

Chapter 7 – Examination of actual practice – quantitative analysis

7 Examination of actual practice – quantitative analysis

7.1 Chapter overview

This chapter presents and discusses the findings of a quantitative examination of the effectiveness of a case of e-mentoring practice. The chapter undertakes an assessment of the validity and reliability of the quantitative measuring instrument and, having established that there is initial evidence for the instrument’s reliability and validity, addresses the evaluation questions set out in Section A. The process is informed and supported by items 6-7 of the research process set out in Chapter 4 and re-stated below. The chapter investigates whether the data provide empirical support for the DeLone and McLean (1992) dimensions as analysis categories appropriate to the quantitative measurement of the e-mentoring effectiveness construct, potentially providing a valid basis for constructing the sample of effective and ineffective mentee-mentor partnerships to be used in Chapter 8. **The small size of the sample being considered means that the analysis set out in this chapter is intended for the purposes of establishing initial linkages, and the triangulation of qualitative effectiveness data explored in Chapter 8.** That the sample is limited and insufficient for establishing statistical significance is acknowledged, and the findings are explicitly qualified by this limitation. **The chapter is not intended to stand alone, but to be considered in conjunction with the qualitative analysis set out in Chapter 8.**

Specifically, it will undertake each of these steps as follows:

1. using a selection of items from the survey questionnaire, undertake an initial assessment of its validity and reliability properties. This will in turn address the evaluation questions set out below:

Evaluation purpose for broader purpose of thesis	Section A – Evaluation questions
<p>To describe, categorize and interpret observations in line with DeLone and McLean’s model of IS effectiveness and the proposed framework. This includes examining the linkages between the dimensions with a view to identifying factors which are likely to influence effectiveness, and exploring the influence of context on outcomes</p> <p>To link outcomes arising from identified antecedents</p> <p>To explore any other issues which may be relevant to an understanding of effectiveness</p>	<p>Does this evaluation research confirm or disconfirm the existence of linkages between the dimensions of the DeLone and McLean model and effectiveness?</p> <p>Does the proposed taxonomy sufficiently accommodate the data, and if not, how?</p> <p>How were the mentoring partnership, program structure, user satisfaction, use and impact linked to effectiveness?</p> <p>Is there evidence of any linkage between Use and User satisfaction?</p> <p>Is there any evidence of any linkage between System and Information quality?</p>

2. using the effectiveness score, classify the sample into very effective, effective, partly effective and ineffective partnerships.

The data collection and analysis will be guided by the process set out in items 6-7 set out Framework guidelines, Chapter 4 and restated here for convenience:

6. Collect data

- *Obtain information from a variety of sources*
- *As appropriate, employ a variety of data collection methods*
- *Document and report information sources*
- *Document, justify and report data collection techniques and information sources*
- *Include data collection instruments in a technical appendix to the evaluation report*

7. Make findings

- *Present observations, descriptions, classification, categorisation, analysis and interpretation of data according to relevant framework or taxonomy*
- *Describe the program and its relevant pedagogical, technical, social, political, organisational and economic features/context*
- *Describe how the program actually functioned against how it was intended to function and discuss discrepancies*
- *Discuss issues of rigour*
- *Document and report any biasing features in the obtained information*
- *Report on reliability and validity - assess and report factors that influenced both*
- *Estimate and report the effects of validity and reliability in the data on the overall judgment of the program*
- *Make judgements about the program with reference to evaluative referent*
- *Consider alternative ways of interpreting evaluation findings*
- *Report limitations of the referenced information, analyses and inferences*
- *If appropriate, make recommendations*
- *Derive conclusions and demonstrate their meaningfulness*

7.2 Introduction

On the basis of the systematic classification of mentee experiences into degrees of effectiveness, a comparative analysis of the more effective and less effective mentee experiences will be undertaken and reported in Chapter 8 using qualitative inquiry. This will involve description and classification of quotational data arising from semi-structured interviews. The qualitative data will provide a basis for a detailed analysis and interpretation of the linkages between the mentoring process, the content or structure provided, use, user satisfaction, impact and effectiveness, against the correlations indicated by the quantitative inquiry. In this way, the associations explored in the qualitative analysis will be aligned with the associations suggested by the quantitative inquiry.

A validated measuring instrument provides a standardised evaluation mechanism (Scott 1995). Instrument validation helps “build a cumulative research tradition, provides improved measurement of research variables, helps improve the clarity of research questions and results in more meaningful variable relationships” (Scott 1995 p.44). This phase of the research, in Scott’s terms, required a standardised means by which the e-mentoring experiences of participants

could be classified into very effective, effective, partly effective and ineffective and to establish possible relationships between the dimensions of the construct of structured e-mentoring effectiveness as a precursor to the second qualitative phase. A search of the literature provided no published examples of validated reliable measurement instruments previously used in the context of evaluating structured e-mentoring effectiveness in the small business context. To advance research, the decision was made to rely on a newly developed instrument.

The survey instrument used in the quantitative analysis was based on the questionnaire used by the researcher in her role of practitioner evaluating the effectiveness of the e-mentoring program on behalf of the host organisation. The researcher mapped the questions and operationalised the effectiveness construct in line with the DeLone and McLean model of IS effectiveness to define a set of quantitative measures of effectiveness.

7.3 Limitations of quantitative analysis

As set out in Chapter 5, the lack of methodological sophistication arising out of small sample size and the failure to establish statistical significance means that it is not possible to generalise on the basis of these findings.

7.4 Findings - general

7.4.1 Response rate

The quantitative study was based on the survey questionnaire developed and administered by the program manager. A total of 32 mentee questionnaires were sent out by email over the course of the five years of the program's operation. A total of 20 mentee questionnaires were returned and form the basis of this quantitative study. The response rate was 62.5 per cent. A response rate of 30 per cent is generally regarded as acceptable in small business research (refer Curran & Blackburn 2001 in section 5.4.5). This response rate while acceptable in these terms, remains small therefore limiting the generalisability of the findings. Because this is an exploratory study seeking an understanding of influences on effectiveness which will focus on theory-building rather than theory-testing, the sample is regarded as adequate for making initial exploratory investigations.

7.4.2 Characteristics of the sample for the quantitative study

This section sets out the characteristics of the sample and considers how the sample compares with self-employed contractors generally - it defines the characteristics of the sample which are shared by self-employed contractors as a group.

7.4.2.1 Gender and age of mentee respondents

Table 54 - Comparison of percentage distribution of self-employed contractors and self-employed contractors working as Professionals by gender characteristics for Australian population at August 1998 compared with those participating in e-mentoring program, 2002-2006

Gender	Self-employed contractors in Australian population (per cent)	Professionals operating as self-employed contractors (per cent)	Self-employed contractors in Mentoring program (per cent) (n=20)
Male	70.7	65.0	65.0 (n=13)
Female	29.3	35.0	35.0 (n=7)

As set out in Table 54, in terms of gender breakdown, women professionals operating as self-employed contractors appeared to be slightly over-represented in the sample of mentoring program participants when compared with self-employed contractors in the general Australian population. However, when compared with self-employed contractors working as professionals, there was no difference.

Table 55 - Comparison of percentage distribution of self-employed contractors and self-employed contractors working as Professionals by age characteristics for Australian population at August 1998 compared with those participating in e-mentoring program, 2002-2006

Age	Self-employed contractors in Australian population (per cent)	Professionals operating as self-employed contractors (per cent)	Self-employed contractors in Mentoring program (per cent) (n=20)
15-44	59.2	57.0	70.0 (n=14)
45 and over	40.8	43.0	30.0 (n=6)

(Source of Australian population information set out in Tables 54 and 55: ABS data derived from Productivity Commission Report, 2001, p. 41 and data specific to Professionals provided to Professional Association by L. Will, Productivity Commission report co-author, 2002)

The age of participants was recorded when they began the mentoring program, and the age ranges of participants compared with Self-employed contractors in the Australian population generally, and Professionals as a specific occupational category is set out in Table 55. The data indicate that for those in the 15-44 age range, the participation rate in the mentoring program was higher than their incidence in the Professionals operating as self-employed contractors (70 per cent compared with 57 per cent) and slightly higher than their incidence in the Self-employed casuals in the Australian population generally (70 per cent compared with 53.9 per cent). ABS data for professionals was used to compare the age range of participants in the mentoring program with the age distribution of Professionals operating as self-employed contractors. Those in the 45 and over range participating in the mentoring program were underrepresented in the sample compared with Professionals operating as self-employed contractors (30 per cent compared with 43 per cent) and Self-employed contractors in the Australian population (30 per cent compared with 40.8 per cent).

7.4.2.2 Geographical location of mentee respondents

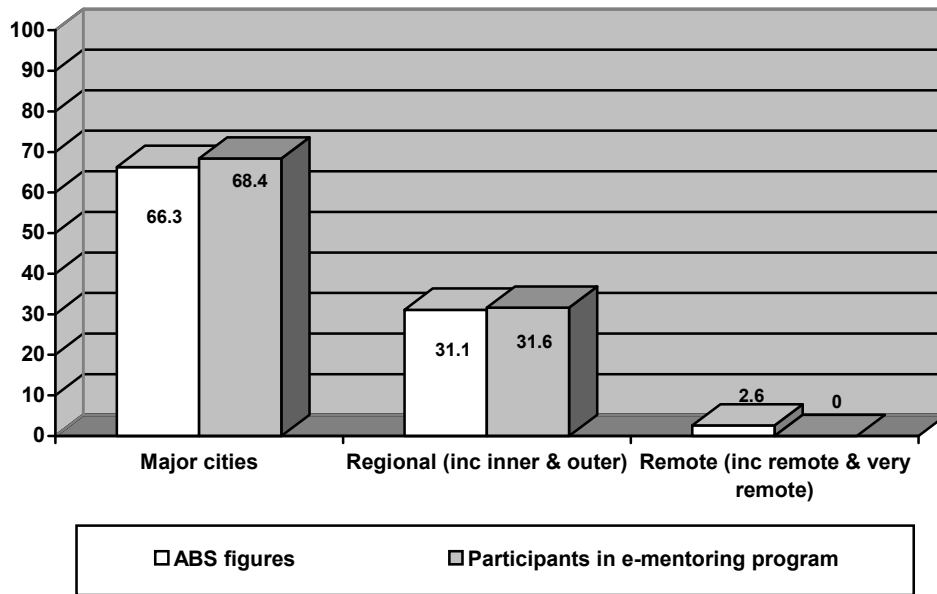


Figure 4 – Geographical location of e-mentoring program participants
 (Source of Australian population information: ABS data derived from ABS Cat. No. 1379.0.55.001 (National Regional Profile, Proportion of population in remoteness area))

31.6 per cent of the sample were located in rural or regional areas. 68.4 per cent were located in rural or regional locations. There is no significant difference between the distribution of those participating in the e-mentoring program according to their geographical location compared with the Australian population more generally (Note n=19 as one mentee was located overseas and this mentee was not included in the remote category). Again, the small number of data points under consideration limits this comparison.

7.4.2.3 Profession groups

The ABS uses the Australian Standard Classification of Occupations (ASCO) (ABS Cat. No. 1221.0 to define occupations.

