ASEAN Multidisciplinary Research Journal
Vol. 3, No. 1, s. 2019

Web-based *Sentheme Mining* on Abstracts of Junior Researchers: A Thematic Analysis for 'Attractive' Writing for Beginners

JUN S. CAMARA
Faculty, Pangasinan State University, Pangasinan, Philippines
& Philippine Association of Research Practitioners, Educators,
and Statistical Software Users, Inc. [PARESSU, Inc]
www.paressu.org; jcamara@psu.edu.ph

Publication Date: December 1, 2019

Abstract

Improving academic writing is a key towards successful publishing, and assessment of pre-published write-ups is the beginning of that ending. Words have 'emotions', and the ability to control it in write-ups is an evolving industry skill. This paper aimed to profile and analyze through web-based sentiment analysis and text mining with theme generation of the scientific paper abstracts of junior researchers in the Philippines. Significant findings showed that majority of the specimen abstracts conformed with academic writing standards, and obtained a passing rating from conference judges. Sentiment Analysis revealed an equal writing behavior of junior researchers in wording abstracts positively or negatively, even if majority were found to have written abstracts neutrally. Text Mining reported the frequent use of the words 'effective' and 'potential' for positively-worded and negatively-worded conclusions of abstracts, respectively. Generating themes from these premises, thematic analysis proposed four principles of writing abstracts to attract acceptance of papers in science fairs. 'Sentheme Mining' as a research procedure was recommended, among others things, for use in future studies involving qualitative studies of paper abstracts or full articles.

Keywords: Abstracts, Sentiment Analysis, 'Sentheme Mining', Text Mining, Writing

ISSN [online] *2672-2453* ISSN [print] *2672-2445*

INTRODUCTION

As early as in the secondary education in the Philippines and in other countries, the concept of writing research papers has been in place amongst learners. Research papers submitted are generally scientific in nature because normally research writing is taught in science subjects. Erum & Pambid [1] found research or manuscript writing as a less employed science process skill by both students and teachers in open science inquiry process compared with other open science processes, and concluded with a possible misconception among the students that researching merely involves going to the library and finding the meaning of scientific terms, among other conclusions. Ulla [2] noted that there were only a handful of studies concerning the challenges and benefits of doing research in the Philippines and the ASEAN Region despite the positive impact of conducting research on teacher's teaching pedagogies and professional development. Further, his study revealed that among other challenges, the lack of research skills and knowledge was one discovered factor experienced by teachers at this modern times. Camara [3] noted that during the time that the Philippine Special Science Program was on its birth implementation in the Philippines, no syllabus nor curriculum for use by Special Science Research Teachers were readily available which led him to developing, refining, and validating from 2015 to 2017 a spirally progressive and contextualized research competency checklist. In another study, Camara [4] pioneered the development and qualitative assessment through content analysis of a simplified and practical tool for beginning researchers in conceptualizing a research topic, and proposed the term 'START Approach' which means *See, Think, Aim, Refine* and *Tell*. Qasem & Zayid [5], having found that to write in English was the predominant challenge experienced by majority of his study participants in conducting research, proposed to give more focus in academic writing with more activities and tasks and training workshops. In writing research articles, Kurniawan *et al.* [6] found in his study, however, that research article publication is not only the content truthfulness, but also the informative and attractive value of the paper itself.

Thus, the challenge to write research articles effectively using the English language is one problem, and to determine the sense of 'informative and attractive' value is another story. In this paper, we dealt with how a research article is found 'informative and attractive' through sentiment analysis, and analyze what words or phrasing patterns make these articles 'attractive' to conference judges, or readers.

The researcher believes that writing is truly arbitrary, and setting rigid standards is crushing the ability of full expression itself. There is a need to revisit the standards we once held the 'best' and compare it with the kind of millennial writers of today. This is a pioneering research.

