Aftershocks of an electronic publishing experiment: a review of the EPICentre project[1]

* Mr Andrew Treloar - Deakin University (andrew.treloar@deakin.edu.au)
<http://www.deakin.edu.au/~aet>
Dr Jenny Cameron - CSIRO Publishing
Mr Tom Denison - Whiteroom Publishing (tdenison@vicnet.net.au)
Dr Bob Jansen - CSIRO DIT (bob.jansen@dit.csiro.au)
<http://mac145.syd.dit.csiro.au/>
Mr Paul Mercieca - INFORMIT (paulm@informit.com.au)
Mr Paul Reekie - CSIRO Publishing (paul.reekie@publish.csiro.au)
Professor Don Schauder - Monash University (dschauder@vicnet.net.au)
Mr Patrick Wilken - Monash University (pwilken@peg.apc.org)
<http://www.cs.monash.edu.au/~patrickw/>

* Please direct questions about this paper to Mr Treloar.

Abstract

This article will outline the project recently undertaken by the Electronic Publishing Innovations Centre (EPICentre) as the result of project funding from the AVCC through the Standing Committee on Information Resources, Library infrastructure projects, Program 3 - Electronic Publishing of Full Text Materials. This Round 1 project worked with two existing journals: *Psyche - an interdisciplinary journal of research on consciousness* (an electronic journal) and the *Australian Journal of Chemistry* (a print journal). Both of these journals were converted into a range of electronic forms (HTML, SGML, PDF) for delivery via network and CD-ROM. The lessons learnt from the journal conversion process and its implications for sustainable electronic publishing will be summarized. The project also surveyed users of about their attitudes to and uses of the new technologies.The results of the survey phase will be presented. .

[1]**A note on presentation**: This article has been designed for parallel delivery on paper and online. The compromise I have developed for citing the enlarged sphere of possible objects that this makes possible is as follows. Links to Internet URL's will be indicated in the text by a footnote marker, with the URL itself as the body of the footnote. This unfortunately means that the scope of the link source cannot accurately be indicated. References to published works will be made using the Harvard system with the works themselves listed in the bibliography at the end of the paper. Those works which are known to be available online will have their URLs

1. Context

Researchers in the field of scientific communication have been envisaging the advent of electronic journals since at least the work of Senders (1976). Early electronic journal (ejournal) projects in the 1980s started to put some of these ideas into practice, but were limited in what they could achieve by the available technological infrastructure (Shackel, 1991).

In the early 1990s a number of researchers began to argue strongly that the existing system of print publication for journals was flawed in a number of ways:

- print journals are slow to appear;
- they cannot be directly searched, leading to a large market for secondary abstracting and indexing services;
- they are limited to information that can be represented statically in print;
- their mechanisms for hyper-linking are clumsy at best;
- they are costly to produce, distribute and store. (Harnad, 1991 and Odlyzko, 1996).

Their suggested response was a shift toward ejournals. The increase in the popularity and availability of the Internet coupled with the explosive growth in the World Wide Web from 1993 onwards removed many of the technology infrastructure barriers that had hampered the early ejournal projects. From 1990 onwards an increasing number of ejournals appeared, both refereed and non-refereed. At the same time, the amount of literature discussing electronic publishing and ejournals also expanded dramatically (Bailey, 1996). From 1994 onwards a number of pilot projects took place in the UK, the US and Australia. This paper describes one of these projects.

2. EPICentre

The Electronic Publishing Innovations Centre (EPICentre) formed in late 1994 as the result of a successful bid for project funding from the Australian Vice-Chancellor's Committee through the Standing Committee on Information Resources, Library infrastructure projects, Program 3 - Electronic Publishing of Full Text Materials. This program is managed by the Standing Committee on Information Resources, Working Group on Electronic Publishing. A number of projects were funded in the first round of funding, but EPICentre received the largest grant of A$100,000.

EPICentre was structured from the outset to draw on the expertise and strengths of a number of organisations. The Division of Information Technology (DIT) within the Commonwealth Scientific Industrial and Research Organisation (CSIRO) provided expertise on electronic publishing technologies, document models, and project management. CSIRO Publishing were publishers of a number of print journals in the Science and Technology Materials (STM) area and were actively investigating electronic publishing. Monash University were publishing electronically and wished to enhance this. The Royal Melbourne Institute of Technology University (RMIT) had significant expertise in CD-ROM based publishing and information organisation.

