Abstract

To avoid the subjectivity and a risk of reliability, selecting the most suitable method of scoring written texts is important in the evaluation of academic writing ability. Thus, previous authors proposed several methods of marking, namely holistic method, the impression method, error count method, analytic method, etc., for scoring academic writing texts in different genres at various levels. However, very little attention has been paid to identifying effective methods of scoring research papers writing ability. Therefore, this paper aims to propose using analytic method including an attempt to separate various aspects of a composition. Specifically, in this study, analytic scale is proposed to evaluate research papers writing ability of EFL/ESL undergraduate students. According to this marking scale, writing components, namely content (relevant ideas), organization (structure and coherence), language use (vocabulary and grammar choice), and mechanics use (punctuation and spelling based on APA style) are graded separately. Delphi technique (DT) was used to validate it through the interviews of experts including two boards of ten experienced and qualified lecturers of TESOL and curriculum studies in Can Tho University (CTU), Vietnam and Universiti Sains Malaysia (USM). The results from the interviews revealed that scoring rubrics for the assessment of research papers writing ability were determined. The results from pretest and posttest on research paper writing in English indicated that the process genre approach to research paper writing instruction significantly improved university students’ research paper writing ability in terms of content, organization, language and mechanics. This paper is, therefore believed to contribute to practical applications for research papers reviewers, lecturers, examiners and undergraduate and postgraduate students in EFL/ESL contexts.

Keywords: Research papers writing ability; Scoring Rubrics; EFL/ESL undergraduate students

1. Introduction

1.1 Background of the study

A number of researchers have investigated different parts of research papers such as abstract, introduction, literature review (LR), results and discussions. Alhuqbani, M. N. (2013) examined the rhetorical structures of 40 abstracts across four different disciplines in Arabic: law, linguistics, medicine and police. Both qualitative and quantitative data were analyzed using Bhatia's (1993) four-move structure and Hyland's (2000) five-move structure. The results showed that these four disciplines vary in their adherence to these models. However, abstracts in medicine strictly adhered to either Bhatia's or Hyland's model. Abstracts in law, linguistics and police had no conventional move structure. Regarding the preferred verb tense, researchers used the present tense in the introduction, purpose and conclusion moves, and the past tense in the method and result moves. Likewise, Samraj (2005) compared the generic structure of RA introductions and abstracts in the field of conservation biology and wildlife behavior, which indicated that RA introductions and
abstracts in conservation biology have a greater similarity in function and organization than the same two genres in wildlife behavior.

Bunton (2002) analyzed PhD theses introductions and found that most of all introductions employ Swales’ CARS model although there are several variations, and the moves are cyclical. Bunton also emphasized that PhD introductions are longer than research articles introductions, and there is no significant difference between native and non-native speakers of English, and no major difference between research papers in different disciplines. Moreover, Árvay-Tankó (2004) did a contrastive analysis of two languages regarding research article (RA) introduction and revealed some variations between English and Hungarian introductions. For example, English introductions are much longer and tend to be divided into paragraphs compared to Hungarian ones; and Hungarian writers tend to raise questions instead of indicating a gap. It was also found that the structure of Hungarian introductions are different from that of English introductions.

Futász (2006) then investigated the rhetorical structure of theoretical RA introductions written by undergraduate students in the field of English linguistics and literature and showed that there were discipline-specific variations although Hungarian introductions displayed the elements typically identified in research articles written by native expert writers. Also, Ozturk (2007) explored the degree of variability in the structure of RA introductions within a single discipline through the analysis of 20 research articles. This study revealed a variation between the two sub-disciplines of applied linguistics (L2 acquisition and L2 writing research) regarding CARS model. Some sub-disciplinary variation was also identified. The two sub-disciplines employ different and almost unrelated move structures. For example, in L2 acquisition corpus, one type of move structure was predominant, whereas in L2 writing corpus, two different types of move structure were almost equally frequent.

A study of Geçikli, M. (2013) analyzed the introduction sections of 20 PhD theses in Turkish and English in the field of English Language Teaching (ELT), aiming to explore whether the authors from different academic institutions within the same discourse community perform the same rhetorical strategies in the introduction parts of doctorate dissertations. Qualitative research design was used through a content analysis including both genre and discourse analysis based on CARS model 2004 version. The findings showed that although the dissertations are written by the authors within the same discourse community, English thesis introductions have a more complex rhetorical organization than the introduction parts of Turkish theses.