Table 56 - Summary of ASCO minor group occupations eligible for professional association membership (n=20)

ASCO minor group classification	Profession	Number of professionals participating in mentoring program
2124, 2125, 2126, 2127, 2129	Professional Engineers	9
2111, 2113, 2114 & 2119	Scientists	5
2121	Architects	0
2231	Computing Professionals	3
2382	Pharmacists	1
2392	Veterinarians	1
Other	eg Managers	1
Total		20

As set out in Table 56, mentees who participated in the mentoring program were drawn from the membership of a special interest group for self-employed contractors. As set out in Table 55, they were predominantly professional engineers and scientists, but included a pharmacist, veterinarian and information technology professionals. The pool of mentors was drawn from both within and outside the Association’s membership, all held tertiary qualifications in their specialist areas and had themselves run a consultancy for a minimum of three years.

The evaluation may be limited in being a sample of predominantly engineers. Data on how these individual professions may generalise to professionals operating as self-employed contractors more generally is not available.

7.4.2.4 Education level of mentee respondents

ABS data on post-graduate education level of self-employed contractors is not available meaning a comparison of the education level of the sample with the population of self-employed contractors is not possible. This information is nonetheless provided to define the sample. (Note: Professionals eligible to join the professional association must have a minimum Bachelor degree.)

Table 57 – Education level

Education level	Number of participants in the sample (n=20)
Bachelor degree	9
Post-graduate diploma or certificate	2
Masters degree	3
Doctorate level	6
Total	20

As set out in Table 57, of the sample, 9 had a Bachelor degree, 2 also had a post-graduate diploma, 3 had a Masters degree and 6 had a PhD or doctorate.

7.4.2.5 Representativeness of sample

On the basis of the data set out in Tables 54, 55, 56, 57 and Figure 3, it is possible to make claims as to its representativeness in terms of gender and geographical location and in terms of these characteristics, the sample potentially provides a basis for generalising findings to professionals operating as self-employed contractors. It should again be noted however that claims of representativeness are limited by small sample size. The sample’s representativeness in terms of age, profession, post-graduate education, membership of a professional association and stage in the business life cycle cannot be established. As a minimum, these factors threaten the generalisability of the evaluation findings, and the external validity of claims made in relation to self-employed contractors. However because this is an exploratory study and the

sample is small, the decision was made to include all the questionnaires returned by respondents rather than further reducing the sample size by using stratified sampling. This means that the external validity of the evaluation was compromised in favour of maximising the size and diversity of the sample.

Eligibility requirements of the mentoring program

Two other significant characteristics comprised the eligibility requirements for the program and therefore the demographic characteristics of the sample. The first was the requirement (beyond the pilot program) to be a current financial member of the professional association hosting the program, and the second was to be within the first three years of business startup. Both these factors further particularise the sample. Claims made will therefore be largely in relation to this sample rather than to Professionals operating as self-employed contractors or self-employed contractors.

Heterogeneity and representativeness

As stated previously, the small business sector is characterised by the uniqueness of each business in terms of size, profit and turnover, whether home or office-based, the type of business engaged in, the industry sector in which they operate, in the products and services produced, in the processes and level of technology used, and in the specific community and business environment in which they are located (Tolentino 1998 p.3, Devins & Gold 2000 p.251). To accommodate such diversity would require a very large sample. Obtaining sufficiently large samples is of course a difficulty in this research as it is with much small business research (Curran & Blackburn 1994). This heterogeneity potentially compromises inferential power and the capacity to make generalisations to self-employed professionals, and this is acknowledged as a limitation of this study as it is in much small business research.

7.5 Main findings

The questionnaire responses were coded, scored and are summarised in Table 60 – Participant scores. The complete responses of each participant according to the questions under each of the dimensions is included as Appendix 6. As stated in Chapter 5, Patton considers evaluation research to involve a “tradeoff between breadth and depth” (Patton 1990 p.165) and suggests that “[t]he design issue is how much time and effort we are willing to invest in trying to increase our understanding about any single person’s experience” (Patton 1990 p.165). To reflect each person’s experience in the data presentation, the decision was made to present an aggregate summary of the numerical data arising from the questionnaire alongside a descriptive summary of effectiveness for individual participants which is presented in Table 61.

7.5.1 Data summary

Table 58 sets out the key to the scoring system. In the case of each of the dimensions, the thresholds were set with reference to what were regarded as appropriate minima in the particular program setting. Key indicators were selected according to the program goals and the thresholds were in some cases subject to review and realignment.

In the case of the dimension of System quality (or Quality of mentee/mentor interaction), a mentee's experience would be rated as poor if they indicated that they failed to have experiences in common with their mentor, did not build a good relationship with their mentor, were not satisfied with the advice offered by the mentor and saw their mentor as having limited competencies. Conversely, the scoring system was designed that the mentee's experience would be regarded as excellent if they indicated that they had experiences in common with their mentor, built a good relationship with their mentor, were satisfied or very satisfied with the advice and assistance provided by the mentor, and indicated that their mentor had broad-ranging quality mentoring competencies.

In the case of the dimension of Information quality (or Program support quality), a mentee's experience would be rated as poor if they did not value the content of the facilitator's messages, did not set program goals, did not develop the skills they nominated in their registration form, found the program's duration inappropriate, found the pre-program training unsatisfactory or limited, did not find that the email-based nature of the program facilitated their participation, and were not satisfied with the matching process or access to further resources. Conversely, a mentee's score was regarded as appropriately high if they indicated broad-ranging high-level satisfaction in each of these areas.

In the case of the dimension of User satisfaction, the scoring system was designed so that a mentee's experience was rated as poor if they indicated that they did not find the program useful or relevant, would not participate again nor recommend the program to a colleague, did not feel they developed personally or professionally, rated the value of the service as poor or very poor and did not use the opportunity to ask questions or bounce ideas off their mentor. Conversely, the mentee's score was graded as high if they indicated high levels of satisfaction in each of these areas.

Because the previous literature indicated that regular contact was critical to effectiveness, as a minimum, the scoring system was designed so that satisfaction with the level of contact from the program host or the mentor was required for the mentee's experience to be scored as reasonable or above. Where a mentee was satisfied with the level of contact from both the

program host and mentor, the scoring system was designed to grade the mentee’s user satisfaction as high.

The thresholds for the Impact dimension were set with reference to questions in which mentees described on nominated outcomes against a pre-set list. Where a mentee nominated broadranging outcomes and indicated that they reviewed their business plan, kept in contact with their mentor beyond the three-month program, and felt they benefited from the program, the scoring system was designed to grade Impact as reasonable or above. Conversely, where the mentee did not nominate positive outcomes in the specified areas, the scoring system was designed to reflect the fact that outcomes were limited.

It is important to note when considering the rationale for scoring systems that the process of setting thresholds is an iterative one necessarily involving the judgement of the program developer. It is also a process which should be guided by the particular goals and anomalies of a program and one which should be regarded as legitimately discretionary mindful of the need for an open, accountable and ethical approach by the program developer.

The summary score sheets attached in Appendix 6 provide explicit detail on the practical implementation of the rational behind the scoring system.

Table 58 - Key to scoring system – E-mentoring Scoring Scale

System quality – quality of relationship between mentee and mentor	
Excellent	18-21
Good	15-17
Reasonable	12-14
Poor	0-11
Information quality – engagement with and adaptation of program structure and content	
High	53-92
Reasonable	31-52
Limited	0-30
User satisfaction	
High	10-12
Reasonable	6-9
Low	3-5
Use	
High	6-7
Reasonable	4-5
Low	0-3
Impact	
Positive	17-24
Reasonable	8-16
Limited	0-7
Total score range	Description
0-45	Ineffective
46-97	Partly effective
98-129	Effective
130-156	Very effective

Table 59 sets out an overview of the nature of scores in quantitative terms.

Table 59 – Summary of scores

Maximum score	141
Minimum score	37
Mean	97.8
Standard deviation	33.2

The mean as the measure of central tendency was 97.8. Scores ranged from 37 to 141. The standard deviation as a measure of dispersion was 33.2. This indicates that two-thirds of respondents' scores lie between 64.6 and 131.0.

Table 60 sets out the scores for each dimension and the total effectiveness score by participant in ascending order. The scoring sheets from which this data is derived are set out in Appendix 6.

Table 60 – Participant scores

Participant number	Year of participation	System quality	Information quality	User satisfaction	Use	Impact	Total score
9	2002	4	27	3	2	1	37
17	2005	20	6.5	8.5	1	7	41
12	2003	15	8	9	4	5	41
5	2002	16	23	11	2	12	64
8	2002	19	31	11	1	9	71
19	2005	19	38.5	6.5	1	7	72
1	2002	13	61	11	4	3	92
15	2004	6	82	3	2	3	96
4	2002	9	60	11	5	12	97
7	2002	19	51	11	5	14	100
10	2003	20	53	11	2.5	17	103.5
6	2002	14	80	8	0	7	109
14	2004	20	74.5	12	3	10	119.5
11	2003	20	74	10	6	12	122
2	2002	20	61	12	6	25	124
13	2004	21	74.5	10.5	5	17	128
18	2005	19	79	12	6	13	129
3	2002	16	82	12	5	17	132
16	2005	21	79	12	6	18	136
20	2006	21	91.5	9.5	5	14	141

Summary of effectiveness for participants

Table 61 represents an interpretation of effectiveness for individual participants in descriptive form. The qualitative descriptions are defined in the key outlined in Table 58. The data is set out in ascending order from lowest effectiveness score to highest.

Table 61 – Data summary - description of effectiveness across the DeLone and McLean dimensions

Participant number	System quality	Information quality	User satisfaction	Use	Impact
9	Poor quality relationship between mentee	Limited engagement with the program	Low level of user satisfaction	Low level of regular interaction	Limited outcomes

	and mentor	structure and content			
17	Excellent relationship between mentee and mentor	Limited engagement with the program structure and content	Reasonable level of user satisfaction	Low level of regular interaction	Limited outcomes
12	Reasonable relationship between mentee and mentor	Limited engagement with the program structure and content	Reasonable level of user satisfaction	Reasonable level of regular interaction	Limited outcomes
5	Good relationship between mentee and mentor	Limited engagement with the program structure and content	High level of user satisfaction	Low level of regular interaction	Reasonable outcomes
8	Excellent relationship between mentee and mentor	Reasonable engagement with the program structure and content	High level of user satisfaction	Low level of regular interaction	Limited outcomes
19	Excellent relationship between mentee and mentor	Reasonable engagement with the program structure and content	Reasonable level of user satisfaction	Low level of regular interaction	Limited outcomes
1	Reasonable relationship between mentee and mentor	High level of engagement with the program structure and content	High level of user satisfaction	Reasonable level of regular interaction	Limited outcomes
15	Poor relationship between mentee and mentor	High level of engagement with the program structure and content	Low level of user satisfaction	Low level of regular interaction	Limited outcomes
4	Poor relationship between mentee and mentor	High level of engagement with the program structure and content	High level of user satisfaction	Reasonable level of regular interaction	Reasonable outcomes
7	Excellent relationship between mentee and mentor	Reasonable level of engagement with the program structure and content	High level of user satisfaction	Reasonable level of regular interaction	Reasonable outcomes
10	Excellent relationship between mentee and mentor	Reasonable level of engagement with the program structure and content	High level of user satisfaction	Low level of regular interaction	Positive outcomes
6	Reasonable relationship between mentee and mentor	High level of engagement with the program structure and content	Reasonable level of user satisfaction	Low level of regular interaction	Limited outcomes
14	Excellent relationship between mentee and mentor	High level of engagement with the program structure and content	High level of user satisfaction	Low level of regular interaction	Reasonable outcomes
11	Excellent relationship between mentee and mentor	High level of engagement with the program structure and	High level of user satisfaction	High level of regular interaction	Reasonable outcomes

		content			
2	Excellent relationship between mentee and mentor	High level of engagement with the program structure and content	High level of user satisfaction	High level of regular interaction	Positive outcomes
13	Excellent relationship between mentee and mentor	High level of engagement with the program structure and content	High level of engagement with the program structure and content	Reasonable level of regular interaction	Positive outcomes
18	Excellent relationship between mentee and mentor	Excellent relationship between mentee and mentor	High level of user satisfaction	High level of regular interaction	Positive outcomes
3	Good relationship between mentee and mentor	High level of engagement with the program structure and content	High level of user satisfaction	Reasonable level of regular interaction	Positive outcomes
16	Excellent relationship between mentee and mentor	High level of engagement with the program structure and content	High level of user satisfaction	High level of regular interaction	Positive outcomes
20	Excellent relationship between mentee and mentor	High level of engagement with the program structure and content	Reasonable level of user satisfaction	Reasonable level of regular interaction	Reasonable outcomes

The following figure sets out a frequency distribution of scores for the ranges 0-45, 46-97, 98-129 and 130-141.

