The Efficacy of Hartley's "Structured Format" in the Teaching and Assessment of Abstract Writing

Cathy Ann Radix
Department of Electrical and Computer Engineering
The University of the West Indies
St. Augustine, Trinidad and Tobago
CathyAnn.Radix@sta.uwi.edu

Crista Mohammed
Department of Electrical and Computer Engineering
The University of the West Indies
St. Augustine, Trinidad and Tobago
Crista.Mohammed@sta.uwi.edu

***Abstract*— The abstract is the quintessential technical writing genre, because abstract writing involves judiciously selecting content and conveying such content in lean, focused writing. This study tested the efficacy of the structured abstract content framework, proposed by Hartley, in the teaching of abstract writing in Electrical and Computer Engineering (ECE). Subject-matter experts and a technical editor independently assessed anonymised abstracts written by four (4) sophomore cohorts based on two (2) source texts. Findings demonstrate that instruction on the structured abstract content framework, as opposed to general abstract instruction, impacts neither content coverage nor writing clarity. However, the study has enabled the authors' to articulate the requirements for the ECE discourse community in a generic, text-blind assessment rubric for abstracts and has provided students with a formal schema for reading technical papers.**

Keywords— electrical and computer engineering; structured abstracts; writing in the engineering sciences; writing for academic purposes; writing-to-learn

INTRODUCTION

If effective writing in the engineering sciences is precise, succinct and purposeful, then the abstract is the quintessential technical writing genre, because abstract writing involves judiciously selecting content and conveying such content in lean, focused writing.

This study investigated the extent to which the content framework of Hartley's structured abstract [1] can support Electrical and Computer Engineering (ECE) sophomores in writing for their discipline, as evidenced by the quality of abstracts that they produced.

The paper begins with a literature review of abstract types, the needs of the ECE discourse community, the impact of structured abstract use, and methods for assessment of abstracts. This is followed by a description of the abstract-writing task and the supporting instruction provided to students in four (4) successive sophomore cohorts – two of which received explicit guidance concerning content selection and sequencing.

The next section describes the experimental procedure, which was took account of published assessment methods for abstracts. The overarching purpose was to determine whether instruction on Hartley's content framework impacted student performance in either content selection or content sequencing. This is followed by a summary of the results and a discussion of the adequacy of the content framework as an instructional tool.

LITERATURE REVIEW

A. Types of Abstracts

Abstracts are categorized in various ways. Lay [2] identifies two types: descriptive abstracts, which describe the purpose, scope and organization of the article without supplying findings or conclusions; and informative abstracts which give the purpose, method, findings and significant conclusions of the study. Zhang [3] classifies abstracts as: indicative abstracts which provide an overview of the study without giving specific outcomes; informative abstracts which share the purpose, method and significant results; and informative-indicative ones which convey the primary elements of the paper and provide an overview of other aspects of the study using indicative statements.

Other descriptions of abstracts focus on the format: Abstracts can be classified as either structured or traditional. In structured abstracts, content is organized under various headings, determined by the information needs of the discourse community. For example, in medical contexts, particularly clinical tests, the structured abstract may include "study design" and "sample" [4], and in leading psychology journals simply "method" is used [5].

In contrast to the structured abstract, the traditional abstract is presented as continuous prose. Traditional abstracts are conceived as "brief summaries that stand alone…and provide a glimpse of the entire document" [2]. In other words, traditional abstracts are less prescriptive about the information required by the readership.

In this study, we tested the content framework of Hartley's structured abstract – background, aims, method, results and conclusion (BAMRC), as a rubric for content selection and sequencing for a continuous prose abstract.

978-1-5090-5920-1/17/$31.00 ©2017 IEEE

B. Information needs of various discourse communities

The abstract is typically associated with technical reports and research articles, where it helps readers quickly select research papers, and orient themselves to key research findings.

Some discourse communities have either switched to (e.g. medical, psychology) or recommend switching to (e.g. software engineering) a structured abstract format for the following reasons:

- Unmasks methodological problems [4]
- Facilitates consistent peer review [4,5]
- Supplies more information than traditional abstracts [4]
- Promotes recall [4]
- Facilitates evidence based research [6].

In a meta-analysis of research on the structured abstract, Hartley et al [5] found that structured abstracts do have some drawbacks – they tend to be longer and their typographical layout sometimes confuses readers. Also, authors using structured abstracts are just as prone to errors of omission and distortion as if they were using a traditional format [5].