Monreal et al., (2011) also analyzed the introductions of 20 doctoral theses on computing written in Spanish and in English to ensure whether the theses in the same scientific-technological area but written by authors from different cultural and linguistic backgrounds employed the same rhetorical strategies to introduce the work. The step analysis suggests that introductions in both languages rely mainly on the presentation of background information and the work carried out. However, English introductions tend to focus on the writer’s own work, its originality and its contribution to the field of study. They also present more embedding and overlapping of steps and sub-steps than Spanish texts.

Duszak (1994) also examined language-related RAs written in English and Polish, and found that although the moves established by Swales appear in the RAs, their order is not straightforward. Cyclicality appears in both English and Polish RAs. Another problem is that Polish texts are often not divided into sections. It is therefore difficult to determine where a particular “introduction” ends.

Turner, E., and Bitchener, J. (2008) evaluated the effectiveness of LR writing teaching course in a New Zealand university including 15 hours within six-days. The study found a clear evidence of improvement in all areas. Randolph (2009) then summarized some main information on how to write a good LR of thesis. This study suggests that a thesis
should begin with a discussion of the purposes of a review, present a taxonomy of LRs, and then discuss the steps in conducting a quantitative or qualitative literature review. The study also concludes with a discussion of common mistakes and a framework for the self-evaluation of LR.

Bitchener & Basturkmen (2006) focused on students’ difficulties in writing the discussion of results section (DRS) of the thesis and investigated the extent to which the perceptions of the students match those of their supervisors. The study found that: (1) students had a more limited understanding of DRS function compared to that of their supervisors; (2) a common understanding between the supervisors and the students about the nature and cause of the students’ difficulties was limited; and (3) the students tended to use limited proficiency as a default mode of explanation of their difficulties, whereas three out of the four supervisors offered explanations not related to second language proficiency.

1.2 Problem statement

Identifying valid and reliable ways to evaluate undergraduate students’ research papers writing ability is limited although testing plays a vital role in the educational system as well as in the teaching and learning process. It is actually required in all classrooms and becomes a key predictor of future professional or academic success. It is also important to students due to its impacts on their education and their daily life (William 1996; Brown 1996; White 1994; Sahin 2007). According to Hughes (1989), testing can provide both lecturers and learners with beneficial backwashes as well as harmful backwashes. A good classroom assessment can bring benefits to lecturers in some ways; for example, it helps the lecturers see how well learners have learnt. Moreover, a valid assessment can also help the lecturers to evaluate the effectiveness of the syllabus as well as the methods and materials the lecturers are using. Consequently, the lecturer can select their own teaching methods to make sure that it corresponds to the syllabus, the objective of the curriculum, learners’ needs and learners’ levels. A good classroom assessment also provides a good feedback for learners. As a result, once the task has been scored and evaluated, learners can learn something about their areas of strengths and weaknesses. Additionally, learners have a better chance to assess and self-regulate their learning process. In other words, assessment plays a decisive role in deciding the result of students’ learning process.

However, assessment can lead to harmful backwashes if it lacks validity and reliability. Especially, it is challenging for teachers of English to evaluate students’ assignments in a reliable way, especially students’ academic writing ability (Alderson et al., 1995) because it needs to take into considerations task variables, test taker variables, rater variables, and rating scales (Bachman & Palmer 1996). Therefore, to avoid the subjectivity and a risk of reliability of assessment, selecting the most suitable method of scoring students’ assignments including English written assignments is necessary. So far, a number of methods of scoring have been presented in literature (e.g., Hamp-Lyons, 1991; Shohamy, 1995) to improve both accuracy and consistency (Brown, 1996; Wiseman, 2012). Nevertheless, there is a need to take into account the rating scale. According to Park (2004), to assess writing ability, scoring rubrics should be used, namely analytic, holistic and primary trait are mentioned (Bachman & Palmer 1996; Weigle 2002; Alderson, 1995). In addition to the use of such scoring rubrics as scoring guilds to evaluate writing ability, the impression method, and error count method are used (Heaton, 1974; Cohen, 1994) for scoring academic writing texts in different genres at various levels. Yet, Becker (2010/2011) revealed that U.S. universities usually adopted an existing scale, and very few educational institutions designed their own scoring rubrics.
1.3 Research objectives
For the above reasons, this paper first aims to propose a valid and reliable scoring rubric to assess research papers writing ability of undergraduate students who major in English. Then, it aims to examine if there is a difference in the effectiveness of research paper writing program using the proposed Process Genre Model on undergraduate students’ overall research paper writing ability in terms of content, organization, language use and mechanics use skills between the control group and experimental group.