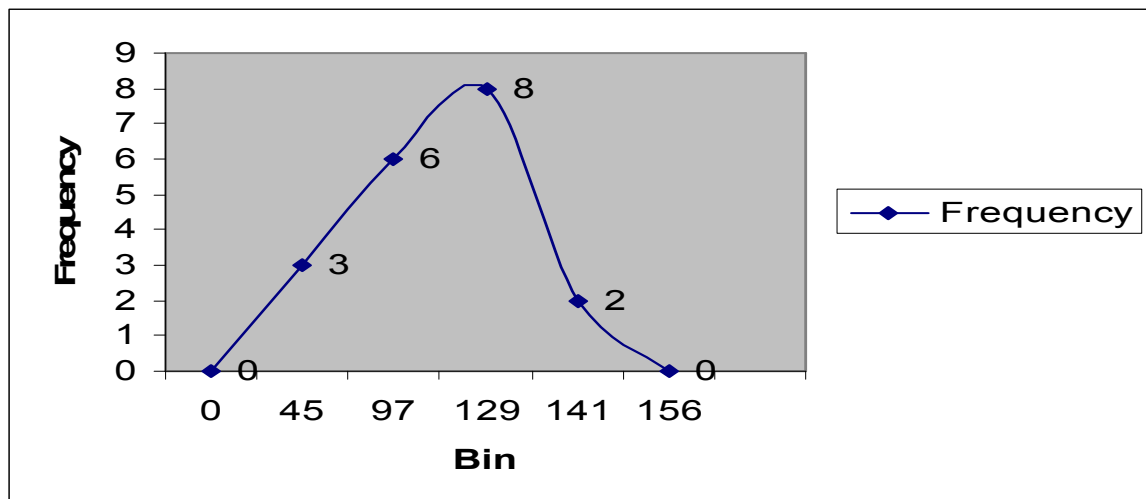


Figure 5 – Spread of effectiveness scores

Figure 5 shows a reasonable spread of total effectiveness scores between the minimum value of zero and the maximum value of 156 indicating the measurement instrument has, as a whole, appropriately measured variance in responses. Responses are skewed to the right showing a potential measurement bias, but again, the small number of data points limits the claims which can be made about the data. The ranges are based on the ranges set out in Tables 58 and 70.

7.5.2 Data analysis – validity and reliability of survey instrument

Because of the sample size, it is not possible to statistically establish whether or not the instrument has validity and reliability. Because the survey instrument was developed and used for the first time to evaluate the effectiveness of this e-mentoring program, its properties have not statistically been rigorously assessed. The tests and discussion of the instrument which follow do not assess conclusively the reliability and validity of the instrument. For such claims to be maintained with any confidence, each item and then each dimension of the survey instrument would need to be subjected to statistical procedures with a larger sample. The validity and reliability of the scores are therefore open to challenge on the basis of measurement error and/or measurement bias.

However, the following discussion and rudimentary statistical analyses of a selection of the questionnaire items used to determine the effectiveness score provide preliminary evidence of the validity and reliability of the instrument, and provided a basis for identifying some of its possible limitations or sources of invalidity and unreliability.

The validity of the quantitative measurement instrument will be considered in relation to the level of agreement between scores and the judgements of the program manager and mentors and with reference to inter-item Pearson correlations between the DeLone and McLean dimensions. The reliability of the measurement instrument will be considered with reference to Cronbach alpha and comparing the use of the measurement instrument in an international context. The main sources of validity and unreliability will then be discussed.

To assist with clarity in section 7.5.2, the following sets out the structure of section:

7.5.2 Data analysis – validity and reliability of survey instrument

7.5.2.1 Evidence of validity

- 7.5.2.1.1 *Level of agreement between scores and judgements of program manager and mentors*
- 7.5.2.1.2 *Rating of mentors*
 - 7.5.2.1.2.1 *Instances of disparity*
- 7.5.2.1.3 *Inter-item correlations*
 - 7.5.2.1.3.1 *Aim of inter-item correlations*
 - 7.5.2.1.3.2 *Limitations with correlation analysis*
 - 7.5.2.1.3.3 *Scatterplot graphs*
 - 7.5.2.1.3.3.1 *Relationships between effectiveness and the DeLone and McLean dimensions*
 - 7.5.2.1.3.3.2 *Relationships between Impact and other DeLone and McLean dimensions*
 - 7.5.2.1.3.3.3 *Discussion of relationships found in scatterplot graphs*
 - 7.5.2.1.3.4 *Conclusions about inter-item correlations*

7.5.2.2 Evidence of reliability

- 7.5.2.2.1 *Cronbach alpha*

- 7.5.2.2.2 *Use of survey questionnaire in international context*
 - 7.5.2.2.2.1 *Definition of the effectiveness construct*
 - 7.5.2.2.2.2 *Formative emphasis*
 - 7.5.2.2.2.3 *Comparison with data arising from use of quantitative instrument in international setting*
 - 7.5.2.2.2.4 *Testing*
 - 7.5.2.2.2.5 *Program implementation*
 - 7.5.2.2.2.6 *Summary of evidence of reliability in international setting*
- 7.5.2.2.3 *Does the measurement instrument behave as expected?*

7.5.2.3 Limitations or sources of invalidity or unreliability

- 7.5.2.3.1 *Construct underrepresentation*
- 7.5.2.3.2 *Aim of the survey instrument – to measure or rank*
- 7.5.2.3.3 *Contextual analysis*
 - 7.5.2.3.3.1 *Gender*
 - 7.5.2.3.3.2 *Geographical location*

7.5.2.1 Evidence of validity

This section will consider whether or not the evidence supports the claim that the quantitative measurement instrument measures what it is intended to measure.

7.5.2.1.1 Level of agreement between scores and judgements of program manager and mentors

Brualdi (1999) suggests that “[e]mpirical evidence in support of criteria-related validity may include a comparison of performance on the test against performance on outside criteria” (p.1). In order to test the validity of the effectiveness score, the score was compared with data from additional sources. Table 62 sets out the effectiveness of the program for the participant against the program manager and mentor’s rating of the effectiveness of the program. These ratings are in qualitative form rather than in the form of numerical criteria. A discussion of the disparities follows.

7.5.2.1.2 Rating of mentors

The judgements of mentors were provided to the researcher further to an email specifically asking mentors to nominate the degree of effectiveness for the partnership or partnerships in which they had participated. No current contact details were available for the mentor of Participants 5, 7 and 8 so the judgement was recorded as unavailable.

Table 62 – Triangulation of data around effectiveness

Participant	Score	Mentee’s perceived level of effectiveness	Program facilitator’s judgement	Mentor’s judgement
Participant 1	92.0	Partly effective	Partly effective	Partly effective
Participant 2	124.0	Effective	Effective	Effective
Participant 3	132.0	Very effective	Very effective	Partly effective
Participant 4	97.0	Partly effective	Effective	Partly effective
Participant 5	64.0	Partly effective	Ineffective	Data unavailable
Participant 6	109.0	Effective	Effective	Ineffective
Participant 7	100.0	Effective	Effective	Data unavailable
Participant 8	71.0	Partly effective	Ineffective	Data unavailable
Participant 9	37.0	Ineffective	Ineffective	Ineffective
Participant 10	103.5	Effective	Very effective	Very effective

Participant 11	122.0	Effective	Effective	Effective
Participant 12	41.0	Ineffective	Effective	Very effective
Participant 13	128.0	Effective	Effective	Effective
Participant 14	119.5	Effective	Effective	Ineffective
Participant 15	96.0	Partly effective	Partly effective	Partly effective
Participant 16	136.0	Very effective	Very effective	Effective
Participant 17	41.0	Ineffective	Ineffective	Very effective
Participant 18	129.0	Effective	Very effective	Very effective
Participant 19	72.0	Partly effective	Effective	Partly effective
Participant 20	141.0	Very effective	Very effective	Partly effective

7.5.2.1.2.1 Instances of disparity

Participant 12

Participant 12's score indicates that the mentee found the mentoring partnership ineffective. In contrast, the program manager's view was that the partnership was effective. The program manager's view was informed by the knowledge that an ongoing business strategic alliance between the mentee and mentor arose out of the mentoring partnership which has continued since the mentee's participation in the program in 2003: "My previous mentee and I are planning joint consultancies together in the asset management field" (email July 2006) and "[t]he mentorship was very effective, and [the mentee] is now a very successful consultant who I work in alliances with [interstate]" (email January 2007). Interestingly the mentor, as well as the program host, judged the mentoring program to be effective. This is the clearest example of a disparity between the mentee's score and the views of the program manager and mentor. Time constraints and competing priorities, which led to a failure on the mentee's part to engage with the program structure during the course of the program as reflected in the questionnaire administered immediately after the program, are the most likely explanations for the low score. The mentee commented that: "Good information was provided. Unfortunately, I didn't make as much use of it due to hectic time" (response to open question on questionnaire). It is possible that, in this case, the value of the mentoring partnership may have been better measured over a longer period rather than immediately following the program to capture evolving or accruing benefits which have obviously arisen out of the program.

Participant 5

Participant 5's score indicates that the program experience was partly effective while the program manager viewed it as ineffective. The disparity in the program manager's view of the partnership as less effective than indicated by the score arose out of discussions with the mentor who indicated that the mentee had not been in regular contact and had failed to set firm program goals by the checkpoint at Week 7 (refer 6.4.3.1.2). On being prompted, the mentee did formulate a specific program goal and the mentor acknowledged this progress. The following exchange occurred in June 2002:

Prompt from Mentors Online:

Will you confirm with me as soon as possible that you are still interested in participating in Mentors Online or whether you'd prefer to withdraw at this stage. Further to the Partnership Status Checklist sent to Mentees and Mentors on Wednesday 12th, indications are that you and your mentoring partner have not yet set firm program goals - these need to be in place at this stage for the program to function effectively. You can opt to complete all four structured exercises in preference to setting your own personal goals if that is your preference. It is important to invest some time in the mentoring relationship to ensure that you get the most out of it.