OBJECTIVES OF THE STUDY

Generally, this study, an excerpt, aimed to profile and analyze through web-based text mining and sentiment analysis, to be known as *Sentheme Mining*, of paper abstracts of junior researchers submitted in a national research competition in the Philippines. and generate themes on the principles for academic research yet 'attractive' writing:

Specifically, this study aims to:

- profile the paper abstracts of junior researchers in terms of regional distribution, length of title, length of abstracts, and remarks on initial scores;
- perform web-based sentiment analysis on the paper abstracts as a whole and on the conclusions and recommendations section to categorize those which are *positive, negative,* and *neutral* in writing;
- perform web-based text mining on the *positively, negatively* or *neutrally*-categorized paper abstracts to determine the frequency of common words or phrases in the word cloud; and
- thematically analyze the results to generate principles for 'attractive' writing.

ISSN [online] *2672-2453* ISSN [print] *2672-2445*

MATERIALS AND METHODS

Materials

Materials used as sources of data in this study included fifty-five abstracts [7][10], which were coded [4] by an independent assistant researcher from Code 1 to Code 55 submitted as entries to a national research competition in the Philippines in 2018, and are thus, representative for the study in a larger scale. For the web-based tools, the generic sentiment analysis and voyant tools were employed. The generic web-based sentiment analysis is found at *https://app.monkeylearn.com/main/classifiers/cl_pi3C7JiL%EF%BB%BF/* while the voyant tool for text mining is found at *https://voyant-tools.org/*. Both of these tools were retrieved for direct web use from October to September 2019 and are still available up to this writing. The abstracts used in this study are the official abstracts in a national research competition in the Philippines sent to the email of the researcher. The researcher was one of the invited judges of the said competition. For the purpose of confidentiality, the titles, name of abstract authors, the winners, rankings, scores and other details of the research competition are not given in this study – thus, there is no conflict of interest this study would pose to the general public.

Methods

Mixed method is employed in this study. This study used a qualitative design because it utilized content analysis [4], more specifically 'summative content analysis', and it generated writing principles using thematic analysis. A Summative content analysis involves counting and comparisons, usually of keywords or content, followed by the interpretation of the underlying context [8]. Further, this study employed quantitative method because data are analyzed quantitatively using checklist in deciding some aspect of the data analysis, not to mention the data produced by both the sentiment analysis and text mining through voyant tools are in percentages. The use of quantitative approach to data analysis increases the reliability of the findings, and is regard as triangulation.

The study followed three (3) phases.

Phase 1. ***Sentiment Analysis.*** The study started with preparing the abstract for sentiment analysis using the generic sentiment analysis tool. The whole abstract was first analyzed, then the section on 'conclusions and recommendations'. For both analysis, the results were tabulated, and the percentages of sentiment as revealed by the tool were recorded.

Phase 2. ***Text Mining.*** The processed abstracts were categorized based on the results of the sentimental analysis on 'conclusions and recommendations', i.e. *positive, negative, or neutral*. The positively-worded conclusions were combined in one file. The same was done with the negatively-worded, and the neutrally-worded. Next, each of these categories were analyzed for text mining by processing them in the voyant tool. The world cloud with the word frequency were recorded.

Phase 3. ***Theme Writing.*** Using the word clouds and frequency counts, themes about 'attractive' writing were written as a list of favorable academic writing principles.

The researcher personally referred to this methodology as ***Sentheme Mining***, and follows this order: Sentiment Analysis > Text Mining > Theme Writing.