1.4 Research questions
Based on the above objectives, the research questions of this study are addressed as follows:
Q1. Which core components of English academic writing should be proposed in the scoring rubric of assessing research papers writing ability of English majored students of Can Tho University, Vietnam?
Q2. Is there any difference in EFL/ESL undergraduate university students’ overall research paper writing ability in terms of content, organization, language use and mechanic use between the control group and experimental group?

2. Literature review

2.1 Research paper and its components
English research paper writing, which emerged in the 14th century, is generally defined as scientific writing, primarily focusing on writing for publication, which provides new scientific findings of any research work. This genre is also viewed as the transmission of a clear signal to a recipient. Thus, the meaning of words in research papers must be clear, simple and well-ordered. Metaphors, similies, idomatic expressions should not be used in all research papers (Day & Gastel, 2012). Likewise, Swales (1990) viewed research paper writing as a genre “complexly distanced reconstructions of research activities,” rather than “simple narratives of investigations” (p.175). The Introduction, Methods, Results, and Discussion (IMRAD) format is a standard format (e.g., Day & Gastel, 1998; Pyrczak & Bruce, 2005; Glasman, 2010; Swales & Feak, 2004; Day & Gastel, 2012), providing a clear outline for editors, referees, and ultimately readers. However, it is not the only format for all research papers because the structure of research papers may vary within different disciplines (Swales & Feak. 2004, p.222; Day & Gastel, 2012).

2.2 Process Genre Model of research paper writing
Adapted from writing model of previous researchers (e.g., Badger & White, 2000; Yan, 2005; Nordin & Mohammad, 2006) views, Tuyen (2016 et al.,) proposed the Process Genre Model of research paper writing including the three main stages below:

Prewriting/Planning stage: includes the following activities and strategies:

Text analysis: refer to the analysis of writing samples to provide input for students in order to raise their awareness of the schematic (or generic) structure, discourse structure, linguistic conventions of a particular genre regarding the purpose and audience. This activity also helps select the best ideas of a piece of written work.

Noting down: involve the process of noting down keys words, collocations and core ideas, etc.
Generating: include activities which help writers determine what they are going to write about. It is effective if there is a provision of a practical purpose for discussion to help learners share their experiences, leading to the enhancement of motivation for each write.

Reorganizing: consist of activities which help writers to identify priorities in what they have to say and help writers give emphasis to the most important parts of their arguments to make sure that what is being written about is relevant to the potential readers.

Ordering: refer to a review of the way in which writers organize their texts to effectively communicate with potential readers.

While-writing stage: Translating/drafting/ writing practice: includes activities and strategies such as practicing writing in groups, in pairs or individually. A set of writing tasks (i.e., controlled writing and free writing tasks) is provided. The writers then translate plans and ideas into the provisional text.

Post-writing stage: includes the following activities and strategies.

Self-revising and Editing: Revising refers to reading back over the text, helping maintain an overall coherence of the text. Good writers should focus on getting the content right first and then correct language spelling, punctuation and grammar. The checklist of guidelines for writers to edit their work is provided. Students practice in pairs in this activity.

Feedback: includes peer feedback and teacher feedback. Peer feedback aims to provide input and authentic audience. Teacher feedback helps reduce mistakes or errors on content and organization of the text, word choice, language use and mechanics use.

Rewriting/ redrafting: After getting feedback, writers make an effort to rewrite based on the provided feedback.

Self-Evaluating: Writers learn how to evaluate their writing based on the checklist provided for assessing their writing. Students can work in pairs in this activity.

Publishing: involves teaching students know how to determine and select potential journals or conferences to submit their research papers.