You may prefer to consider participating in next year's program if other commitments are affecting your participation in the pilot program. Please be assured that a withdrawal from this year's program would in no way be seen as a reflection on your commitment to a future program.

Please arrange with your Mentor to establish some firm program goals and some discussion topics which you think will be of value. Then contact me as a priority so I know that your program is either back on track or we can close the mentoring partnership down. If I don't hear back from you by June 21st, I'll advise your Mentor that we should formally conclude your partnership (email 14.6.2002).

Mentee response (extract)

The goal of our mentoring relationship is to put together an initial website for my business. I have been working on content and have about 70% of this done. My Mentor is assisting me to get the rest together. Work is variable, and I am doing what I can when possible. I suppose the next step is registering a web address with a Host. I am still looking at who to host it" (email June 2002).

Mentor response

That is certainly great progress ... though she has not communicated with me since 4 June. I am happy to continue in this role ... I just wanted to ensure that she received some value from the mentoring relationship (email June 2002).

Participant 8

Participant 8's score indicates that the program was partly effective while the program manager viewed it as ineffective. Because there was limited engagement with the program host by this participant, the program host's summation of effectiveness was based on the fact that in spite of regular encouragement to provide feedback to and engage with the host, there was only basic communication between host and participant in the case of this partnership. This suggests that the program can be effective for the participant without regular or indeed any interaction with the program host. In turn the program manager's view of effectiveness may be subject to bias and error in these terms.

Participant 4

Participant 4's score indicated that the program was only partly effective while the program manager viewed it as effective. The disparity between the score and judgement of the program manager can be accounted for by impressions made by comments of the mentee to the host:

The exercises that you provide have really helped me. X and I had discussed my business plan quite a bit and I had been wondering 'what next'. So I mentioned networking as per the next exercise. The discussions with X really opened my eyes as to how to maximise opportunities. I've realised that this is really the area that will help my business to grow. In general, it's just been great to have someone to bounce ideas off" and "Thanks for the prompt; your timing is impeccable. This is an excellent time to review progress and future direction" (emails to host, June 2002).

Interestingly, the mentor suggested that the mentee would be likely to perceive the partnership as effective to very effective (email January 2007). This may suggest that the survey instrument may have led to an understatement of effectiveness for this participant.

Participant 10

Participant 10's score indicated that the mentee found the program effective while the program host and mentor classified it as very effective. The disparity between this level of effectiveness and that judged by the program host can be accounted for by comments made to the program manager: "I guess it has come at just the right time for me, which is why I am excited. And I have a couple of friends in similar positions and I will be sharing what I learn with them ... so it could be an all round growing experience" (email to host, July 2003). It is possible that the program did not meet the high expectations of this mentee, and that this was reflected in the lower than expected effectiveness score.

Participant 18

Participant 18's score indicated that they found the program effective while the mentor and program manager described the partnership as very effective. The disparity between this level of effectiveness and that judged by the program manager can be accounted for by comments made by the mentee to the program manager which would suggest that the mentee found the program very effective rather than effective:

I felt the existence of a shared journey even though our areas of expertise are 180 degrees apart. My mentor had travelled the same paths as me in terms of family, demanding schedule, lots of overseas travel, working as an employee when my skill set would be more suited to running my own company. I experienced renewed interest in business and learning how to run and manage a company, had been feeling jaded and isolated, but realised these feelings were experienced by others who had similar energy levels and drive but were not in the optimum work situation.

I also experienced a major and long overdue shift in my perception of my working life and realised I had been stuck and become disinterested due to a work environment which is based on micro-management, bottom line and politics. My enthusiasm is for the top line. I am also involved in a pre-seed research project which meant I was effectively working 1½ jobs – with the accompanying tiredness. My interactions with [my mentor] renewed my energy and enthusiasm and the courage to persist.

Her advice on time and financial management was spot on and I am still integrating her suggestions into my daily activities. The fact that she had experienced almost identical

workplaces, life experiences and challenges established an immediate and strong rapport, without which I think I would have not had so much trust and confidence in her advice (response to open question on evaluation questionnaire, November 2005).

This suggests that, as was the case with Participant 10, the evaluation instrument may have understated the level of effectiveness for this participant.

Participant 14

In this case, the mentor considered the program for the mentee to be ineffective because the partnership was concluded prior to the end of the program. In spite of the fact that the program was concluded early, the mentee via the questionnaire indicated that the program was effective and this was also the view of the host. The score in this case appeared to be a better indicator of effectiveness than the mentor's judgement.

Participants 6 and 17

Participant 6

In this case, the mentor considered the program for the mentee to be ineffective while the score and host indicated that the program was effective. Of note in this instance was the high level of engagement with the program structure alongside only a reasonable relationship with the mentor.

Participant 17

In this case, the mentor considered the program for the mentee to be very effective while the score and the host indicated that the program was ineffective. Of note in this instance was the limited level of engagement with the program structure alongside an excellent relationship with the mentor. This instance of disparity more than any other suggests an inadequacy of the instrument to sufficiently capture the extent of the System quality – the nature and quality of relationship between mentee and mentor - where the program structure is essentially disregarded in the program structure adaptation process.

The most significant disparities between the judgements and score occurred in the cases of Participants 6 and 17. The disparity in the case of Participant 17 suggests that effectiveness can be marked by a very strong relationship with the mentor (where the mentor indicates that they believed the partnership was very effective for the mentee) alongside a very low level of engagement with the program structure. This was mirrored by the disparity in the case of Participant 6 which was marked by only a reasonable relationship with the mentor (where the mentor indicates that they believed the partnership was ineffective for the mentee) alongside a

high level of engagement with the program structure. The scores failed to adequately capture the complexity of effectiveness in these areas.

7.5.2.1.4 Summary – evidence of validity

The general convergence of the program manager and mentors' perceptions of the effectiveness of the program for the mentee with the effectiveness score arising from the measurement instrument confirms that there is a level of convergence and agreement between the judgements and the effectiveness score arising out of the questionnaire.

The process of data triangulation established instances where the effectiveness score was confirmed and disconfirmed by the judgements of mentor and program manager and the disparities were discussed. Most critically, the instrument failed to adequately capture the very effective relationship between mentor and mentee in the case of Participant 17 suggesting that the instrument was limited as a means of quantifying effectiveness in the area of the relationship between mentee and mentor. Analysis of this initial data also suggested that while there was a relationship between System quality and effectiveness, and Information quality and effectiveness, positive outcomes in *both* these areas was not necessarily required for the program to be effective for the mentee. There was also some evidence to suggest that the threshold score ranges set out in Table 70 could be revised downward to better concord with the stated views of mentees on their perceptions of effectiveness. This evidence is, however, in contrast to the possible bias indicated in the spread of effectiveness scores set out in Figure 5. Overall, there is some empirical support for the validity of the survey questionnaire in measuring what it is intended to measure.

7.5.2.1.3 Data analysis - inter-item correlation

7.5.2.1.3.1 Aim of inter-item correlation

Pearson correlations between each of the DeLone and McLean (1992) dimensions and effectiveness will be computed to confirm or disconfirm the strength and direction of relationships between each of the DeLone and McLean dimensions. If positive inter-item correlations are found between each of the dimensions and effectiveness, the proposition that the effectiveness score is a measure of the dimensions of the construct of effectiveness would be confirmed.

7.5.2.1.3.2 Limitations with correlation analysis

There is an inadequacy in the calculations regarding effectiveness in the correlation analysis because the variables of effectiveness and the other dimensions are not measured independently. The scores for the individual dimensions are also components of the effectiveness score. While

this means the integrity of the data is compromised, the inferences made on the basis of the data still have some resonance and basis in empirical data. To investigate the relationships between the DeLone and McLean dimensions and confirm or disconfirm whether the data supports the relationships in the e-mentoring effectiveness context, the correlation analysis was therefore also conducted on the DeLone and McLean dimensions alone measuring the dimensions separately to establish internal consistency and relationships without this limitation or error factor. This is a means of demonstrating that the interdependent relationships set out in the DeLone and McLean model hold in the e-mentoring context without the limitation outlined above.

7.5.2.1.3.3 Scatterplot graphs

7.5.2.1.3.3.1 Relationships between Effectiveness and the DeLone and McLean dimensions

The aim of this data analysis was to test the validity of the questionnaire and scoring system as a basis for assessing and quantifying effectiveness for respondents. The question this analysis sought to answer was: “With reference to the dimensions set out in DeLone and McLean’s model of IS effectiveness, was effectiveness appropriately represented by the total effectiveness score?”

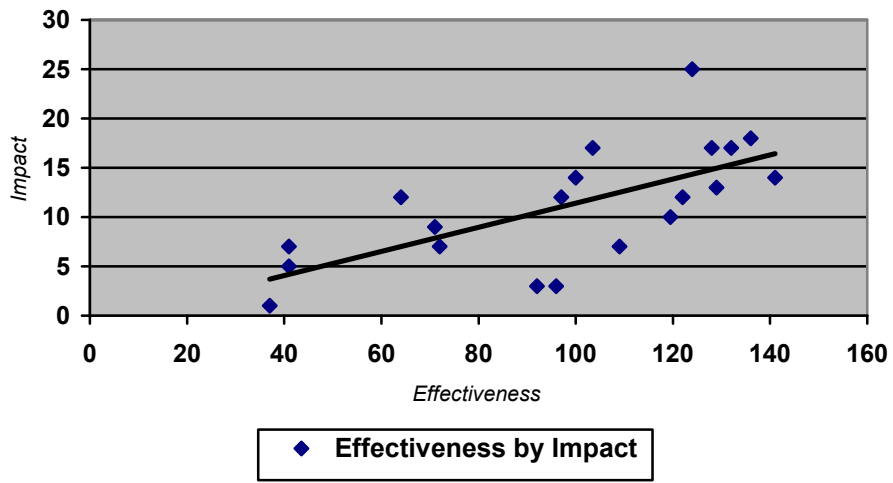
If there are strong relationships between effectiveness and each of the dimensions, it would be appropriate to draw the inference that the measurement of the construct of e-mentoring effectiveness was appropriately and validly measured with reference to the five DeLone and McLean dimensions.

If there are no relationships between the e-mentoring effectiveness variable and the DeLone and McLean dimensions as measures of association, it would be reasonable to infer that the construct was not appropriately measured with reference to the DeLone and McLean dimensions.

The correlation coefficient was calculated to provide a preliminary understanding of the strength and direction of each of the bivariate relationships.