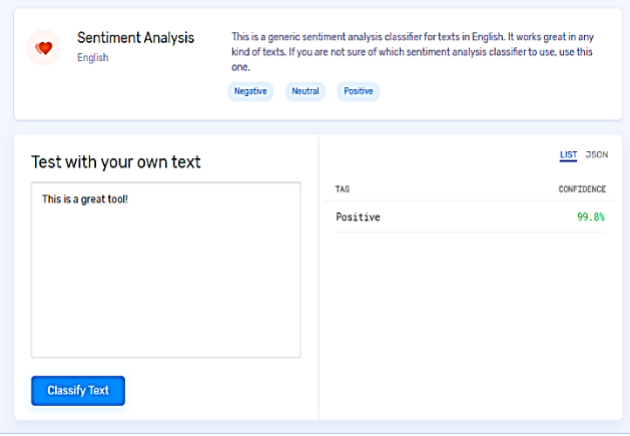

Figure 1. Generic Sentiment Analysis Tool

ISSN [online] *2672-2453* ISSN [print] *2672-2445*

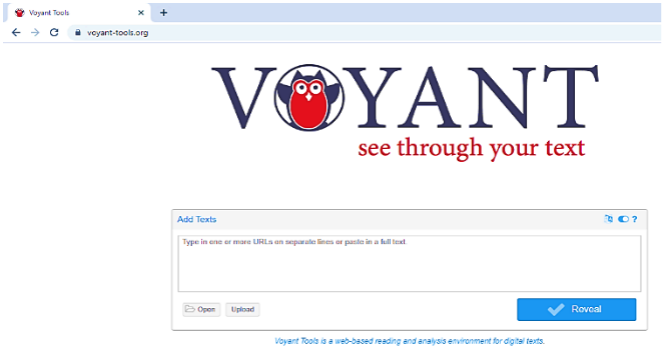

Figure 2.. The Voyant tool for Text Mining

RESULTS AND DISCUSSION

Distribution of the Junior Scientific Papers per Region in the Philippines

Table 1. Number of Junior Scientific Papers (JSP) per Region

Region	*f*	*%*
I	3	*5.45*
II	2	*3.64*
III	7	*12.73*
IV-A	14	*25.45*
IV-B	15	*27.27*
V	1	*1.82*
NCR	13	*23.64*
Total	**55**	**100.00**

Table 1, displays the regional distribution of the fifty-five (55) specimen abstracts analyzed in this study. The table clearly shows that majority of the abstracts were submitted by competitors coming from Region IV-B *(27.27%)*, followed by Region IV-A *(25.45%)*, then NCR *(23.64%)*. Further, the table shows that one-third of the population comes from combined competitors from Regions I, II, III, and V. Other regions are not already listed in the table because no paper on scientific category was registered from their regions.

Profile of the Junior Scientific Papers

Table 2. Summary on the Profile of the Abstracts

Parameter	*f*	*%*
Length of Title		
Fine	15	27.27
Medium	19	34.55
Long	21	38.18
Length of Abstract		
Prescribed	49	89.10
Not Prescribed	6	10.90
Rating Remarks		
Passed	52	94.55
Failed	3	5.45
Total	**55**	**100.00**

Table 2 presents the summary on the profile of the specimen abstracts. The full-length description of all the abstracts are found in the appendices of this article. In terms of *length of title*, the table reveals that majority of the specimen abstracts are long-worded *(38.18)*. More than one-third are medium in length *(34.55%)*, and a little more than one-fourth are considered worded just fine *(27.37%)*.

Bavdekar [10] stated that title is an important part of the article, and it condenses article content in a few words and captures readers' attention. According to him, a good title for a research article should be able to introduce the research work to the fullest extent, but in a concise manner, and accepted the fact that writing scientific titles that are informative and attractive is a challenging task.

In terms of *length of abstract*, almost all of them *(89.10%)* followed the prescribed 250-word rule for abstract, while a handful did not follow the prescribed number of words *(10.90%)* for whatever reason. In terms of *Rating Remarks*, almost all of them 'Passed' *(94.55%)* the screening by the conference judges, while a very few were given a failing score *(5.45%)*. The cut-off score was 48 points of 65 total points.