2.3. What is an effective scoring rubric of assessing English academic writing ability?

Scoring rubrics are descriptive scoring schemes developed by teachers or other evaluators to guide the analysis of the products or processes of students' efforts (Brookhart, 1999). Scoring rubrics are typically employed when a judgement of quality is required and may be used to evaluate a wide range of subjects and activities. One common use of scoring rubrics is to guide the evaluation of writing samples. Judgements concerning the quality of a given writing sample may vary depending upon the criteria established by each evaluator. One evaluator may heavily weigh the evaluation process upon the linguistic structure, while another evaluator may be more interested in the persuasiveness of the argument. A high quality essay is likely to have a combination of these and other factors. By developing a predefined scheme for the evaluation process, the subjectivity involved in evaluating an essay becomes more objective. Schafer (2004) viewed rubrics as tools that are used to assess the quality of student work in a range of excellent to poor performances. A rubric has a criteria that corresponds to a scale of possible points to score spoken or written performances. The highest point refers to the best performance, whereas the lowest one refers to the worst performance on the scale. Various levels of proficiency are included in the scale. It can be
generic enough to be used with various types of writing. Four common scoring types which use rubrics include holistic scoring, analytical scoring, weighted trait scoring, and primary trait scoring (Campbell et al., 2000). Herman, Aschbacher, and Winters (1992) proposed four characteristic features of a rubric such as criteria, standards, scale, and examples. An effective rubric has a clear criteria for the test-takers to know what is expected and for the raters to assess the answers. It also has a good standard for various levels of performance, and has a valid and reliable scale to meet the standard of writing performance. It is possible to include an example of expected performance at the different levels on the scale.

Types of assessment rubrics

2.4.1 Advantages and disadvantages of holistic scoring rubrics

Relevant literature refers to the important role of three basic assessment rubrics for evaluation, namely primary trait analytic, and holistic scoring rubrics in the evaluation of written proficiency (e.g., Cumming, 1990; Weigle, 2002; East & Young, 2007). There is a difference in impact, discriminatory power, inter-rater reliability, the degree of bias, and the cost-effectiveness in terms of time, effort and money among the three basic assessment rubrics (Kuo, 2007). Holistic scoring takes into account the general response of written work and provides an overall score to the performance (White, 1994; Weigle, 2002; Hyland, 2002). According to Park, T. (2004), holistic scoring is economical compared to analytic scoring because a single score is used to grade writing work. A single score gives useful ranking information but no details. Therefore, holistic scoring is widely used to assess writing in large-scale due to its general guideline that defines good performance at each score point. According to White (1994) and Cohen (1994), holistic scoring has more advantages than disadvantages. In addition to its advantages, holistic scoring rubric has its drawbacks. For example, diagnostic information about students’ writing cannot be provided because specific aspects of writing (e.g. organization, grammar, vocabulary etc.) are not provided. Moreover, it lacks reliability because it scores the text generally (Ferris & Hedgcock, 1998; Park, T., 2004).

2.4.2 Advantages and disadvantages of analytic scoring rubrics

Using analytic rubrics to assess students’ writing may lead to the improvements in learners’ writing skills (Anderson, 1995) and critical thinking subskills of writing (Becker, 2011). In this regard, writing components (i.e., unity, coherence, flow of ideas, formality level and so on) are analyzed. Based on this type of scoring rubric, aspects of writing such as organization, ideas, spelling and so on are graded separately (Cohen, 1994). According to Park, T. (2004), writing components such as content, organization, cohesion, register, vocabulary, grammar, or mechanics are graded. In this sense, more detailed information about a test taker’s performance in different aspects of writing is provided. Unlike holistic scoring, analytic rubrics provide more useful diagnostic information about students’ writing abilities including information about the strengths and weaknesses of students. Therefore, instructors and curriculum developers could give the instruction to meet the students’ needs. It is stated that L2 learners get more benefits from analytic scoring, especially for some learners who may perform content and organization of writing well, but may have more errors on grammars; others may perform sentence structure well, but may not good at organize their writing coherently. Likewise, less inexperienced raters find it easier to grade students’ writing based on analytic scale. Finally, a clear analytic scoring with clear criteria helps writers get consistent and direct feedback. However, it has several disadvantages. It is impossible to evaluate a piece of good writing based on 3 or 4 criteria. Each scale may not be
used separately. It is difficult to grade writing performance effectively based on the description for each scale (e.g. what does ‘adequate organization’ mean?’). Indeed, if the ‘idea’ scale has high scores, other scales are influenced.