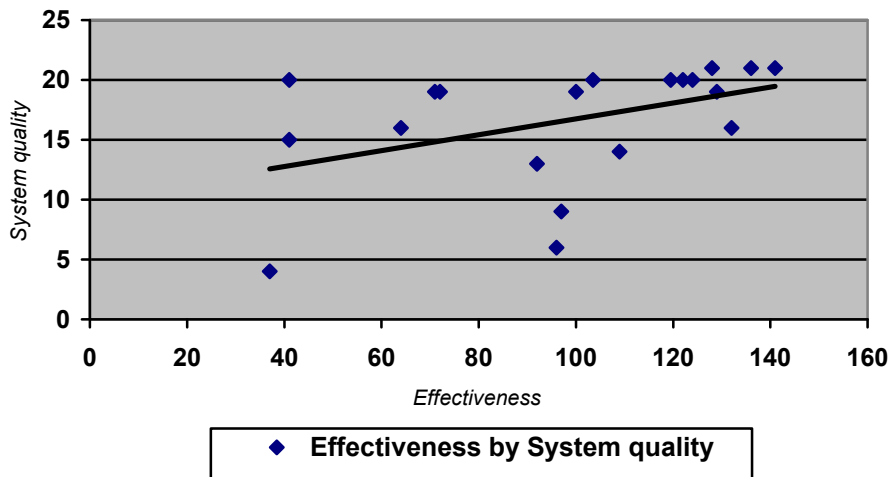
The following figures present scatterplot graphs in an effort to establish whether the data indicated that there was a positive linear relationship between effectiveness and the particular dimension being considered. If the relationships between effectiveness and each of the dimensions exist, this would be regarded as an indication of the relationship between the effectiveness score and scores for the particular dimensions, and confirm that the effectiveness score was an appropriate (though not necessarily adequate) measure of effectiveness.

Figure 6 – Effectiveness (total score) (X axis) by Impact (Y axis) with trendline



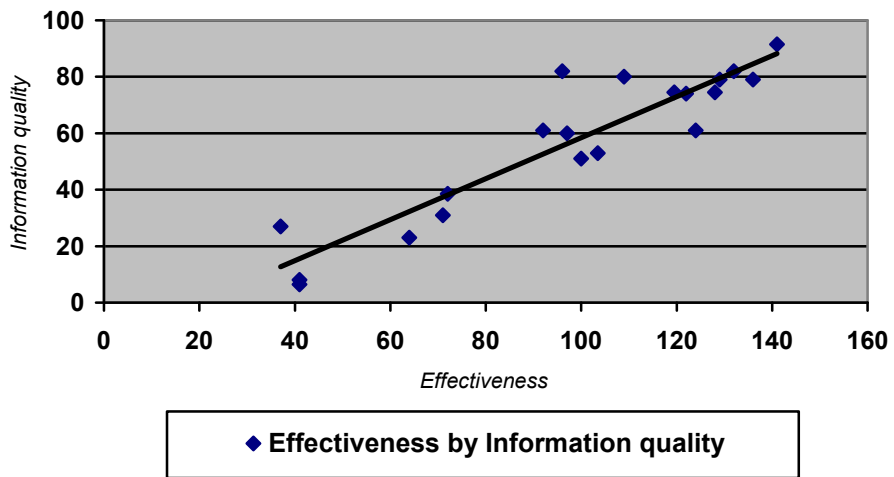
(Correlation coefficient 0.67)

Figure 7 – Effectiveness (total score) (X axis) by System quality (Y axis) with trendline



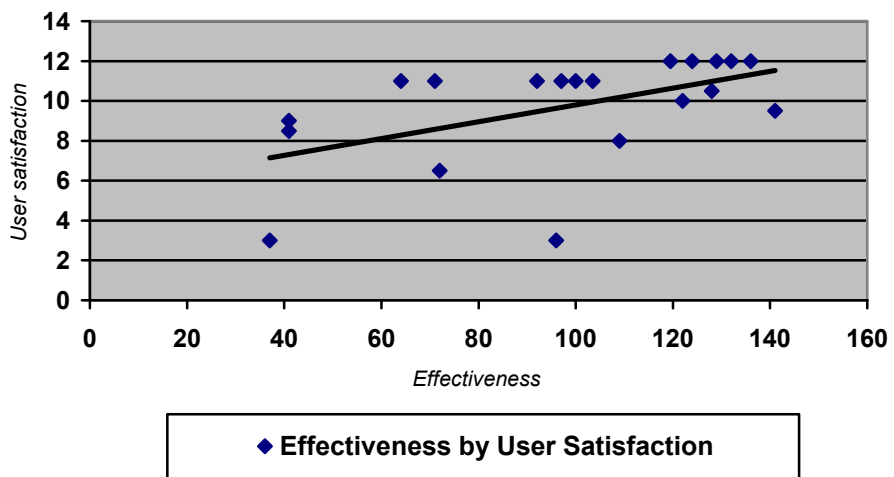
(Correlation coefficient 0.43)

Figure 8 – Effectiveness (total score) (X axis) by Information quality (Y axis) with trendline



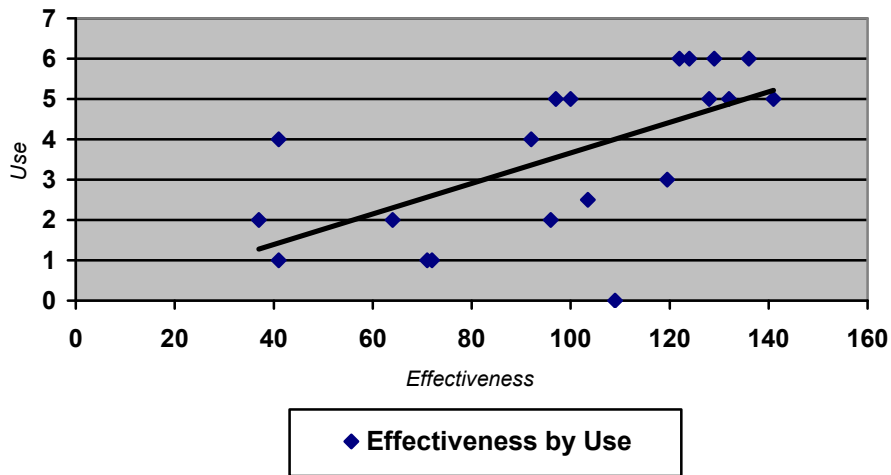
(Correlation coefficient 0.92)

Figure 9 – Effectiveness (total score) (X axis) by User Satisfaction (Y axis) with trendline



(Correlation coefficient 0.51)

Figure 10 – Effectiveness (total score) (X axis) by Use (Y axis) with trendline



(Correlation coefficient 0.63)

7.5.2.1.3.3.2 Relationships between Impact and other DeLone and McLean dimensions

The aim of this data analysis is to establish whether the data confirms the relationships which comprise the DeLone and McLean model of IS effectiveness. The correlation coefficient will be calculated to provide a mathematical indicator of the strength and direction of each of the bivariate relationships.

The question this analysis seeks to answer is: “Does the data suggest that the relationships between the DeLone and McLean dimensions hold in the e-mentoring context?”

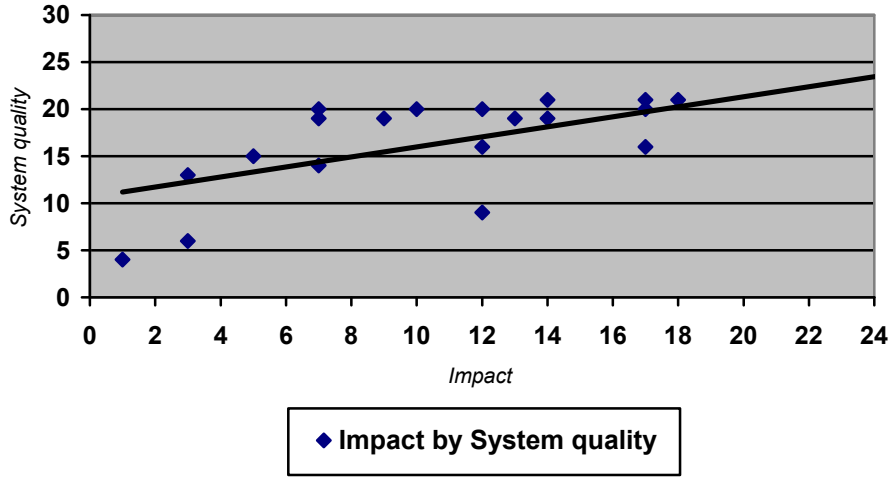
If there are strong relationships between impact and the other dimensions, it would be appropriate to draw the inference that the measurement of the construct of Impact is appropriately measured with reference to the other four DeLone and McLean dimensions in the e-mentoring context.

If there are no positive relationships between the Impact variable and the other DeLone and McLean dimensions, it would be reasonable to infer that the Impact construct is not appropriately measured with reference to the DeLone and McLean dimensions in the e-mentoring context.

The following figures present scatterplot matrices in an effort to establish whether the data indicates that there is a positive linear relationship between impact and the other DeLone and McLean dimensions. If the relationships between impact and each of the dimensions exist, this

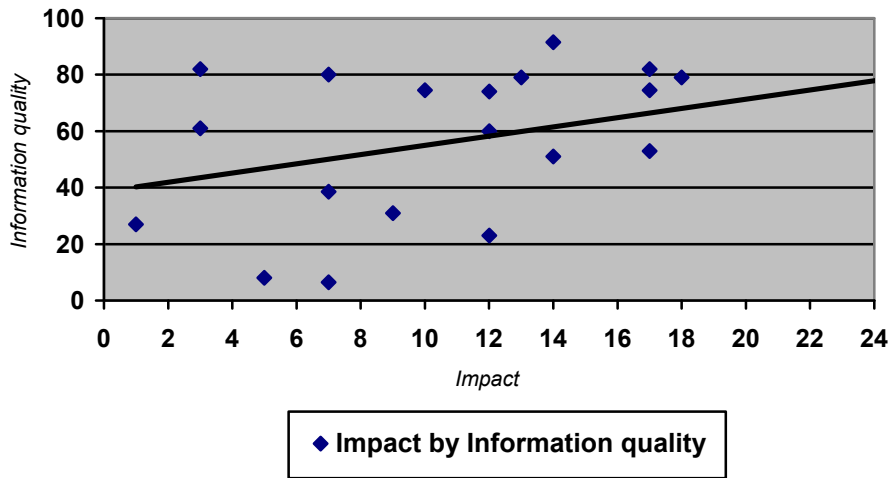
will be regarded as an indication that the DeLone and McLean dimensions are an appropriate (though not necessarily adequate) measure of Impact in this particular context.