Abstracts in the ECE discipline remain traditional in form, consisting of continuous prose. Technical reports written by and consumed in the ECE and wider engineering discourse communities are organized into sections - Introduction, Methods, Results and Discussion (IMRaD) [7]. IMRaD, in terms of content and sequence, aligns closely with the BAMRC framework. If the engineering discourse community has come to accept IMRaD as an acceptable framework for conveying a technical report, then it is likely that an abstract fashioned using BAMRC will meet the information needs of the target discourse community.

C. Novice Writers and the Structured Abstract

Of particular relevance to the current study is research on the use of the structured abstract by novice writers. While there are many studies on the use of structured abstracts, we limited our review to empirical investigations of student use.

Budgen et al [6] compared traditional and structured abstracts written by software engineering students. They found that the structured abstracts were judged to be more complete and clear.

In the field of psychology, Hartley et al [5] found that there is some support for the use of structured abstracts by student-writers, as such abstracts were judged to be of better quality than the traditional abstracts.

There is, to the best of our knowledge and our literature search, no study on the use of structured abstracts in ECE. This study proposes to open up a discussion on whether or not structured abstracts are supportive of the discourse needs of the ECE community.

III. INSTRUCTION AND ABSTRACT PRODUCTION

The abstract-writing task, which the study investigated, is one element within a programme-wide Communication-across-the-Curriculum (CaC) intervention. The task is located in a sophomore course entitled “Introduction to Microprocessors”.

The specific task required each student to produce a 250-word continuous-prose abstract based on third party, magazine-type engineering articles, rather than self-authored material. This way, students read technical content (about microprocessor technology), for the purpose of writing. The learning outcomes were thus two-fold: writing-to-learn (WTL) and writing for academic purposes (WAP).

A. Text Selection for WTL

Texts were selected based on the following criteria:

- Source – Must be a scholarly, leading publication in its field. Toward this end, the exercise drew on IEEE Spectrum (ISSN: 0018-9235, http://ieee.spectrum.org)
- Timeframe – Articles must be recent, that is published within 3 years of the task.
- Subject – Articles must treat substantially with microprocessors
- Length – Articles must pose a light reading load, which we defined as ranging between 3000-4000 words

Students were randomly assigned one (1) of four (4) to six (6) texts. Based on the cohort size (50 – 110 students), between 8 and 20 students within a cohort may have been assigned the same text.

B. Teaching Instruction for WAP

The control groups (2 cohorts) received instruction on the traditional abstract. That instruction focused on the purposes and uses of abstracts. Students were directed to select, summarize and collate the main findings of the articles assigned to them.

In the test groups (2 cohorts), students were introduced to the BAMRC framework, as proposed by Hartley [1]. However, the assignment modified Hartley's layout, by asking students to eliminate headings within the final abstract. The abstracts were presented as continuous prose, as obtains in traditional abstracts, rather than separate content items preceded by sub-headings.

The deliberate use of continuous prose by the test group served three (3) purposes: the final abstract was presented in a layout that is typical in the engineering sciences; students had an opportunity to work on textual coherence, which is a vital writing skill; and there were neither visual nor discourse cues that marked the abstracts as having been produced using Hartley's content framework.

IV. EXPERIMENTAL PROCEDURE

This study examined whether instruction on Hartley's content framework impacted student performance in either

content selection or writing coherence, giving rise to these questions:

1. Did abstracts meet the information requirements of the ECE discourse community?
2. Did the test groups' abstracts more effectively meet information requirements than the control groups' abstracts?
3. Did students produce clear, coherent abstracts?
   Did the test group abstracts produce clearer, more coherent abstracts than the control group?

A. Abstract Evaluation Method

Schriver [8] argues that comprehensive text evaluation should supply information about global aspects of text-quality and information about audience response to the text. Therefore two (2) sets of evaluators were used: a technical editor and six (6) subject matter experts (SMEs).

The technical editor assessed global text quality, as signaled by clarity (Criterion 1: appropriate lexis, Criterion 2: acceptable sentence structure) and coherence (Criterion 3: appropriate use of transitions and Criterion 4: logical sequencing of sentences). Each of these 4 criteria was assessed using a 3-point Likert scale (0 – No evidence/Unacceptable, 1 – Some evidence/Acceptable, 2 – Ample Evidence/Very Good) to produce a final score (maximum 8).

SMEs assessed the relevance of the material selected and presented in the abstracts. Budgen et al [6] used undergraduate evaluators to assess student abstracts. However, students who are not yet full-fledged members of the discourse community cannot be appropriate judges, where the pedagogical goal is to have students meet the discourse needs of full-fledged members of the ECE discourse community. Therefore, this work used faculty members as evaluators.

