boundary for the non-parametric estimate of the correlation. Across the five coders the non-parametric estimate was between 0.27 and 0.41, while for the four consistent coders it is between 0.38 and 0.55. In both cases the Pearson estimate falls within the range of the estimates from non-parametric sources.

These results are evidence that four coders were able to achieve assessments that were consistent with each other and with the self-report estimates by students. Consequently it is concluded that the results provide satisfactory evidence of the concurrent validity of the Law form of the scale. In other words, the Law form of the ISS provides estimates of Information Literacy which are consistent with those made by four out of five coders.

The correlations between coder assessments of information literacy and student self-report estimates were also considered for each standard, with and without the one coder whose assessments were judged to be inconsistent. These results are reported in Table Thirty.

Table Thirty: Correlations between coder assessments of Information Literacy and Law student self-report estimates, by Standard

Standard	Number	4 coders		Number	5 coders	
		Pearson's r	Spearman's r		Pearson's r	Spearman's r
1	31	0.22	0.22	43	0.24	0.23
2	31	-0.01	0.13	44	-0.06	0.00
3	32	0.15	0.21	45	0.11	0.20
4	29	0.14	0.10	42	0.10	0.04
5	33	0.35*	0.35*	47	0.29	0.29
6	32	0.39**	0.45**	43	0.44**	0.51**

** beyond 0.01 level of significance
* beyond 0.05 level of significance

One aspect of these results is that for five of the six standards, the reliance on four coders rather than five results in a more positive correlation between coder assessments and student self-report. The exception is standard six where both with and without the fifth coder, there is a statistically significant correlation between coders and students.

Table Thirty-one: Comparison of observed Product Moment Correlations and Estimated Maximum Correlation for Law

Standard	Number of items	Kronbach Alpha	Pearson's r (4 coders)	Expected maximum correlation
1	4	0.46	0.22	0.35
2	4	0.58	-0.01	0.45
3	4	0.52	0.15	0.40
4	4	0.41	0.14	0.32
5^	6	0.63	0.35*	0.49
6^	6	0.58	0.39**	0.45

^ six items
*beyond 0.05% confidence interval
** beyond 0.01% confidence interval

Coders did not agree with self-assessments by students for Standard Two. This Standard addresses the search strategy, which is an area where the technical expertise of coders may contrast with the 'whatever works' approach of some students. A second aspect to note is that for Standards Five and Six, there were six items used to estimate student self-reports. This may have contributed to the higher correlation with coder assessments for these two Standards. Finally, it was noted that the Pearson product moment correlation for Standard Four was low. It is noted that the coders were assessing against the Outcomes specified in the first edition of the Information Literacy Standards. The student self-report estimates are based on items appropriate to the second edition of the Information Literacy Framework. Outcome Three from the original standards, that dealt with application of copyright requirements was relocated to Standard Six in the revised Standards. In addition it was the observed that some students had not been exposed to bibliographic software and techniques when they completed the survey, and hence the items that were used were more appropriate to students who had completed an initial semester of legal studies. These two issues may account for the lower correlation.

Another aspect is the comparison between the estimate of reliability for each Standard for the ISS and the correlation between coders and students. This comparison is also reported in Table Thirty-one. If the inter-coder reliability between coders is again estimated to be 0.8, then the expected maximum correlations are as reported in Table Thirty-one. In interpreting these estimates it should be noted that the expected correlations are estimates based on an assumption about the inter-coder reliability.

8. OTHER VALIDITY CONSIDERATIONS

It was noted in the introduction to the discussion of validity that Lally and Myhill (1994, 17) had identified four aspects of validity that were necessary to establish the quality of a survey. The above analysis has addressed the content validity of the survey, the construct validity of the scales and sub-scales and the concurrent validity as evidenced by Librarian assessments of a subset of the students who responded to each form of the ISS. Predictive validity is the fourth aspect to be considered.

8.1 Predictive Validity

The benefits claimed for enhanced information literacy include better learning outcomes for students. The available evidence was examination results. There was also the question of whether entry scores could predict the level of information literacy skills. Entry scores are also largely examination based.

For Education students, data was available for grades and for the ISS for two hundred students. For the comparison with tertiary entrance scores and scores on the ISS, data was available for one hundred and eighty-seven students. In both cases, there were zero order correlations with information literacy self-report. For entry score ($r = 0.04$) and for subject result ($r = -0.08$).There was however a positive and significant correlation of 0.275 between entry score and subject result.

For Law students data on the 28 item form of the information literacy scale were compared with the grades in the relevant law subject, and also with the tertiary entrance score for each student. Complete data was available for grades and for the ISS for one hundred and forty five students. For the comparison with tertiary entrance scores and scores on the ISS, data was available for one hundred and seven students. In both cases, zero order correlations were obtained.