Figure 11 – Impact (X axis) by System quality (Y axis) with trendline



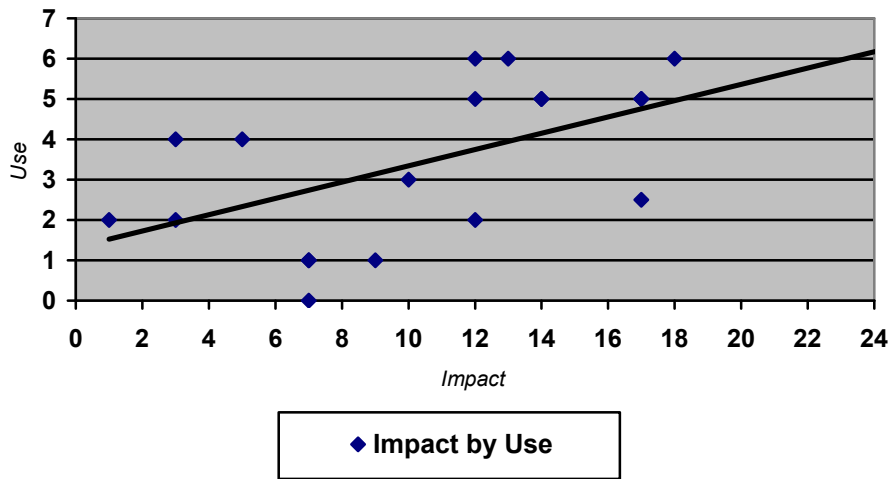
(Correlation coefficient 0.63)

Figure 12 – Impact (X axis) by Information quality (Y axis) with trendline



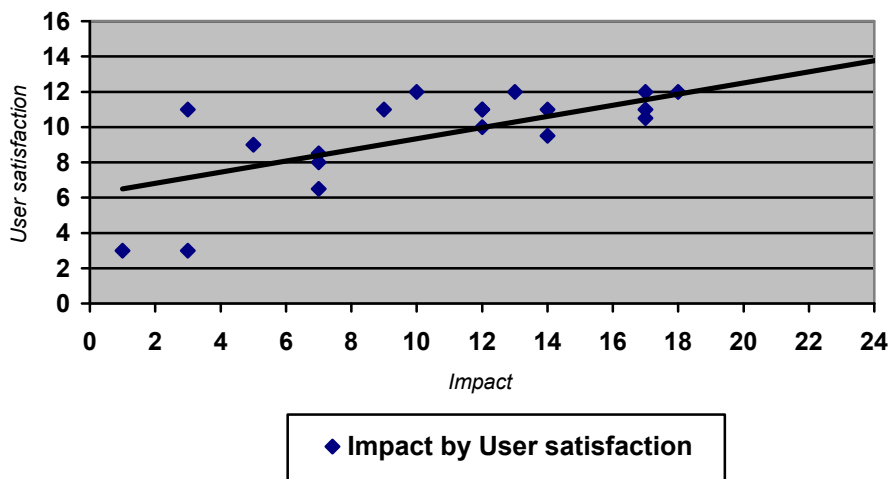
(Correlation coefficient 0.37)

Figure 13 – Impact (X axis) by Use (Y axis) with trendline



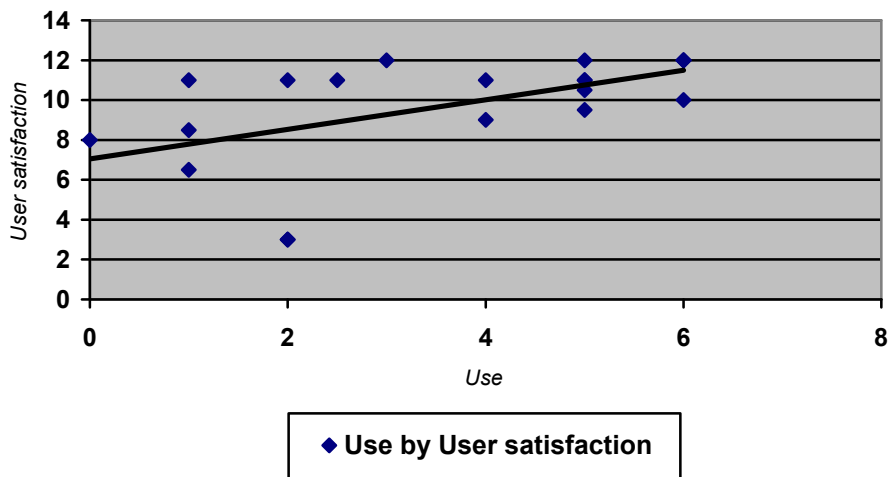
(Correlation coefficient 0.61)

Figure 14 – Impact (X axis) by User Satisfaction (Y axis) with trendline



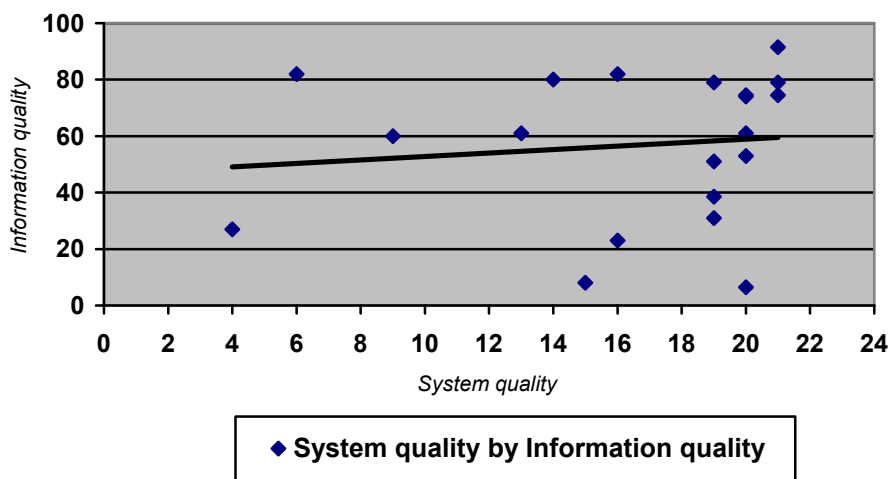
(Correlation coefficient 0.70)

Figure 15 – Use (X axis) by User Satisfaction (Y axis) with trendline



(Correlation coefficient 0.54)

Figure 16 – System quality (X axis) by Information quality (Y axis) with trendline



(Correlation coefficient 0.12)

7.5.2.1.3.3.3 Discussion of relationships represented in scatterplot graphs

The following section discusses the relationships represented in the scatterplot graphs. It should be noted that p-values are not reported due to sample size.

Effectiveness by Impact

The data and wide scatter around the line suggest a reasonable positive correlation between Effectiveness and Impact scores. The correlation coefficient is 0.67.

Effectiveness by System quality

The data and wide scatter suggest a weak positive correlation between Effectiveness and System quality scores. The correlation coefficient is 0.43.

Effectiveness by Information quality

The data and small scatter suggest a strong positive correlation between Effectiveness and Information quality scores. The correlation coefficient is 0.92.

Effectiveness by User satisfaction

The data and wide scatter suggest a weak positive correlation between Effectiveness and User Satisfaction scores. The correlation coefficient is 0.51.

Effectiveness by Use

The data indicate a reasonable positive correlation between Effectiveness and Use scores. The correlation coefficient is 0.63.

Impact by System quality

The data suggest a reasonable positive correlation between Impact and System quality. The correlation coefficient is 0.63.

Impact by Information quality

The data suggest a very weak positive correlation between Impact and Information quality. The correlation coefficient is 0.37.

Impact by Use

The data suggest a reasonable positive correlation between Impact and Use. The correlation coefficient is 0.61.

Impact by User satisfaction

The data suggest a reasonable positive correlation between Impact and User satisfaction. The correlation coefficient is 0.70.

Use by User satisfaction

The data suggest that there is a weak positive correlation between Use and User satisfaction. The correlation coefficient is 0.54.

System quality by Information quality

The data suggest that there is no positive correlation between System quality and Information quality. The correlation coefficient is almost zero at 0.12. This suggests that while effectiveness may be influenced by both system quality and information quality there is not necessarily a relationship between the two – that is, the existence of high scores in both areas is not required for a partnership to be measured as effective for the mentee.

This confirms the findings of the analysis of the judgements of mentees, mentors and program manager set out in the earlier section of this chapter (refer to section 7.5.2.1.2.1) which found that while System quality and Information quality were both related to effectiveness, there was not necessarily a relationship between the two, and positive outcomes in both areas were not necessary antecedents to an effective e-mentoring partnership.

7.5.2.1.3.4 Conclusions to inter-item correlation analysis

For the interpretation of the effectiveness score to be considered indicative of the level of effectiveness of a mentoring partnership for a mentee, relationships between effectiveness and the five DeLone and McLean dimensions would need to be confirmed by the data. The data represented in Figures 6-10 above suggest positive relationships between each of the five DeLone and McLean dimensions and effectiveness. There is some support for the claim that the selection of questions from the questionnaire classified and scored according to the DeLone and McLean dimensions, on the basis of the data analysis undertaken above, appropriately represents the construct of e-mentoring effectiveness.

For the DeLone and McLean dimensions to be considered a sound and appropriate means of evaluating effectiveness in quantitative terms, the relationships as set out in DeLone and McLean's model of IS effectiveness would need to be confirmed by the data. The relationships between the dimensions held in all but one case (System and Information quality) as set out in Figures 11-14.

The inter-item correlation analysis provides some support for the e-mentoring effectiveness scale to be considered a valid initial measure of the e-mentoring effectiveness construct in this setting.

7.5.2.2 Evidence of reliability

7.5.2.2.1 Cronbach alpha

The reliability of any new scale should be tested by establishing the Cronbach alpha. Bryman and Cramer (1990) suggest that Cronbach alpha of 0.8 or over indicates an acceptable level of reliability while Nunnally (1978) suggests that a figure of 0.6 may be satisfactory in the

case of an initial investigation (Bryman & Cramer and Nunnally in Cavana et al. 2000 p.320). Cronbach alpha of greater than 0.6 would confirm the reliability of the measuring instrument in relation to the proposition being considered. The alpha coefficient was calculated using the Cronbach alpha formula (Cavana et al. 2000) and with the aid of Statistical Package for the Social Sciences (SPSS) software.

While sample size (n=20) compromises the confidence in any claim made arising out of this data analysis, it is nonetheless an important step to be undertaken when considering the reliability of a measuring instrument which has not be used previously.

The Cronbach alpha coefficients will firstly be calculated to test the reliability of the items which together measure each dimension. The Cronbach alpha coefficients were computed as follows:

Table 63 – Correlation matrix – reliability for the five effectiveness dimensions

	Cronbach alpha
System quality	0.81
Information quality	0.55
User satisfaction	0.87 (see Note)
Use	0.67
Impact	0.77

Note: Because there was no variance in responses to Item 4 (all respondents indicated that they would recommend the program to another professional), this item was deleted from the computation of the Cronbach alpha. In any future redraft of the questionnaire, this item would either be amended so that responses better capture variance in the opinions of respondents or deleted as an item measuring User satisfaction.

This analysis demonstrates an alpha coefficient of over 0.6 in each of the dimensions with the exception of Information quality which had an alpha coefficient of 0.55. Based on Nunnally’s (1978) suggested minimum coefficient of 0.6 for an initial study, this provides some tentative evidence that measurement of at least four the five dimensions of the effectiveness construct is internally consistent and has a satisfactory level of reliability.

Cronbach alpha will also be calculated to investigate the following propositions:

Proposition 1

The dimensions measuring effectiveness as operationalised by the DeLone and McLean dimensions of System quality, Information quality, User satisfaction, Use and Impact reliably measure the construct; and

Proposition 2

The Effectiveness score reliably represents Effectiveness as comprised by the DeLone and McLean dimensions.