ISSN [online] *2672-2453* ISSN [print] *2672-2445*

Sentiment Analysis of the Abstracts Profile

Table 3. Individual summary on the Sentiment Analysis of the Specimen Abstracts for 'Conclusions and Recommendations'

Code No.	Reg	Abstract			Con-Rec		
		(+)	(-)	0	(+)	(-)	0
1	1		1			1	
2	1	1					1
3	4-B		1				1
4	3	1				1	
5	1		1			1	
6	NCR	1				1	
7	NCR	1				1	
8	NCR		1				1
9	3	1				1	
10	4-A	1					1
11	4-B	1				1	
12	4-B	1				1	
13	4-B			1			1
14	3			1		1	
15	4-A			1			1
16	4-A			1			1
17	NCR			1			1
18	NCR	1					1
19	4-A		1		1		
20	4-A			1	1		
21	4-A			1			1
22	4-A	1			1		
23	NCR		1		1		
24	NCR			1	1		
25	4-B	1					1
26	4-B			1		1	
27	4-B			1			1
28	4-A	1			1		
29	4-B			1	1		
30	4-A			1			1
31	NCR		1			1	
32	4-A	1			1		
33	4-B	1				1	
34	NCR	1					1
35	NCR	1			1		
36	*		1				1
37	NCR	1				1	
38	3	1			1		
39	2	1			1		
40	4-B			1		1	
41	4-B	1				1	
42	NCR	1			1		
43	4-A	1			1		
44	*	1					1
45	5			1			1
46	4-A	1			1		
47	*	1			1		
48	3	1					1
49	3	1			1		
50	4-B			1			1
51	NCR	1				1	
52	3	1			1		
53	4-A	1				1	
54	4-B	1					1
55	4-B			1	1		
	f	***31***	***8***	***16***	***0***	***17***	***17***	***21***
	%	***56***	***15***	***29***	***0***	***31***	***31***	***38***

ISSN [online] *2672-2453* ISSN [print] *2672-2445*

Table 3 shows the distribution of the fifty-five specimen abstracts in terms of sentiment analysis. In terms of the *Abstract*, the table clearly reveals that more than a half of the specimen abstracts are positively-worded *(56%)*. Many are neutrally-worded *(29%)*. And, a handful is worded negatively *(15%)*. In terms of the *Conclusions and Recommendations*, the percentages are close to one another, but majority are worded neutrally *(38%)*. The percentage of positively-worded to negatively-worded portion is perfectly equal *(31 % - 31%)*. It is very interesting to note that Abstracts Coded 36, 44, and 47 which failed the screening, as revealed in the Appendix, resulted to a 'Neutral' wording. This implies that in scientific paper writing, neutrality is not encouraged. Note that the region where each abstract belongs was purposely not written, for confidentiality.

Text Mining Results on Conclusions and Recommendations of the Specimen Abstracts

Figures 3, 4 and 5 display the cloud word through Cirrus tool of the voyant tool for text mining. The displayed words appear to be different for each figure, though there are words that are present in all three like 'results', 'showed'. This implies that these two words would normally appear in a section on Conclusions and Recommendations.

Positively-worded conclusions and recommendations, based on Figure 3, revealed that the most common term found in the specimen abstracts that possibly made it positive is the term 'effective' – an adjective in use. Other terms include soil, contaminated, results, and product, among others more. Interestingly, the word 'effective' is found as one of the top adjectives used by both industry and non-industry reporters in writing abstracts in a study [10] involving 306, 007 publications reviewed.

Negatively-worded conclusions and recommendations, based on Figure 4, revealed that the most common term found in the specimen abstracts that possibly made it negative is the term 'potential' – a noun in use. Other terms include results, extract, study, and agent, among others more. Interestingly, the word 'potent/ial' is discovered among published abstracts [10] as one of the top

Figure 3.. Cloud Word for 'Positive' Sentiments

Figure 4.. Cloud Word for 'Negative' Sentiments

Figure 5. Cloud Word for 'Neutral' Sentiments

ISSN [online] *2672-2453* ISSN [print] *2672-2445*

words used, and was more commonly used by industry authors compared with non-industry authors.