2.4.3 Advantages and disadvantages of primary trait scoring

Primary trait scoring was developed for scoring essay writing performance (Lloyd-Jones, 1977). Primary trait scoring, known as holistic scoring, is less common than other methods (Becker, 2011). It is similar to holistic scoring; however, it is required to focus on individual characteristics of writing task. It deals with the core features of particular genre of writing; for instance, by considering differences between several types of essays (Weigle, 2002). According to Lloyd-Jones (1977), like analytic scales, primary trait scoring yields a quite reliable score thanks to sufficient training which is provided for raters. Primary trait scoring guides focus on the rater's attention to the features of a piece of writing which are relevant to the kind of discourse (i.e., audience, speaker role, purpose, and subject). A unique quality of primary trait scoring is that scoring guides are constructed for a particular writing task set in a full rhetorical context. However, it also has its drawbacks due to this characteristics. For example, it will ignore other mistakes or errors on other aspects of writing. Moreover, it takes a longer time to design because the scoring guilds are long. Perkins. K (1983) states that it has not been widely used in the classroom due to a lack of feedback of students' writing.

3. Methods

3.1. Research Design

In this study, regarding research question 1, Delphi Technique (DT) was first conducted within two rounds to validate scoring rubric for evaluating research paper writing ability of EFL/ESL undergraduate students. To save time, a questionnaire with structured questions was used through face to face discussions to survey two groups of experts’ opinions about the criteria description. The questionnaire was developed by researcher based on literature on scoring rubrics of academic writing ability. The questionnaire includes necessary criteria of evaluating research paper writing and thus categorized into 4 structures.

- Structure 1 involves evaluation criteria related to the content of research paper writing.
- Structure 2 involves evaluation criteria related to the organization of research paper writing.
- Structure 3 involves evaluation criteria related to the language use of research paper writing.
- Structure 4 involves evaluation criteria related to the mechanics use.

The scoring rubric of evaluating research paper writing ability through the questionnaire was then introduced to the expert jury. These experts were asked to evaluate and validate it based on the proposed questionnaire. They were also encouraged to make any addition, omission, corrections, and change if necessary. After all their comments and suggestions were consolidated, the scoring rubric was adjusted by the researcher and then sent to the expert jury again to confirm the final scoring rubric of evaluating research paper writing.

Regarding research question 2, the non-equivalent control group experimental design with pretest and posttest was used as a research method for data collection in order to (1) examine if there is a significant difference in the effectiveness of the research paper writing program on undergraduate students’ overall research paper writing ability in terms of the content, organization, language use and mechanics use skills between the control group and experimental group.
3.2 Sample and Sampling Methods

Different methods have been used to select a sample from a particular population. For example, probability sampling is one of the methods that is classified into five categories (Zikmund, 2000), namely the simple random technique, the systematic technique, the stratified technique, and the cluster and multistage techniques. However, this study used the simple random technique supported by Gay and Airaisian (2002) due to the representative of the larger group. Several factors (i.e., research types, research hypotheses, financial constraints, variables, data collection methods and accuracy level etc.,) should be taken into account when determining the sample size of the study. The correlational research should have a minimum of 30 subjects, and in research comparing groups, there should be at least 15 subjects in each group (McMillan & Schumacher, 1993).

In this study, scoring rubric was designed to evaluate research paper writing ability of EFL/ESL undergraduate students, who were the third year university students in academic year (2016-2017), majoring in English. They are studying in English Department, School of Education, Can Tho University (CTU), a multidisciplinary university, which is the biggest public university in the Southeast of Vietnam. It has a mission to develop scientific research projects and get an access to scientific and technological knowledge for problematic solutions to science, technology, economics, culture and society in the region. These students are taught four skills of English as their major subjects (i.e., speaking, listening, writing, and reading). They are compulsory to study genres of academic writing (i.e., sentences writing, letters, paragraphs writing, essays writing etc.) in the first and second academic year. In the third academic year, these students are encouraged to attend a course of research paper writing to prepare for their thesis writing in the final year of their undergraduate program.