Table 64 – Correlation matrix – summary of five inter-item correlations

	System quality	Information quality	User satisfaction	Use	Impact
System quality	1				
Information quality	0.22	1			
User satisfaction	0.67	0.18	1		
Use	0.29	0.35	0.54	1	
Impact	0.63	0.33	0.70	0.61	1

Mean inter-item correlation = 0.453

Cronbach’s alpha = 0.81

Table 65 – Correlation matrix – summary of six inter-item correlations

	System quality	Information quality	User satisfaction	Use	Impact	Effectiveness
System quality	1					
Information quality	0.22	1				
User satisfaction	0.67	0.18	1			
Use	0.29	0.35	0.54	1		
Impact	0.63	0.33	0.70	0.61	1	
Effectiveness	0.43	0.81	0.51	0.63	0.67	1

Mean inter-item correlation = 0.51

Cronbach’s alpha = 0.86

Proposition 1

Cronbach’s alpha was 0.81 in relation to Proposition 1. This result indicates that there is some support for the items selected to measure Impact, as operationalised with reference to the dimensions of DeLone and McLean’s model of IS success, to be considered as measuring the e-mentoring Impact construct with a degree of reliability. Proposition 1 is therefore supported.

Proposition 2

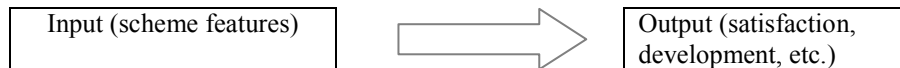
Cronbach’s alpha was 0.86 in relation to Proposition 2. This result indicates that there is some support for the items selected to measure Effectiveness, as operationalised with reference to the dimensions of DeLone and McLean’s model of IS success, to be considered as measuring the E-mentoring Effectiveness construct with a degree of reliability. Proposition 2 is therefore also supported.

7.5.2.2.2 Use of survey questionnaire in international context

The survey instrument was applied in a similar context in the United Kingdom on behalf of Business Link Surrey, and the evaluation reported on a selection of the results (Megginson et al. 2003).

7.5.2.2.2.1 Definition of the effectiveness construct

The Megginson et al. (2003) taxonomy for evaluating effectiveness can be represented as follows:



The Megginson et al. (2003) taxonomy set out above does not include the temporal or process characteristics of the DeLone and McLean model (refer to section 1.13.2.2) nor the interdependent nature of the dimensions of effectiveness (refer to section 1.13.1). The rationale behind the taxonomy adopted by Megginson et al. as the effectiveness construct, and how input and output were operationalised, is not explicitly set out in the study. The variables referred to development, whether the mentee would recommend the program to others and satisfaction. Given the correlational analysis presented, program benefit, willingness to participate in a similar program again, professional development and contribution to professional identity were also regarded as indicators of effectiveness.

7.5.2.2.2.2 Formative emphasis

The emphasis in this study is on identifying trends in respondents' perceptions of the scheme features and their links to "key output variables". The approach is founded on establishing antecedents to outcomes or "input" to "output". The focus on identifying factors which influence effectiveness, on "lessons for the scheme to be learned" and the identification of critical success factors suggest that a formative approach underpins this study. In this way, the study attempts to pinpoint factors which influence effectiveness with a view to informing program development and maximising program effectiveness. This is in contrast to the focus of the quantitative and qualitative phases of this thesis which, while similarly attempting to understand the relationships between the dimensions of the effectiveness construct, is ultimately a summative analysis of effectiveness.

7.5.2.2.2.3 Comparison with data arising from use of quantitative instrument in international setting

As a means of testing the reliability of the measurement instrument, the correlations proposed in the international context will be compared with the correlations arising from the data from the

Australian study outlined in this thesis. If correlations of similar strength and direction are found, this would indicate that the measuring instrument has some reliability and is measuring the same way across the different settings. If correlations of similar strength and direction are not found in the Australian context, this lack of reliability must be accounted for as a threat to the reliability of the measurement instrument.

While not including correlation coefficients, the Megginson et al. (2003) study reported the following correlations significant to the .01 level (p.33):

1. Program benefit and Contact frequency
2. Program benefit and Supplementing email contact with other modes of communication
3. Program benefit and Self-managed learning
4. Willingness to participate in a similar program again and Addressing the skills nominated in the registration form
5. Professional development and Value of online tutorial
6. Professional development and Mentors providing mentees with options
7. Contribution to professional identity and Integration with day to day business activities

The results of comparative data analysis for the same bivariate relationships in the Australian context is set out in Table 66 as follows:

Table 66 – Summary of correlation coefficients and strength of relationship between nominated variables in Australian context

	N (number of responses in data set)	Correlation coefficient for variables in Australian context	Strength of relationship in Australian context	Strength of relationship in UK context	Relationship confirmed?
Program benefit and contact frequency	20	0.28	Weak	Strong	No
Program benefit and Supplementing email contact with other modes of communication	20	0.07	No relationship	Strong	No
Program benefit and Self-managed learning	12 (Question not included in 2004, 2005 and 2006)	0.63	Reasonably strong	Strong	Possibly
Willingness to participate in a similar program again and Addressing the skills nominated in registration form	19 (Question not answered by one participant)	0.34	Weak	Strong	No
Professional development and Value of online tutorial	20	0.49	Weak	Strong	No
Professional development and Mentors providing mentees with options	20	0.22	Weak	Strong	No
Contribution to professional identity and	18 (Question not answered)	0.50	Reasonable	Strong	Possibly

Integration with day-to-day business activities	by two respondents)				
---	---------------------	--	--	--	--

The disparities in correlation should therefore be accounted for and will be discussed in sections 7.5.2.2.2.4 and 7.5.2.2.2.5 and summarised in 7.5.2.2.2.6.

7.5.2.2.2.4 Testing

The UK research population was involved in pre and post-testing while the Australian group undertook post testing only. Because of this, it is possible that pre-tests sensitised the respondents to the post test (Cavana et al. 2001). This is a variable which may also impact on the validity of the comparison between UK and Australian data.

7.5.2.2.2.5 Program implementation

The UK and Australian programs differed in implementation in relation to the frequency and nature of contact with the program host. While the Australian program provided fortnightly emails to mentees and mentors, the report on the UK program indicates that “[o]ngoing communication was kept at a minimum throughout the program” (Megginson et al. 2003 p.15). While based on the Australian program, the absence of regular contact with the program host as part of the program structure is a critical difference between the programs. It compromises the validity of the comparison between the programs and highlights the need to adequately define the construct of structured e-mentoring.

7.5.2.2.2.6 Summary of evidence of reliability in international setting

The strong correlations found in the UK study are confirmed in the Australian context in only two instances, that between Program benefit and Self-managed learning, and between Contribution to professional identity and Integration with day-to-day business activities. The other relationships were either weak or non-existent. It is possible to conclude that the different contexts, different testing procedures, differences in program implementation and definition of the construct of structured e-mentoring, or the unreliability of the measuring instrument, are all potential sources of these disparities. There is therefore evidence of the potential unreliability of the measurement instrument, much scope for further comparative research and, most critically, grounds for caution in relying on the measurement instrument as the sole means of measuring effectiveness.

7.5.2.2.3 Does the measurement instrument behave as expected and do the results confirm previous research findings?

Brualdi (1999) suggests that a means of establishing validity of a measurement instrument is to determine whether the test behaves as one would expect a measure of the construct to behave, and whether or not the internal structure of the test is consistent with what is known about the internal structure of the construct. So did the data arising out of the survey confirm or disconfirm what is known about e-mentoring effectiveness?

The importance of program structure to the effectiveness of structured e-mentoring programs is widely acknowledged in the literature (Single & Single 2005). The data analysis confirmed a strong relationship between effectiveness and the program structure or Information quality. In particular, satisfaction with program features was correlated with program effectiveness for mentees (correlation coefficient of 0.90).

The mentoring literature however indicates that in the small business context there is limited value in providing generic program content because of the heterogeneity which characterises the small business population (Atterton 2002, Tolentino 1998). Megginson et al. also point to this in their analysis of the UK e-mentoring program citing Garvey (1995 cited in Megginson et al 2003) who emphasise the importance of voluntarism and uniqueness in mentoring relationships and cautioned against “being too prescriptive and standard-driven about ... advice given” (Megginson et al. 2003 p.35). The intention of the Australian program was to provide a basic structure which the participants would adapt to their own purposes rather than to be prescriptive or standard-driven. It would therefore be reasonable to expect a correlation between effectiveness and those who indicated that they adapted the generic content to their own purposes. The study however found only a low positive correlation between effectiveness and participants setting their own program goals with a correlation coefficient of 0.25. When considered in relation to the strong link between program structure and program effectiveness discussed previously, it may be that respondents found provision of structure and informal adaptation of the generic content more important than setting their own program goals at the outset of the program in this form of e-mentoring. The data would suggest that these questions need to be considered together for the measuring instrument to return the results which the informing literature would lead us to expect, and that there is further scope for research into how participants in e-mentoring programs adapt generic content to their own purposes.

The relationship between use and effectiveness is widely acknowledged in the e-mentoring literature. Single and Single (2005) refer to a study in which ready access to technological hardware in the form of provision of portable keyboards rather than limited hardware

dramatically improved the quality of e-mentoring relationships and the skills of proteges because it provided for more regular contact (Friedman 2004 in Single & Single 2005). This study is supported by a range of other studies which discuss the link between contact frequency and improved outcomes. Single and Single refer to a range of studies which confirm that involvement operationalised as frequency and duration of e-mentoring interactions was a variable related to positive e-mentoring outcomes (Single & Single 2005) while Bierema and Merriam (2002) suggest that “successful mentoring involves frequent and regular interaction” (p.214). Single and Muller (2001, p.118) problematise the relationship between effectiveness and use suggesting that causal direction has not been clearly established – that is, do effective partnerships lead to greater email exchange frequency?, or does a higher rate of email exchange bring about an effective partnership? The data, while not providing any clarification on this particular issue, confirmed a positive correlation between use and effectiveness and in this way demonstrates the relationship between the two dimensions which would be expected based on previous studies.

Studies of face-to-face mentoring have shown that failure to meet is one of the key reasons for mentoring partners not developing or maintaining their partnership (Noe 1988, Dickey 1997). The measurement instrument showed that all but one participant indicated the email-based structure of the program facilitated their participation. In this way the measuring instrument behaved as expected and in line with previous research.

The test or application of the measurement instrument conforms with previous research which found information and psychosocial benefits as outcomes of the mentoring process. The test confirmed that most respondents were referred to further useful information and resources (15 out of 20 respondents), and that they experienced psychosocial support and reinforcement in the form of personal and/or professional development (15 out of 20 respondents). In this way then, the measurement instrument performs in the way it could be expected to perform in the context of previous research.

The literature indicates that the value of impartiality is important as a benefit of structured e-mentoring. Rather than being mentored within an organisation which may lead to a reluctance to discuss personal or professional weaknesses, the literature indicates that participants in e-mentoring programs value the opportunity to discuss issues with a neutral party outside their existing network. The data arising from the measurement instrument confirmed that most participants valued the opportunity to bounce ideas off a neutral third party and discussed issues which they would not normally do within their existing network.

Table 67 – Mentor as sounding board (n=20)

	Yes	No	Total
Opportunity to bounce ideas off a neutral third party?	18	2	20
Opportunity to discuss issues which they would not normally discuss within existing network	15	5	20

The data arising out of the test therefore confirmed the results which could be expected in line with previous research.