Neutrally-worded conclusions and recommendations, based on Figure 5, revealed that the most common term found in the specimen abstracts that possibly made it neutral is the term 'showed' – a verb in use. Other terms include study, results, extract, and activity, among others more.

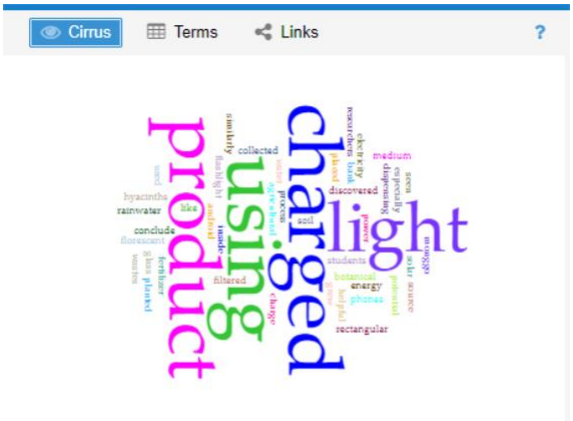

Figure 6. 'Failed' but 'Neutrally' worded

Interestingly, Figure 6 displays a 'non-direction' view of conclusions and recommendations without any trace of positive or negative – the words do not measure up to how a conclusion and recommendations section is written. No signal verbs for findings are found.

Themes found in 'Attractive' Writing

Using ***'Sentheme Mining'*** – a combined methodology first used and applied in this study, the following themes are generated:

1. *Find which is, and expand with, 'effective' area or section of the study.*
   There is a sense of positive writing when adjectives like 'effective' or its variant 'efficacious' [10] is used to describe the results of the findings. This theme suggests that to write 'attractive' abstracts is to find wherever possible the area which the study became effective, and expand on it.

2. *Use 'potential' – not at the end.*
   Recency effect – the effect that enables you to remember what is seen or read or recognized last in a series – could impact the 'image' of the Abstract. The 'recent' image of the Abstract have to provide an image that it has 'done something' or contributed to science, and not another possibility without end. To use the word 'potential' is not forbidden as it is frequently used in writing [11] but to write it at the end sentence of the Abstract may mistake 'association for causality' – they could be associated, but not actually caused by the findings. This implies that recommendations are 'not' required in abstract, and that conclusions are already enough.

3. *No clear verbs make it 'neutral failed'.*
   Verbs indicate much about the findings in the study, and the choice of verb usage to introduce the findings could impact the acceptance or non-acceptance of a paper. When the conclusion contains too much noun, it becomes neutral, and it will fail acceptance of the paper, based on Table 3.

4. *Choose which way to earn an 'aye'.*
   'Aye' means yes, or approval. Sentiments performed revealed that you either choose to be positive in writing, or negative, and you earn an approval in a scientific paper. The positive or negative tone of the conclusions and recommendations will give you a sense of 'aye' in a research fair, based on Table 3.

CONCLUSIONS AND RECOMMENDATIONS

Using '***Sentheme Mining'*,** the researcher concludes the following about the study:

- Majority of the specimen abstracts are written by researchers from the MIMAROPA Region *(27.27%)*, with long-worded titles *(38.18%)*, conforming the 250-worded abstract rule *(89.10%)*, and obtained a 'Passed' rating from the conference judges *(94.55%)*;

ISSN [online] *2672-2453* ISSN [print] *2672-2445*

- Web-based Sentiment Analysis of Abstracts revealed that majority are positively-worded *(56%)*, but when only the conclusions and recommendations sections were considered, the number of positively- and negatively-worded sections is equal *(31%)* even if majority are worded neutrally *(38%)*;
- Web-based Text Mining reported clear and most frequent used of the term 'effective' for positively-worded sections, and 'potential' for negatively-worded sections, while no clear and frequent words were shown for neutrally-worded sections on conclusions and recommendations;
- Thematic Analysis showed that 'neutrally-worded' sections on conclusions and recommendations all failed the screening process based on the data; and, further,
- Thematic Analysis generated four (4) principles in attractive abstract writing including *expanding the presence of effective section of the study, using the word 'potential' not at the end of the abstract, use of clear verbs to avoid neutrality and non-acceptance of paper,* and *choosing verbs to introduce findings appropriately*.