Previous studies, of abstracts written by novices, employed checklists, against which abstracts were scored [5, 6]. These made performance criteria transparent. However, neither of these checklists were wholly applicable to ECE. There is, to the best of the authors' knowledge, no existing articulation of what constitutes an effective ECE abstract. Following Sless' [9] participatory approach to defining heuristics for assessing texts, eight SMEs produced a rubric (see Appendix 1) against which the abstracts were scored.

The rubric assesses six (6) aspects of content, using a 3-point Likert scale to produce a final maximum score of 12. In addition, the SME group, by consensus, judged "quality writing" to be signaled by text clarity and coherence. The final quality measure was graded 0 (Unacceptable) – 3 (Exemplary).

Readability formulae or fog indices have also been employed to evaluate structured abstracts [1, 5]. However the usefulness of such measures have been questioned extensively, largely because these text-based evaluations focus at the word and sentence levels and not on how the text is functioning globally to produce or not produce meaning for the reader [8].

Budgen et al [6] used such metrics to determine if the selected sample is representative of the population, and whether the samples from different cohorts exhibit similar characteristics. Following Budgen et al [6], the word count and Fleish Reading Ease score [10] for each submission was determined in order to answer this associated question:

Was the sample representative of the larger population (by cohort and by instruction received)?

B. Experimental Design

The study involved assessment of submissions from four (4) successive cohorts – where the instruction method changed for the third and fourth cohorts. There were multiple issues, related to the way in which the task was administered, which influenced the selection of an appropriate sample size and the number of assessors.

The first issue was that no single text was used for all four cohorts. Due to the timeframe selection criterion, texts were used in at most 3 cohorts. Therefore in selecting items for this study, it was necessary to select items that were written by students in response to different texts. We note that Budgen et al [6] and Hartley [1] ignore the relevance/complexity of the source text/subject in their studies. Two texts, ranging between 3000-4000 words, were selected based on their longevity of use – one spanning the former three (3) cohorts, and the other spanning the latter three (3) cohorts.

The second issue was that of assigning subject-matter experts to assess sample items. Because the task necessarily required reading the text before assessing the samples, it was decided that each assessor should assess the set of items related to one of the two texts. As with Hartley et al [5], assessors were asked to assess all items in the set associated with the text, and then if needed multiple assessors were asked to work with each set. The deliberate confounding of the assessor and text selection suggested a final experimental question:

8. Did the selection of text/assessor impact the outcome of Questions 1 – 4?

The third issue was that of balancing sample selection. A balanced sample by cohort for either text would necessarily be unbalanced with respect to the instructional method. A balanced sample by cohort across both texts would bias the first and fourth cohorts by text. Because the experimental questions involved the instructional method, and assessors would work with all items associated with a given text, the decision was therefore taken to balance the sample overall by text and instructional method, while accepting unbalanced samples across cohorts.

The final issue was that of sample size and the significance of the statistical analysis. Budgen et al [6] selected 20/79 and 20/88 thesis abstracts for their study, while Hartley et al [5] selected 50/112 and 50/115 report abstracts. On the basis of these samples, both studies report at the 95% significance level. In the study by Budgen et al [6] ANOVA with a bi-variate model was used to discount the effect of single, independent assessments of 40 items (20 structured, 20 traditional) by 20 student assessors. In contrast, Hartley et al [5] utilized balanced t-tests for samples with unequal variances for 100 items (50 structured, 50 traditional), each scored by 4 faculty assessors (2 using rubric, 2 independent of rubric).

In this work the population size was significantly smaller. The maximum possible balanced sample across texts and instructional method was 18 samples each. An a-priori investigation of t-tests for instructional method, using G*Power, established that values of α=0.05 could not be achieved with the available sample size. Therefore the experiment was designed to provide α=0.10 and power (1-β)=0.80, in balanced t-tests for samples with unequal variances, where a large effect (d=0.8) was presumed. This required a minimum sample size of 30, which was achievable.

In the final selection, for each text, an unbalanced sample was chosen, such that overall 15 abstracts were each obtained from the test and control groups. The number of available student submissions, and the number of samples selected for each text and cohort appear in Table 1.

TABLE 1: POPULATION AND SAMPLE SIZES

	2010	2011	2012	2013	Total
Submissions B	8	11	9		28
Submissions C		17	12	19	48
Sample B [Wehner]	6	3	6	-	15
Sample C [Patterson]	-	6	3	6	15
Sample Total	6	9	9	6	30

C. *Sample Selection and Preparation*

Electronic student submissions from the four cohorts associated with the two selected texts were identified and downloaded from the Campus Learning Management System.

Each file was assigned a random ID#. Samples were selected by sorting the ID#'s, and selecting the first 30 samples that met the year and article requirements specified in Table 1. The text from each sample submission was then extracted, anonymised and randomly assigned a number (e.g. B4) for the associated text (B or C).