The original proposal was to undertake a number of sub-projects to develop an understanding of various aspects of electronic publishing. The most significant of these were to:

- convert or enhance two journals;
- survey readers of the journals;
- investigate various document models.

3. The journals

The two journals selected were *Psyche - an Interdisciplinary Journal of Research on Consciousness*

(Psyche) and *The Australian Journal of Chemistry* (AJC). These journals were selected with great care to provide the greatest possible contrast.

The AJC is a traditional print journal from the physical sciences moving into parallel electronic delivery. In terms of its formattingrequirements it presents significant challenges for online delivery. In common with much STM (Science, Technology and Medicine) material it contains in its articles formatted text, diagrams, photographs, mathematical equations and tables.

Psyche is a science publication that started life in electronic form and is moving into parallel print delivery through MIT Press. Its formatting requirements are much less demanding, and until recently were handled using ASCII text only.

Articles from both of these journals were converted into a range of electronic forms. These forms ended up including Hypertext Markup Language (HTML[2]), Standard Generalised Markup Language (SGML[3]), and Adobe's Portable Document Format (PDF), usually referred to as Acrobat[4], although this is technically the name of the Adobe family of products for manipulating PDF.

Psyche now has a fully functional Web site[5] which is used to deliver the journal articles. ASCII versions of the articles are also available from an FTP site. The Psyche web site also provides links to related:

- conferences by subject and date;
- courses;
- discussion groups;
- electronic journals;
- print journals;
- bibliographies, papers, preprint services, and publishers
- useful web & gopher sites.

The AJC is continuing to appear primarily in print only, but is using its Web site[6] to provide:

- supplementary information about the journal (Notice to Authors, Journal Style, Reprint Information, Editorial Committee);
- links to some external sites including the Australian Chemistry Network, the Cambridge Crystallography Data Centre and the EPICentre project;
- contents pages of recent issues following (or even preceding) release of the paper publication.;
- some selected articles published in the journal;
- information about special issues.

4. Reader survey

The survey was structured to determine:

- broad demographics for the readership of each journal;
- readers' access to technology;
- their familiarity with different forms of electronic publishing;
- their perceived advantages for electronic publishing;
- their perceived disadvantages for electronic publishing.

The survey instrument was distributed to Psyche readers via email. The instrument was also

---

[2] http://www.w3.org/pub/WWW/MarkUp/
[3] http://www.w3.org/pub/WWW/MarkUp/SGML/
[4] http://www.adobe.com/acrobat/main.html
[5] http://psyche.cs.monash.edu.au/

posted on the EPICentre Web site so that people could download the text, complete it and send it in. Use of the Web forms facility to gather responses was not used due to lack of time to set up the data-handling on the server. The survey instrument was distributed to AJC readers as an inclusion in the print journal. The email version of the survey instrument is included in this paper as Appendix 1.

For Psyche, 336 responses were received from a mailing list of 2800 readers (332 received by email, three printed out and mailed in, one faxed in). AJC has a subscription list of around 800. From this readership, a total of three (3) responses were received. As even the most sophisticated statistical techniques are unable to do much with a data set having these characteristics, the following discussion is based on the Psyche responses only.

4.1 Basic demographics

Because the overwhelming majority of Psyche responses were in the form of email it was possible to do some analysis of the geographical distribution of the respondents. Of the four print responses, it was possible to assign an equivalent domain to three of them. Table 1 shows an initial division into US domains, and then all other countries. The .us domain was only associated with k12.xx.us email addresses in this survey. Outside the US, only countries with more than 5 respondents (United Kingdom, Australia, Canada, The Netherlands, Germany and France) have been listed separately. The 52 aggregated country responses included respondents from Argentina, Austria, Belgium, Brazil, Chile, Croatia, Denmark, Estonia, Fiji, Finland, Hong Kong, Hungary, Ireland, Israel, Italy, Japan, New Zealand, Norway, Portugal, Singapore, the Slovak Republic, South Africa, Spain, Sweden, Switzerland, and Taiwan.