With all these conclusions, the researcher recommends the following:

- review of academic verbs could be integrated in undergraduate research subjects;
- '*Sentheme Mining'* could be performed in other sections of the Abstract including the Introduction or Background which draw readers to reading the paper;
- the use of other web-based approaches in sentiment analysis and text mining like rapidminer.com and the like is also recommended.

ACKNOWLEDGMENTS

The researcher is very grateful to the recognition and trust of Dr. Joseph R. Jacob, the Founder and President of the National Science Engineering Fair Philippines (NSEF) who invited my humble person as Conference Judge to the 1st NSEF Philippines, 2018, for both scientific and engineering fair categories. Gratitude is also given to J. P. Pasana and E. A. R. Pasana as manpower, and for A. B. Acosta and J. V. Bernardo for their printing services.

REFERENCES

[1] Erum, G. P. & R. C. Pambid. (2018). Open Inquiry Process in Special Science Class in Grade 8. *PSU Multidisciplinary Research Journal.* Vol 1., No. 1.

[2] Ulla, M. B. (2018). Benefits and Challenges of Doing Research: Experiences from Philippines Public School Teachers. *Issues in Educational Research*. Vol. 28, No. 3.

[3] Camara, J. S. (2018). Spirally Progressive and Contextualized Research Curriculum Competency Checklists for the Philippine Special Science Program. *Asian Journal of Multidisciplinary Studies.* Vol 1., No. 3

[4] Camara, J. S. (2019). The START Approach – A Simplified and Practical Tool for Beginning Researches. *Southeast Asian Journal of Science and Technology*. Vol 4, No. 1

[5] Qasem, F. A. A. & E. I. M. Zayid. (2019). The Challenges and Problems Faced by Students in the Early Stage of Writing Research Projects in L2, University of Bisha, Saudi Arabia. *European Journal of Special Education Research*. Vol 4, No. 1

[6] Kurniawan, A. B., Warsono, D. Sutopo, & S. W. Filtriati. (2019). The Implementation of Effective Method for Writing Research Articles. *International Journal of Scientific and Technology Research*. Vol. 8, No. 9

[7] Tavsancil, E., G. G. Citak, & F. Kezer. (2011). The Investigation of the Abstracts of Theses and Dissertations in the Domain of Measurement and Evaluation. *The International Journal of Educational Researchers*. Vol. 3, No. 1

[8] Hsiesh, H. & S. E. Shannon. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research.* Retrieved on November 2019 at https://journals.sagepub.com/doi/abs/10.1177/1049732305276687?journalCode=qhra#articleCitationDownloadContainer

ISSN [online] *2672-2453* ISSN [print] *2672-2445*

[9] Bavdekar, S. B. (2016). Formulating the Right Title for a Research Article. *Journal of the Associations of Physicians of India.* Vol. 64. Retrieved on November 2019 at https://japi.org/february_2016/08_aow_formulating_the_right.pdf

[10] Cepeda, M. S. *et al.* (2015). Use of Adjectives in Abstracts when Reporting Results of Randomized, Controlled Trials from Industry and Academia. Retrieved on November 2019 at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4359185/

[11] Andrade, C. (2011). How to write a good abstract for a scientific paper of conference presentation. *Indian Journal of Psychiatry.* Vo. 53, No. 2. Retrieved on Nov 2019 at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3136027/