This indicated that student scores on Information Skills Survey were not predicted by tertiary entrance score, and in turn did not predict the grade of result for the subject about which self-reported practice was collected. These findings were confirmed when the data was restricted to students aged under twenty who were presumed to be school leavers. It was concluded that the available evidence did not establish the predictive validity of the ISS. However, further research is needed to determine how information literacy impacts on student learning and outcomes. This therefore is not considered a limitation in the use of the scale, rather evidence of the need for such research to be undertaken.

8.2 Risks of Bias

In standard test design it is normal to investigate whether assessment tools may be biased in relation to gender, age, or ethnicity. This was undertaken to the extent possible with the available data. At both Central Queensland University and at the University of Melbourne, data from their student records were collected with the permission of participants. Student records were accessed to obtain biographical details, residence, entry scores and results in the subject about which they had reported on their information skills.

8.2.1 Assessment of Bias in Generic Form of ISS

In the case of the generic form of the ISS administered at Central Queensland University, data was available from the student records system for most students. After allowing for missing elements of data, analysis was possible on two hundred or more of the two hundred and twenty-two cases.

To establish whether there was a possible source of bias, comparisons were made with other available data. There was no significant difference between students on the basis of gender in terms of either entry score ($p = 0.64$) or results ($p = 0.45$).

TABLE Forty: Comparison of Entry Score and Subject Result for Education Students, Classified by Gender.

Background	Entry Score			Subject Result		
	Number	Mean	SD	Number	Mean	SD
Female	163	81.6	8.16	170	5.68	0.81
Male	36	82.3	7.82	44	5.57	1.00

The estimates of information literacy were compared between males and females. For the self report estimation obtained from the ISS, there was no significant difference in results ($p = 0.10$). As for the subject results reported above there was a non-significant advantage for females. It is concluded that there is no evidence that the ISS is biased by gender of respondents. The librarian estimates of information literacy were also compared by gender but the limited number of estimates available, after allowing for cases with missing values, meant that there were insufficient cases to draw conclusions. Nonetheless the results for both estimates of information literacy are reported in Table Forty-one.

TABLE Forty-one. Comparison of Information Literacy estimates for Education students Classified by Gender.

Background	Self-report			Librarian assessment		
	Number	Mean	SD	Number	Mean	SD
Female	162	64.3	6.35	27	13.71	1.74
Male	45	62.5	7.13	3	13.20	1.57

Small but significant relationships were observed between age and results ($r = 0.18$, $p < 0.01$), between age and entry scores ($r= 0.29$, $p<0.01$) and, as noted above, between Entry Score and results ($r = 0.28$, $p<.01$). Older students had higher entry scores and higher grades on average. However, when Entry score was used as a partial correlate, there was a zero order correlation between age and subject result. In considering the risks of bias in the ISS it was evident that differences in entry scores could be considered as a covariate. There was no significant correlation between ISS estimate and age ($r = 0.03$), and when entry score was used as a covariate the correlation was (-0.01). Hence it was concluded that there is no evidence of bias in the ISS in relation to age of respondents.

For ethnicity there were insufficient students identified as from a NESB and virtually no Indigenous Australians. Consequently, the possibility that the items discriminate on the basis of race or ethnicity could not be explored. An examination of the items identified no obvious sources of bias and hence the face validity of the items is acceptable. The language used is plain English that one might expect would be understood by a student enrolled at an English language institution, but experience suggests that readability should be investigated especially with overseas students from a non-English Speaking Background, and virtually no Indigenous Australians.

8.2.2 Assessment of Bias in Law Form of ISS

In the case of the Law Form of the ISS administered at the University of Melbourne, there was missing data in the biographical information obtained, which reduced the

number of cases for various analyses by between thirty and fifty percent. This data was collected at the commencement of the enrolment of students, and hence was not related to the information literacy data collection. For this reason it is assumed that there was no bias involved in using the cases for which data was available.

There were one hundred and seventy-three students for whom age and results were known. The Pearson's product moment correlation for age and subject results was 0.024 which was not significant. The narrow age range of most students in this cohort meant that the opportunity to identify an age factor may have been affected by the distribution. The minimum age[12] was 19 and the mean was 20.9 with the seventy-fifth percentile at 21. The distribution was markedly skewed and had exceptionally high kurtosis. All but eight of one hundred and seventy-five valid cases was aged twenty-five or younger. The correlation between age and student ISS assessment was of zero-order ($r = 0.09$).