The samples were then ordered according to their sample number before being given to the assessors. Two assessors were each allocated a single text.

V. RESULTS

Table 2 presents the results of the word count and Flesh Reading Ease scores for the sample abstracts and the population for each text. The calculated t-values indicate that the samples are representative of the population (i.e. p>>0.1)

Table 3 summarizes assessor results for candidates in the test and control groups. The calculated t-values indicate that no differences are observed between control and test groups (i.e. p>>0.1)

Table 4 shows inter-correlations between subject matter experts and technical editor scores.

TABLE 2: READABILITY AND WORD COUNT FOR SAMPLE AND POPULATION

	Sample		Population		Welch's t value
Item M s.d.	B (15)	C (15)	B (28)	C (48)	
Word count	273.87 32.66	237.53 33.08	254.89 35.46	231.04 33.78	B t=1.7621 C t=0.6599
Flesch Reading Ease	34.53 11.98	31.93 8.68	31.32 10.58	30.23 8.04	B t= 0.8715 C t=0.5728

TABLE 3: SAMPLE COUNT, MEAN CONTENT SCORES (AND STANDARD DEVIATIONS) FOR EACH RATER AND RATER CATEGORY

Rater		Pre framework	Post framework	Welch's t value
SME-BA	n M s.d.	9 5.67 1.86	6 4.78 2.43	t = 0.761 p = 0.469
SME-CA	n M s.d.	6 5.67 3.39	9 6.44 2.00	t = 0.501 p = 0.632
Both SME	n M s.d.	15 5.8 2.68	15 5.47 2.17	t = 0.371 p = 0.714
Tech. Ed.	n M s.d.	15 2.50 1.32	15 2.53 1.18	t = 0.066 p = 0.948

TABLE 4: INTER-CORRELATIONS, R, BETWEEN TECHNICAL EDITOR AND SUBJECT MATTER EXPERT SCORES

Measures	Pre	Post	Chi squared
SME Content - Tech. Ed.	0.48	0.40	5.03 p=0.081
SME Style – Tech. Ed.	0.24	0.42	
SME Quality – Tech. Ed.	0.44	0.48	

VI. DISCUSSION

Did abstracts meet the information requirements of the ECE discourse community?

Overall, in both the control and test groups, the abstracts met the information needs of the assessors. Only 6 of the 30 (20%) samples were judged as unacceptable by subject-matter experts. The average SME content mark was approximately 6/12 – suggesting that most students to an adequate degree were able to identify content in the required information fields (context; rationale; aim/purpose; method; findings; significance of work).

Did the test groups' abstracts more effectively meet information requirements than the control groups' abstracts?

As it relates to meeting the information needs of the assessors, the test group did slightly better than the control group. Four of the unacceptable samples appeared in the control group, and two in the test group. While the test group appears to have fewer unacceptable samples, this test statistic is not significant.

Did students produce clear, coherent abstracts?

To some extent, students were able to produce clear, coherent abstracts. Twelve (12) of the thirty (30) or 40% of the samples were judged as unacceptable in one or more areas by the Technical Editor. This suggests that further language intervention is required.

Did the test group abstracts produce clearer, more coherent abstracts than the control group?

Samples judged as unacceptable in one or more areas by the Technical Editor were equally distributed between the control and test groups.

Was the sample representative of the larger population (by cohort and by instruction received)?

The sample was representative of the overall population. The samples did not possess significantly different means from the population for both word count and Flesh Reading Ease scores.

Did the selection of text/assessor impact the outcome of Questions 1 – 4?

Text selection and assessor did not impact the outcomes. Results for individual texts followed the same trends with unacceptable samples split between the two texts, 7/5 and 4/2.

CONCLUSION

The inherently dualistic nature of the abstract-writing task requires students to both identify salient content and compose a coherent abstract. This paper has described a study that considered whether the use of a structured abstract content framework could positively impact the quality of abstracts produced by sophomore ECE students for a CaC intervention embedded in a technical course. No impact has been identified.

However, the authors' note that this practice leads to additional intangible benefits, namely:

- a schema for extracting pertinent information from research
- a content framework for abstracts

One unique aspect of our experimental procedure stands in contrast to the work by Hartley [1] and Budgen et al [6]. Rather than have students rewrite existing abstracts—which really is a test of editorial skills, or have them write abstracts for self-authored documents—where the quality of the source document may impact the quality of the abstract, this study used previously published papers. This meant that the quality of the original text was controlled as far as possible for both the test and control groups, so that there should have been no impact of the source text on the quality of the abstract.