Table 1. Survey responses by Country (N=336)

US Domains	*Number*	*Percent*
.com (Commercial)	46	13.7
.edu (Education)	125	37.2
.gov (Government)	4	1.2
.net (Unspecified)	4	1.2
.org (Organisations)	8	2.4
.us (United States)	3	0.9
Countries with more than 5 replies		
.uk (United Kingdom)	36	10.7
.au (Australia)	22	6.5
.ca (Canada)	17	5.1
.nl (The Netherlands)	8	2.4
.de (Germany)	5	1.5
.fr (France)	5	1.5
Other Categories		
Other countries (aggregated)	52	15.5
Unknown	2	0.6
Total:	336	100.0

4.2 Technology access

Respondents were asked to indicate which computer/communications technology they had access to, either at home or at work. The intention was to determine what facilities an

ejournal could assume from its readership. There is little point in adding sound or full-colour video to en ejournal if a majority of readers cannot benefit from this. Table 2 shows the responses to this set of questions, ordered by the sequence of questions in the survey instrument.

Table 2. Access to Technology (N=336)

Response	*Number*	*Percent*
Personal computer	327	97.3
CD-ROM drive	243	72.3
Sound output	238	70.8
Colour screen	307	91.4
Direct network connection	258	76.8
Modem connection	226	67.3
Other	33	9.8

It is clear from table 2 that over two-thirds of the respondents have access to multimedia facilities on their computers. This is a large enough proportion to consider adding features to *Psyche* that take advantage of this technology. This might include supplementary material on CD-ROM, access to interviews (perhaps using a streaming format like Real-Audio), and video simulations to support article text.

4.3 Use of electronic publishing

The possible activities that respondents might take to work with various forms of electronic publishing are:

- subscribing to various electronic publishing forums (such as mailing lists, network news groups, bulletin board systems and the like)
- using FTP to access materials stored on remote ftp servers
- using Gopher to access materials stored on remote gopher servers
- using the World Wide Web to access materials
- accessing materials stored on CD-ROMs
- viewing electronic journals (however retrieved)
- publishing electronically themselves.

Table 3 shows respondents' interactions with these epublishing activities.

Table 3. Use of electronic publishing technologies (N=336)

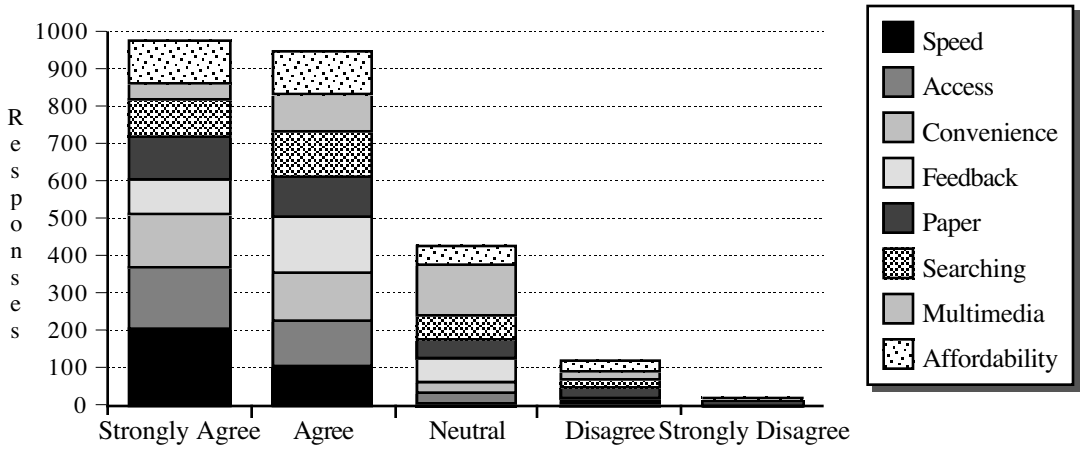
Activity	*Frequentl y*	*Regularl y*	*Occasion ally*	*Rarely*	*Never*	*No answer*	*Wrong*	*Total*
Subscribe to forums	43.8%	20.2%	21.1%	7.4%	2.1%	1.5%	3.9%	100%
Use FTP	24.4%	19.0%	27.4%	11.6%	5.7%	8.9%	3.0%	100%
Use Gopher	7.4%	8.6%	32.7%	31.0%	12.2%	6.2%	1.8%	100%
Use WWW	45.5%	21.1%	18.2%	4.8%	3.9%	1.5%	5.1%	100%
Access CD-ROM	15.2%	2..2%	27.4%	21.1%	16.1%	4.8%	3.3%	100%
View ejournals	15.2%	31.8%	33.3%	12.2%	1.5%	1.2%	4.8%	100%
Publish electronically	4.2%	4.5%	14.6%	23.8%	43.8%	8.3%	0.9%	100%