ISSN [online] *2672-2453* ISSN [print] *2672-2445*

Appendix 1: Profile of Specimen Abstracts

Code No.	*Reg*	*Title No. of Words*		*Meaning*	*Abstract No. of Words*	*Meaning*	*Initial Score*	*Meaning*
1	1	23	1	Long	249	Prescribed	60	Passed
2	1	16	1	Medium	228	Prescribed	59	Passed
3	4-B	17	2	Long	250	Prescribed	55	Passed
4	3	13	2	Medium	146	Prescribed	53	Passed
5	1	17	3	Long	250	Prescribed	59	Passed
6	NCR	14	3	Medium	221	Prescribed	51	Passed
7	NCR	13	4	Medium	301	Not Prescribed	51	Passed
8	NCR	19	4	Long	248	Prescribed	57	Passed
9	3	21	5	Long	247	Prescribed	50	Passed
10	4-A	14	5	Medium	249	Prescribed	57	Passed
11	4-B	15	6	Medium	248	Prescribed	54	Passed
12	4-B	19	6	Long	248	Prescribed	59	Passed
13	4-B	22	7	Long	250	Prescribed	55	Passed
14	3	14	7	Medium	161	Prescribed	52	Passed
15	4-A	9	1	Fine	250	Prescribed	55	Passed
16	4-A	18	8	Long	233	Prescribed	58	Passed
17	NCR	15	8	Medium	406	Not Prescribed	58	Passed
18	NCR	18	9	Long	250	Prescribed	51	Passed
19	4-A	8	2	Fine	212	Prescribed	49	Passed
20	4-A	14	9	Medium	248	Prescribed	56	Passed
21	4-A	12	3	Fine	250	Prescribed	51	Passed
22	4-A	15	10	Medium	223	Prescribed	51	Passed
23	NCR	20	10	Long	226	Prescribed	55	Passed
24	NCR	17	11	Long	249	Prescribed	58	Passed
25	4-B	10	4	Fine	215	Prescribed	49	Passed
26	4-B	16	11	Medium	250	Prescribed	58	Passed
27	4-B	8	5	Fine	248	Prescribed	51	Passed
28	4-A	21	12	Long	250	Prescribed	50	Passed
29	4-B	11	6	Fine	248	Prescribed	50	Passed
30	4-A	14	12	Medium	258	Prescribed	52	Passed
31	NCR	17	13	Long	250	Prescribed	58	Passed
32	4-A	12	7	Fine	290	Npt Prescribed	49	Passed

ISSN [online] *2672-2453* ISSN [print] *2672-2445*

33	4-B	13	13	Medium	250	Prescribed	53	Passed
34	NCR	16	14	Medium	215	Prescribed	54	Passed
35	NCR	17	14	Long	250	Prescribed	55	Passed
36	*	12	8	Fine	241	Prescribed	48	Failed
37	NCR	27	15	Long	308	Not Prescribed	57	Passed
38	3	5	9	Fine	250	Prescribed	54	Passed
39	2	14	15	Medium	250	Prescribed	51	Passed
40	4-B	23	16	Long	258	Not Prescribed	55	Passed
41	4-B	15	16	Medium	214	Prescribed	57	Passed
42	NCR	28	17	Long	204	Prescribed	51	Passed
43	4-A	11	10	Fine	240	Prescribed	58	Passed
44	*	6	11	Fine	250	Prescribed	45	Failed
45	5	14	17	Medium	217	Prescribed	50	Passed
46	4-A	7	12	Fine	248	Prescribed	55	Passed
47	*	4	13	Fine	250	Prescribed	48	Failed
48	3	9	14	Fine	249	Prescribed	57	Passed
49	3	18	18	Long	248	Prescribed	59	Passed
50	4-B	13	18	Medium	192	Prescribed	59	Passed
51	NCR	22	19	Long	220	Prescribed	56	Passed
52	3	19	20	Long	254	Prescribed	58	Passed
53	4-A	28	21	Long	250	Prescribed	57	Passed
54	4-B	15	19	Medium	158	Prescribed	50	Passed
55	4-B	5	15	Fine	257	Not Prescribed	51	Passed

* *those specimen abstracts that obtained a 'failing' score*

ISSN [online] *2672-2453* ISSN [print] *2672-2445*