REFERENCES

[1] J. Hartley, "Improving clarity of journal abstracts in psychology: the case for structure," Science Communication, vol.24, pp. 366-379, 2003.

[2] M. Lay, B. Wahlstrom, C. Rude, C. Selfe, and J. Selfe, Technical Communication, 2nd ed.,Boston: McGraw-Hill, 2000.

[3] C. Zhang, and X. Liu, "Review of James Hartley's research on structured abstracts," Journal of Information Sciences, vol. 37, no. 6, pp. 570-576, 2011.

[4] A.Taddio, T. Pain, F. Fassos, H. Boon, A. Lane, and T. Einarson, "Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal and the Journal of the American Medical Association," Canadian Medical Association Journal, vol. 150, no. 10, pp.1611-1615, 1994.

[5] J. Hartley, J. Rock, and C. Fox, "Teaching pychology students to write structured abstracts: an evaluation," Psychological Teaching Review, vol. 11, no.1, pp. 2-11, 2005.

[6] D. Budgen, A. Burn, and B. Kitchenham, "Reporting computing projects through structured abstracts: a quasi-experiment," Empirical Software Engineering, vol. 16, pp. 244-277, 2011.

[7] Al-Othmany, D.; Solaiman Ali, M., "How to be an effective technical writer?," *IEEE Global Engineering Education Conference (EDUCON), 2012* pp.1-8, 17-20 April 2012.

[8] K. Schriver, "Evaluating text quality: the continuum from text-focused to reader-focused methods," IEEE Transactions on Professional Communication, vol.32, no.4, pp. 238-255, 1989.

[9] D. Sless, "Designing public documents," Information design journal + document design, vol.12, no.1, pp. 24-35, 2004.

[10] D.Hanauer, "Fleish Reading Ease score," Supermagnus Software. Accessed April 4 2016 <http://www.supermagnus.com/mac/Word_Counter/index.html>

Cathy Ann Radix is a longstanding member of the IEEE, having actively participated in the creation of the country-chapter IEEE-TT. Her interest in instructional scaffolding has led to research in the use of visual organizers, concept inventories, and learning preferences for learning and assessment. She is part of the faculty-team responsible for teaching Embedded Systems, within the Department of Electrical and Computer Engineering, The University of the West Indies.

Crista Mohammed leads the Communication across the Curriculum Programme in the Department of Electrical and Computer Engineering, The University of the West Indies. She holds a Post Graduate Diploma in Education: the Teaching of English with distinction (UWI) and an MA in Technical Communication (SHU) with distinction. Her research interests include genre approaches to writing and the assessment of writing.

APPENDIX 1: SUBJECT-MATTER ASSESSOR CHECKLIST

1. Context *The abstract refers to the General Problem Space and/or pre-existing work.* 0-No/Unacceptable 1-Yes/Acceptable 2-Very Good/Comprehensive	
2. Rationale *The abstract details the motivation and/or specific problem to be addressed by this work* 0-No/Unacceptable 1-Yes/Acceptable 2-Very Good/Comprehensive	
3. Aims/Purpose *The abstract details what the work intend(ed) to achieve* 0-No/Unacceptable 1-Yes/Acceptable 2-Very Good/Comprehensive	
4. Method *The abstract identifies specific resource requirements and/or plan of work* 0-No/Unacceptable 1-Yes/Acceptable 2-Very Good/Comprehensive	
5. Findings *The abstract identifies relevant/key results/problems from the study* 0-No/Unacceptable 1-Yes/Acceptable 2-Very Good/Comprehensive	
6. Significance of Work *The abstract makes concluding statements re: impact/significance of the findings* 0-No/Unacceptable 1-Yes/Acceptable 2-Very Good/Comprehensive	
Content Total (12 marks)	

7. Spelling, Grammar *Spelling and grammatical errors (if any) hamper readers' understanding* 0 – Yes; 1 - No	
8. Logical Development *Content items appear in the sequence:* *(1->2/3(interchangeable)->4->5->6)* 0-Sequence cannot be discerned; 1-Partially correct sequencing of not less than 3 areas; 2-Ideal sequence observed	
Writing Style Total (3 marks)	

Overall Abstract Quality Criteria	Descriptor – Score (3 marks)
Unacceptable in 4 or more areas	Unacceptable - 0
Acceptable or better in at least 4 areas AND no more than 3 Unacceptable areas	Acceptable - 1
Very Good in at least 2 content areas AS WELL AS at least Acceptable in Logical development AND no more than 2 Unacceptable areas	Very Good - 2
Very Good in at least 4 areas AS WELL AS Very Good in Logical Development AND No Unacceptable areas.	Exemplary - 3