More than two thirds of the respondents used various electronic publishing fora or the World Wide Web either frequently or regularly. Slightly under half viewed ejournals either frequently or regularly. Last in frequency of use was publishing itself with a combined score (frequently+regularly) of less than 9%. Clearly, for the Psyche respondents the ratio between authors and readers is about 10:1.

4.4 Advantages of electronic journals

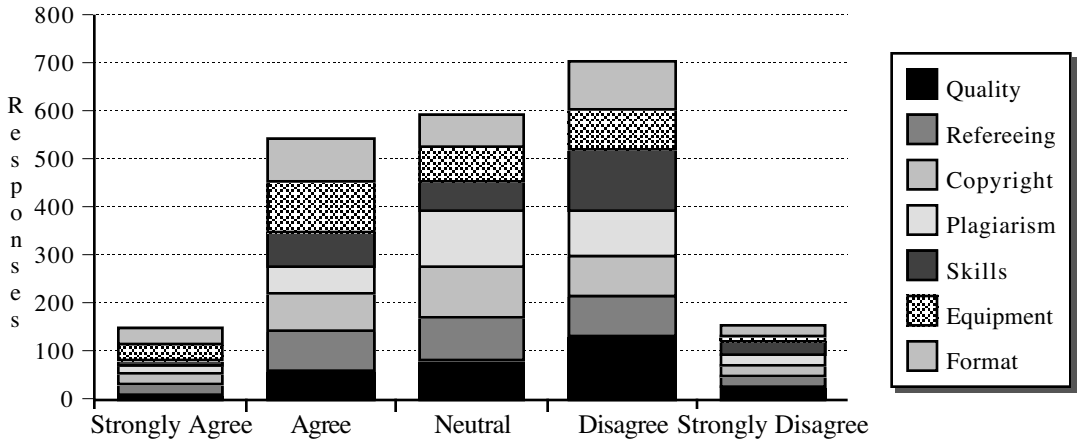
More than half of the respondents either 'strongly agreed' or 'agreed' that speed of publication, 24 hour access, convenience, reduced paper use, ease of searching and affordability were advantages of electronic scholarly journals. The responses were strongly clustered around 'strongly agree' and 'agree', as shown in Figure 1.

Figure 1: Cumulative responses to advantages of electronic publishing



4.5 Disadvantages of electrnic journals

Figure 2: Cumulative responses to disadvantages of electronic publishing



No proposed disadvantage received more than 10% of 'strongly agree' responses. Those disadvantages that received more than 25% of 'agree' responses were lack of refereeing, the

need for special equipment and the unfriendliness of the format. With the disadvantages questions the concentration of answers shifted more towards 'neutral' and 'disagree', as shown in Figure 2.

The general conclusions from the survey sub-phase can be summarised as:
- email surveys have a higher response rate than print;
- respondents use a range of technology to access scholarly ejournals but do little publishing;
- respondents have good access to multimedia technology;
- respondents are generally positive about the advantages of electronic scholarly journals;
- respondents are generally neutral about the disadvantages.

For more detailed discussion of the survey, please consult Treloar (1996) or the online survey results[7] at the EPICentre Web site[8].

5. Document models

The CSIRO Division of Information Technology was charged with investigating document models and/or document type definitions, in relation to the *Australian Journal of Chemistry* and *Psyche*. The investigations concentrated on the Australian Journal of Chemistry, as it was the most challenging of the two journals used in the study. Early on in the research, it became obvious that any activity in defining models/DTD's must take into account end-user requirements as well as the publisher's intent. Without an understanding of these two constraints, any DTD or document model is unrealistic and will necessarily be incomplete. Further research is needed to finalise these requirements, although it may be difficult to elicit responses to hypothetical possible journal forms. The lessons learnt from the document model sub-phase relate to input data formats, document markup, document type definitions (DTDs) and presentation data formats.