Overall results for information literacy were examined by gender using Analysis of Variance. There were ninety-six females and seventy-seven males for whom data was available. There was no significant differences observed between gender and either mean entry scores ($p = 0.53$), or subject results ($p = 0.45$). For scores obtained on the Law ISS, there were one hundred and fifty-four cases with valid data and again there was no significant difference in mean scores between males and females ($p = 0.42$).

The data allowed for consideration also of the effect of English language background classified as either first language as English, or non-English speaking background. For University entry-score there was no significant difference on the basis of language background ($p=0.10$). The actual difference indicated that students from non-English speaking backgrounds had a marginally lower mean and a smaller standard deviation for entry scores than was the case for English speaking students. There were a total of one hundred and twenty-seven students for whom data was available for this comparison. When language background was considered in relation to the subject results, data was available for one hundred and seventy-three students. Non-English speaking background students had on average results that were lower both in terms of statistical significance and in terms of the effect size.

In relation to the Law ISS, data were available for thirty-nine students from non-English speaking backgrounds and for one hundred and fifteen for whom English was reported as their first language. The small difference in mean score in favour of English speaking students was not statistically significant. The effect size was 0.24, which is small and not statistically significant. It is of the same order and in the same direction as the effect size difference on entry scores ($ES = 0.27$) and is small compared to the effect size for subject results (0.80). Hence, the ISS does not appear to be biased in relation to self-reported information literacy skills because the difference based on language background is of the same order and is in the same direction as that obtained for entry scores. The results are reported in Table 42.

[12] For the purpose of the analysis age was calculated as at December 31. Some students would have been 18 when data was collected.

TABLE Forty-two: Comparison of Entry Scores, ISS Assessment and Subject Results for Law students, Classified by Language Background.

Measure	Language	Number	Mean	SD	ES
Entry Score	NESB	29	98.2	1.25	-0.27
	English	98	98.6	1.53	
ISS Assessment	NESB	39	70.72	9.34	-0.24
	English	115	72.91	8.97	
Subject Result[13]	NESB	41	62.39	7.86	-0.80
	English	132	68.63	7.75	

9. CONCLUSIONS

The analysis of the datasets reported above has confirmed that Information Literacy as defined by the Information Literacy Framework is a coherent concept. In each data set, an exploratory factor analysis demonstrated that all the content validated items loaded positively on the first factor with, at worse, a near zero loading.

The surveys have demonstrated statistical coherence sufficient to benchmark group achievement of overall Information Literacy outcomes specified in the Information Literacy Framework, 2nd. Edition. With the exception of Standard One for the general form of the survey, it is possible to describe the performance of groups of students in relation to each standard with moderate levels of reliability.

Within the limitations of the data, congeneric factor analysis has confirmed that items written to describe each Standard can be modelled to provide a sub-scale for each standard. However the design of this exploratory study did not provide for the range of difficulties of skills. Further, some items sought information about Information Literacy strategies to which first year students had yet to be exposed. Hence students could not provide confident responses about the opportunities to perform certain skills. These outcomes indicate that research aimed at further developing assessment tools for institutional monitoring of Information Literacy outcomes should be based on the Rasche model of scaling of latent variables. (Andrich, 1988).

The concurrent validity of both forms of the ISS has been confirmed by comparing the estimate of Information Literacy obtained from student self-reports with coder assessments of individual students. For the social sciences ISS scale, the results indicated substantial agreement between the self-report data collected from students and the assessments of coders. Despite some difficulties due to variations in the implementation of coding procedures, the overall correlation was 0.42 for all coders, and substantial agreement was evident on most sub-scales except where coders had not differentiated among students in terms of performance.

For the Law form of the ISS scale, four coders out of five were able to achieve assessments that were consistent with each other and with the self-report estimates by

[13] Difference was statistically significant beyond p = 0.01.

students. Consequently it is concluded that the results provide satisfactory evidence of the concurrent validity of the Law form of the scale.

Evidence of the predictive validity of the scales was sought by considering the correlation of self-reported information skills with students' entry scores and with their grade results in the relevant subjects. Zero order correlations were obtained. Although it might be expected that information literacy skills were relevant to exam grades, this assumes first of all that there is a relationship between information skills and learning outcomes as evidenced in course grades. In particular, it assumes that the course assessment relates to the ability to use information, as opposed to simply recalling information. There is no substantive research into this possible relationship. Further research is needed into how information literacy affects student performance. Performance in assignments that require searching for and using information, may be more likely to provide results that are correlated with level of information literacy. These investigations need to be made in a range of settings.

In terms of risks of bias, there was no differentiation between results for males and females. Further research is needed to establish in particular whether first nation peoples or overseas students from NESB may be disadvantaged by the language used in the generic self-report scale. However, on the Law scale it was concluded that the ISS is not biased in relation to people from a non-English speaking background.