This discussion demonstrates that there is some support for the claim that the test can be considered to behave, in Brualdi’s (1999) terms, as one would expect a measure of the construct to behave, and that the internal structure of the assessment is consistent with what is known about the internal structure of the construct.

7.5.2.3 Limitations or sources of invalidity

7.5.2.3.1 Construct underrepresentation

Brualdi (1999) defines construct underrepresentation as when “tasks which are measured in the assessment fail to include important dimensions or facets of the construct” (p.3). Such a failing in a measurement instrument would mean that an effectiveness score arising from the test may not reveal the extent of effectiveness indicated by that score or that a dimension is inappropriately weighted within the total score. The data arising from the quantitative analysis, while indicating a correlation between effectiveness and the quality of the mentee/mentor relationship, failed to measure this relationship with any degree of depth or complexity. It is proposed that this is however not only of this particular measurement instrument but of any attempt to measure this dimension of the structured e-mentoring experience in quantitative terms. The literature indicates that the nature and quality of this relationship is central to outcomes and effectiveness (Devins & Gold 2000 et al). The numerical data arising from the use of this measurement instrument confirm that the inferences which can be drawn from the data lack complexity and do not provide a basis for exploring linkages between effectiveness and the dimensions with any depth, sophistication or meaning. The data in this context are useful in confirming a relationship between effectiveness and the nature and quality of the relationship and a possible preliminary basis for classifying the extent of effectiveness; however obtaining richer and more useful data requires a different methodological approach using qualitative data to adequately represent the complexity of this dimension of the construct.

7.5.2.3.2 Aim of the survey instrument – to measure or rank

Another limitation of the quantitative analysis arises out of the inclusion of both ordinal and nominal items included in the items selected to operationalise the dimensions. While ordinal

items can be counted and ordered, they cannot be measured as such. To therefore describe the instrument as a measure of e-mentoring effectiveness is in part misleading.

7.5.2.3.3 Contextual analysis

Because of the difficulty of controlling for contextual variables, the validity of the findings are potentially compromised by the multitude of contextual variables which potentially impact on the effectiveness of the program. While investigating all of the Phase 2 variables (refer Table 48) set out in the proposed framework is not feasible, the available data provides a basis for exploring whether or not a selection of variables impact on effectiveness. The two factors selected from the proposed framework for preliminary investigation are the “Internal mentee factor” of gender and the “External mentee business factor” of geographical location. These were the two contextual factors established as being representative in the sample.

7.5.2.3.3.1 Gender

A comparison of the mean effectiveness scores for males and females will be undertaken to determine whether or not there is evidence of differential outcomes by gender.

Table 68 – Mean effectiveness scores by gender

Gender	Mean score
Female	91.8
Male	93.9

The mean score for females was 91.8 (n= 7) and the mean score for males was 93.9 (n=13). On the basis of this data, and in view of the small numbers involved, there are no grounds for making the claim that gender impacts on effectiveness.

7.5.2.3.3.2 Geographical location

A comparison of the mean effectiveness scores for city and regional/rural location will determine whether or not there is evidence of differential outcomes by geographical location.

Table 69 – Mean effectiveness scores by gender

Geographical location	Mean score
City	91.08
Rural/regional	107.17

The mean score for respondents in city locations was 91.08 (n= 13) and the mean score for those located in rural or regional areas was 107.17 (n=6). On the basis of this data, and in view of the small numbers involved, there are no grounds for making the claim that regional location impacts on effectiveness.

Conclusion

The immaturity of structured e-mentoring effectiveness as a field of evaluation research is evident with extensive scope for further investigation of the myriad of contextual factors which may impact effectiveness. Such analysis will form part of the theory-building stage of the cumulative research process in this emerging area, and establishing, in Carlile and Christensen's (2005) terms, how and for whom any generalised statements of effectiveness apply.

7.6 Extrapolating to program effectiveness

The process of extrapolating from effectiveness for individual mentees to program effectiveness must be defined by the evaluator in conjunction with the program host. As an example of how quantitative measures of effectiveness can be extrapolated to program effectiveness, the effectiveness pentagon set out in Figure 17 visually represents program effectiveness by plotting the mean mentee scores for each of the dimensions against total possible scores (total possible scores set out in brackets following axis label).

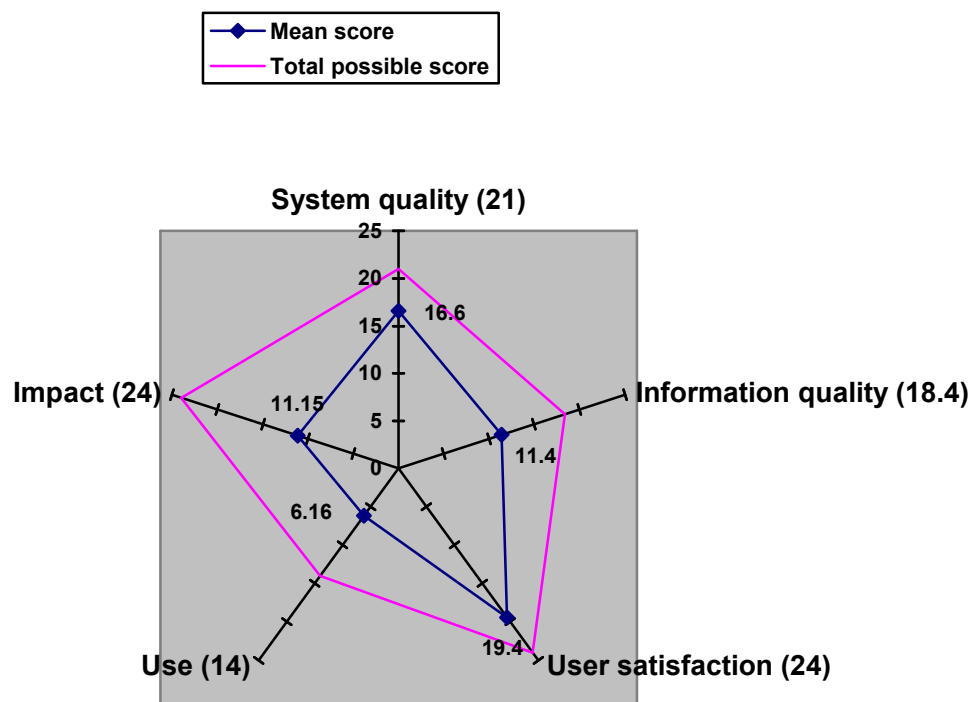


Figure 17 - Effectiveness pentagon

The pentagon is useful in that it represents effectiveness holistically and the dimensions are presented relative to the other dimensions. Using the quantitative data arising out of this survey was problematic because of the varying scales of the data which arose from the secondary data was utilised. To facilitate representation in this format, a subsequent iteration of the scoring scales would amend the scale for Information quality. To facilitate representation for the Information quality scores in this instance, the Information quality scores were reduced by a

factor of 5, and the Use and User satisfaction scores were doubled to better align the scales for those dimension with the other dimensions. As a secondary step, the data around each of the dimensions could be expanded to present individual scores. The program host would make decisions about the scores for each of the dimensions they would judge the program to be effective, and this could be imposed on the Effectiveness pentagon as a means of quantitatively evaluating program effectiveness.

Detailing how effectiveness for individual mentees and program effectiveness are related, suggesting further how practitioners might extrapolate from individualised data to making generalisations around program effectiveness, and in turn, how program effectiveness relates to policy effectiveness, is beyond the scope of this study. The construction of Figure 5 simply demonstrates one way that quantitative data could be used for making evidence-based judgements around effectiveness.

7.7 Conclusions

The immaturity of this field is reflected in the limitations of the measurement instrument. The limited basis upon which the validity and reliability of the instrument have been assessed is acknowledged. There is irrefutably a need for continuing development, refinement and testing and re-testing of quantitative measures of effectiveness in different contexts alongside the use of methodologies which will more adequately represent the construct of effectiveness, chiefly in relation to the quality of the mentee/mentor relationship.

As detailed in the discussion relating to Table 58, the setting of these thresholds involved the qualitative judgement of the researcher as informed by practice and experience in the field. The frequency distribution of effectiveness scores set out in Figure 5, the concordance between scores and judgements of the mentee, mentor and program host set out in Table 62, and the tests relating to validity and reliability set out in this chapter in sections 7.5.2.1 and 7.4.2.2 confirm the credibility of the thresholds supporting the effectiveness classification scheme.

The purpose to which the data arising from a measurement instrument is put is widely acknowledged as underpinning the validity of its use (Brualdi 1999). While it is recognised that caution should be exercised in relying upon the quantitative measures of effectiveness used in this study, it is proposed that there is sufficient empirical data and argument, as outlined above, to suggest that there is some validity to the inferences drawn about effectiveness in this context for the purposes of constructing the sample for the qualitative analysis to be presented in Chapter 8. It is proposed that the measurement instrument forms a basis on which the mentees'

experiences can be classified according to levels of effectiveness with a degree of validity and credibility to be used for this purpose.

In the analysis to be undertaken in Chapter 8, effective and very effective will be compared with the ineffective and partly effective mentee program experience. Total effectiveness scores will therefore be used to classify participants into four groups – Ineffective, Partly effective, Effective and Very effective. This will form the basis for the qualitative comparative analysis of Very effective and Effective with Partly effective and Ineffective partnerships to be undertaken in the next Chapter.

The participants will, on the basis of the quantitative analysis set out above, be classified as follows:

Table 70 – Summary of classification of degree of effectiveness

Score range	Description	Participant number	Frequency
0-45	Ineffective	9, 17, 12	3
46-97	Partly effective	5, 8, 19, 1, 15, 4	6
98-129	Effective	7, 10, 6, 14, 11, 2, 13, 18	8
130-156	Very effective	3, 16, 20	3

The inter-item correlation analysis of this evaluation research confirms the existence of general patterns between the dimensions of the DeLone and McLean model and confirmed the DeLone and McLean dimensions as antecedents to effectiveness. While there is scope for further research into how the proposed taxonomy fails to accommodate or underrepresent the data, the data analysis confirmed as a minimum, modest positive relationships between each of the DeLone and McLean dimensions with the exception of that between System and Information quality. The comparison of correlations between factors influencing effectiveness in the UK program found only tentative confirmation of factors influencing effectiveness in the Australian program. The differences may be attributed to the different program context, differences in program implementation and testing, or of more concern, the unreliability of the measurement instrument. The reliability and validity testing showed grounds for caution in using the survey instrument as the sole means of measuring effectiveness

While the implementation of the proposed framework did not attempt to control for contextual variables, the framework nonetheless provided a basis for considering, selecting and at least acknowledging the impact of the contextual factors of gender and regional location on the validity of the inferences set out in the Findings section, and provided a basis for evaluation researchers to consider, as theory-building is planned and undertaken. Overall, this preliminary process of quantitative analysis provides some empirical support for claiming that effectiveness

scores can be regarded as preliminary though not sufficient measures of effectiveness, and that the DeLone and McLean dimensions were validated as appropriate quantitative analysis categories in the structured e-mentoring context.

An in-depth analysis of factors influencing effectiveness will be undertaken in Chapter 8 with reference to qualitative data.