Input data formats can be a fruitful source of delays and difficulties. The EPICentre recommendations are to
- finalise text and layouts before starting any conversion;
- start with styled text if at all possible;
- determine a consistent editorial policy for submissions to facilitate smooth conversion later;
- be aware that data held in external application (ie. Chemdraw) file formats present a major challenge if they are to be included into the source text.

Marking-up of existing text is much easier said than done. EPICentre found that processing legacy data in an effective manner mandates support from appropriate tools. To give some idea of the way in which appropriate tools can make a difference, to markup one AJC article in HTML took one day without tools - to markup another fourteen articles took four hours using the RTFtoHTML[9] converter and Adobe Pagemill[10] WYSIAWYG[11] editor.

Designing SGML DTDs is recognised as one of the more arcane areas of electronic publishing. EPICentre confirmed that retrofitting a DTD to an existing corpus of documents is a complex task when working with real-world examples. This was exacerbated by lack of consistency in the logical structure of the AJC articles selected. To mark up one AJC article in SGML took one week (without tools). To mark up the same article in HTML took one day

---

[7] http://www.informit.com.au/epicentre/survey/survey.html
[8] http://www.informit.com.au/epicentre/
[9] http://www.w3.org/pub/WWW/Tools/rtftohtml-3.0.html
[10] http://www.adobe.com/prodindex/pagemill/main.html

(also without tools). It was also determined that SGML-compliant editors required graphic displays for novice users. Tables were particularly problematic.

The investigation of presentation data formats concentrated on PDF and HTML. PDF was found to be only really useful for representing printed formats as rendering is both tied to a particular original page size and not sensitive to the characteristics of the display device. Navigation around a PDF document while reading was tedious unless the user had a two page display. On the plus side, creating PDF's from arbitrary Postscript using Adobe Distiller was remarkably painless. In general, PDF was felt to be a useful way to distribute print images in electronic form with the expectation that the document would need to be printed for ease of use.

HTML was found to be unsuitable for any but simple texts. The layout facilities are too constrained, and the document too simple. In particular, HTML is not powerful enough to represent the rich character set required by the chemistry domain without extensive (and time consuming) use of small GIF files. Realistic deployment of HTML requires both powerful WYSIWYG tools for marking up (such as Pagemill) and tools for site management (such as Sitemill).

6. Conclusion

Looking back, with the benefit of hindsight, many of the conclusions reached now seem obvious and in line with other research. It is important to remember that this research commenced in 1994 and predates the arrival of technologies like PDF, and the ready availability of good tools for working with HTML and SGML. Members of the EPICentre team are now actively cooperating on developing tools and methodologies to facilitate network-based scholarly publishing. The data from the Psyche survey is being further processed and analysed statistically, and a second survey round will compare the results with an equivalent population of readers of print journals.

Bibliography

C. Bailey Jr. (1996), "Network-Based Electronic Publishing of Scholarly Works: A Selective Bibliography"*The Public-Access Computer Systems Review* 6, no 1. Available online at <http://info.lib.uh.edu/pr/v6/n1/bail6n1.html>.

S. Harnad (1991), "Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge",*The Public-Access Computer Systems Review* , 2, no. 1, 39-53. Available online at <http://info.lib.uh.edu/pr/v2/n1/harnad.2n1>.

A. Odlyzko (1996), "Tragic loss or good riddance? The impending demise of traditional scholarly journals" in *Scholarly Publishing: The Electronic Frontier*, eds. Robin P. Peek and Gregory B. Newby, Cambridge, MA, MIT Press/ASIS monograph, 91-102. Available online at <http://netlib.bell-labs.com/netlib/att/math/odlyzko/tragic.loss.txt.Z>.

B. Shackel (1991), *Blend-9: overview and appraisal*, British Library Research Paper 82, London, British Library

J. Senders (1976), "The Scientific Journal of the Future", *The American Sociologist* 11 (August 1976): 160-164

A. Treloar (1996), "Attitudes to electronic publishing among readers of an ejournal: the Psyche survey", *Public-Access Computer Systems Review*, forthcoming.