It is therefore recommended that the generic form of the ISS be used with students enrolled in Social Science disciplines and that further data be collected to investigate the validity and reliability of the scale in these and other disciplines. For the Law form of the ISS, it is recommended that use be restricted to students who have completed their first semester of Law studies.

10. REFERENCES

American Psychological Association (APA) (1999). The Standards for Educational and Psychological Testing. AERA, Washington DC.

Andrich, D. A., (1988) Rasch Models for Measurement, Sage London.

Bollen, K. A. (1989) Structural Equations with Latent Variables. Wiley, NY

Bruce, C. (2003) Information Literacy as a Catalyst for Educational Change. Invited Paper, UNESCO Information Literacy Meeting of Experts, Prague. www.nclis.gov/libinter/infolitconf&meet/grouppaper.html

Bundy, A. (Ed) (2003) Australian and New Zealand Information Literacy Framework. ANZIL, Adelaide.

Byrne, B. (2001) Structural Equation Modelling with Amos. Lawrence Erlbaum, London.

Carroll, J.B., (1993) Human Cognitive Abilities. Cambridge University Press, NY.

Catts, R.. (2002) Issues in the Interpretation of the Australian Information Literacy Standards. Refereed paper at 2nd International Lifelong Learning Conference, Yeppoon, June 2002.

Catts, R. (2003a) Evaluation of Information Literacy at Institutional Level. Invited Presentation, Third Nordic Conference on Information Literacy, Akeruyri, Iceland, September 26, 2003.

Catts, R. (2003b) Assessment and Evaluation of Information Literacy at Institutional Level. Invited presentation to National Forum on Information Literacy, October 3, 2003, Washington DC.

Candy P., Crebert, G. and O'Leary, J. (1994) Developing Lifelong Learners Through Undergraduate Education, National Board of Employment Education and Training, AGPS, Canberra.

CAUL. (2001) Information Literacy Standards, 1st Edition, CAUL, Canberra. http://www.CAUL.edu.au/CAUL-doc/publications.html

CAUL. (2003) Information Skills Survey and Administration Manual, Council of Australian University Librarians, Canberra.

Embertson, S. E. and Reise, S.P. (2000) Item Response Theory for Psychologists. Lawrence Ebraum Assoc. NJ.

Ferguson, G. A. (1976) Statistical Analysis in Psychology and Education 4th Edition. McGraw Hill,

Higher Education Council, NBEET, (1992) Higher Education, Achieving Quality. AGPS, Canberra.

Holmes-Smith, P. (1999) Introduction to Structural Equation Modeling. SREMS, Elsternwick, Vic.

Holmes-Smith, P. and Rowe, K. (1994) The Development and Use of Congeneric Measurement Models in School Research. International Congress for School Effectiveness Research, Melbourne.

Hu Li-tzi, and Bentler, P.M. (1995) Evaluating Model Fit, Ch. 5 in Hoyle, R.H. (1995) (Ed.) Structural Equation Modeling. Sage, London

Kunkel, Lilith R., Weaver, Susan M., and Cook, Kim M. (1996) What do they Know? An assessment of undergraduate Library Skills. Journal of Academic Librarianship. 430-434

Lally, M. and Myhill M., (1994) Teaching Quality: The Development of Valid Instruments of Assessment. Dept. of Employment, Education and Training, AGPS, Canberra.

Lave, J. and Wenger, E. (1991) Situated Learning: Legitimate Peripheral Participation. Cambridge Press, UK

Markless, S. (2003) Information Literacy – a Learning-focused Approach. Keynote, Nordic Institute of Information Literacy Conference III, Akureyri, Iceland

MacNair, G (1990) The British Enterprise in Higher Education, Higher Education Management, 2, 60-71.

Marsh H. W. (1984) Students' Evaluations of University Teaching: Dimensionality, reliability, validity, potential biases, and utility. Journal of Educational Psychology, Vol 76 pp707-727.

Ramsden, P. (1979) Student learning and perceptions of the Academic Environment. *Higher Education* 8 (4) 411 - 27.

Ramsden, P. (1994) Using Research on Student Learning to Enhance Educational Quality. In Gibbs, G. (Ed) Improving Student Learning – Theory and Practice. Oxford Centre for Staff Development, Oxford University.

Weil, S. (1992) Creating Capability for Change in Higher Education, in Barnett, R. (Ed) Learning to Effect. Open University Press, Milton Keynes.

Wilson, K.L., Lizzio, A. and Ramsden, P. (1997) The Development, Validation and Application of the Course Experience Questionnaire. Studies in Higher Education Volume 22 (1) 33 - 53.

CAUL (Council of Australian University Librarians)
http://www.caul.edu.au/
caul@caul.edu.au