Appendix - Survey instrument

QUESTIONNAIRE: Electronic Scholarly Publishing

A. INTRODUCTION AND PURPOSE

We would like you to take part in this survey because you receive either the ejournal _Psyche_ or the print journal _The Australian Journal of Chemistry_. This survey is part of a larger joint project between the Royal Melbourne University of Technology (RMIT), Monash University and the Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO). This project has been funded to investigate electronic publishing by the Australian Vice-Chancellor's Committee (AVCC).

This questionnaire is designed to provide a broad picture of the demographics of the readers of the above journals. A sample of this population will then be selected for a more detailed follow-up questionnaire. You will be asked if you wish to be part of such an exercise towards the end of this questionnaire.

Please fill in the blanks and return the questionnaire. A range of return mechanisms (mail, email, fax) are available. Space is provided near the end of the questionnaire for additional comments if required. This questionnaire is also available online at http://www.deakin.edu.au/people/aet/initsurv.txt.

B. BASIC DEMOGRAPHICS

1. Industry Category

Please indicate the industry sector of your employing institution. Choose one category and mark with a cross.

__ Education (Primary/Secondary/Tertiary)
__ Government
__ Industry
__ Consultant/Self-Employed
__ Other (please indicate) _____

2. Position

Please indicate your primary employment role. Choose one category and mark with a cross.

__ Practitioner (e.g. clinical psychologist, industrial chemist)
__ Information Services/Publishing
__ Administration
__ Teaching
__ Student
__ Research
__ Consultant/Self-Employed
__ Retired
__ Other (please indicate) _____

3. Access to Technology

Please mark the pieces of information technology you have ready access to (either at home or at work). Mark all appropriate items with a cross.

__ Personal Computer
__ CD-ROM drive
__ Sound output from computer
__ Colour screen
__ Direct network connection to Internet
__ Modem connection to Internet
__ Other (please indicate) _____

C. ELECTRONIC PUBLISHING

4. Familiarity

Please indicate how often you use the following forms of electronic publishing with the appropriate letter:

a. Frequently b. Regularly c. Occasionally d. Rarely e. Never

__ Subscribe to electronic publishing forums (listservers, BBS, netnews, etc.)
__ Use FTP (File Transfer Protocol) to access materials
__ Use Gopher to access materials
__ Use World Wide Web to access materials
__ Access materials on CD-ROM
__ View electronic journal(s)
__ Publish electronically

5. Advantages and Disadvantages

Below are two possible lists of advantages or disadvantages of electronically published scholarly articles. You may add any items to the lists that you wish. Please indicate your feelings about each advantage or disadvantage to the left of the suggested text according to this scale: SA - Strongly Agree, A - Agree, N - Neutral, D - Disagree, SD - Strongly Disagree.

ADVANTAGES of electronic scholarly journals are:	DISADVANTAGES of electronic scholarly journals are:
__ Speed of publication	__ Poor quality
__ 24 hours a day access	__ Lack of refereeing
__ Convenience	__ Concerns about copyright
__ The way they encourage feedback	__ Increased plagiarism
__ Reduced paper consumption	__ Special skills needed to use
__ Ease of searching	__ Special equipment needed to access
__ Multimedia publications	__ Format is not reader friendly
__ Affordability	__ Communication costs to access them
__ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _	__ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _
__ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _	__ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

D. FOLLOW-UP SURVEY

6. Do you wish to take part in a more detailed survey?

We are seeking a select group of individuals to trial a number of new electronic and print publishing products. You will be asked a more detailed set of questions as part of this trial. If you are prepared to take part in this exercise later this year, please provide your details below. This will be regarded as an expression of interest only, and does not guarantee that you will be selected.

Name: _____
Address: _____
_____
Phone: _____
Fax: _____
Email: _____

****
Any other comments?

****
Please return this questionnaire before December 15, 1995, via any one of:

Post:
Andrew Treloar, School of Computing and Mathematics, Deakin University,
662 Blackburn Road, Clayton, 3168, Australia.

Email:
Andrew.Treloar@deakin.edu.au

Fax:
+61 3 9244 7134

Enquiries to Andrew Treloar at the above addresses or phone +61 3 9244 7461.
****
Based in part on a proposed questionnaire designed by Philip McEldowney,
University of Virgiana (philipmc@Virginia.edu). Last revised September 5, 1995.
****