Regarding research question 2, the participants were divided into two groups: one experimental group (20 students) and one control group (45 students). Two groups were taught by the same lecturer during one semester of 90 meetings (theory, practice and assessment). The experimental group were required to receive a program of research paper writing based on the proposed PGM of research paper writing in the current study. The control group were required to receive the existing program of the university where this study were conducted based on the product approach.

3.3 Research Variables

In this study, there is a dependent variable: (1) research paper writing ability in terms of content, organization, language use and mechanics use skills. Independent variable is research paper writing program designed based on the proposed Process Genre Model (Tuyen et al., 2016).

3.4 Pre and post-tests

3.4.1 Test planning and designing

To establish the test’s validity and reliability, Hughes (1989) suggested that the test should be designed based on what its purpose is, who the test aims at, what content is to be covered, what methods are to be used, how it is marked, and then a detail set of specifications for each section of the test set out. Bachman (1990) also stated that what students have learnt should be tested, and the test tasks should be appropriate to their target language use situations.
For such reasons, in the present study, for pretest, writing a reflection in the form of a paragraph of 200-400 words based on a provided research article will be used for quantitatively measuring academic writing ability of students in both groups. Meanwhile, for posttest, writing a research article based on the thesis provided including: (a) Abstract; (2) Introduction; (3) Literature review (4) Methods; (4) Results; (5) Discussions and (6) Conclusion were required for both groups. Topics for pretest and posttest are presented in Table 3.1 below:

**Table 3.1**
Topics for Pre-test and Post-test

<table>
<thead>
<tr>
<th>Topic for Pre-test</th>
<th>Topic for Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing a reflection in the form of a paragraph of 200-400 words based on a</td>
<td>Based on the thesis provided, writing a research article including the following components:</td>
</tr>
<tr>
<td>provided research article entitled: ‘EFL teaching and learning at a Vietnamese</td>
<td>1) Abstract of 150-300 words including:</td>
</tr>
<tr>
<td>University’ by Hong Thi Nguyen.</td>
<td>(a) Background; (2) Aim; (3) Method;</td>
</tr>
<tr>
<td></td>
<td>(4) Results; (5) Conclusion.</td>
</tr>
<tr>
<td></td>
<td>2) Introduction of 800-1000 words based on CARS model</td>
</tr>
<tr>
<td></td>
<td>3) Literature review of 7000-1200 words.</td>
</tr>
<tr>
<td></td>
<td>4) Methods of 300-500 words including describing research design, participants,</td>
</tr>
<tr>
<td></td>
<td>data collection instruments.</td>
</tr>
<tr>
<td></td>
<td>5) Results and Discussions 500-1000 words</td>
</tr>
<tr>
<td></td>
<td>6) Conclusions</td>
</tr>
</tbody>
</table>

3.4.2 Validity and reliability of tests

In this study, to avoid researchers’ subjectivity and a risk of reliability, five lecturers of TESOL, where this study were conducted were invited to view the topics for pretest and posttests. They all agreed that the topic is relevant to students’ level and the purpose of the test.

3.4.3 Test administration

According to Hughes (1989), the test gives reliable and valid results if it is well-administered. In this sense, in this present work, the pretest was administered to both groups at the beginning of the study. Pretest was administered in the quiet, large and bright classroom within 60 minutes. The candidates were allowed to ask if there is something they would like to make clear about the topic. When the time was up, the candidates were asked to stop writing, and their papers were collected. The posttest were administered after the treatment program in the same conditions as the pre-test, but the allotted time to complete the writing up of research article is 120 minutes.

3.4.4 Methods of scoring academic writing ability

From the obtained data, a valid and reliable valid scoring rubric was proposed to assess research papers writing ability of EFL/ESL undergraduate students. Scoring rubric has also been known as a marking scheme or marking guide. In this study, analytic method, including an attempt to separate various aspects of a composition was used. Specifically, analytic scale adapted from Carroll and West (1989, as cited in Tribble, 1996) was proposed to grade research paper writing ability. According to this marking scale, writing components, namely content (relevant ideas), organization (structure and coherence), language use
(vocabulary and grammar choice), and mechanics use (punctuation and spelling based on APA style) are graded separately.

Based on the evaluation system of Can Tho University, the total point of all subjects is 10.0 points. Thus, in this paper, the total point of research writing ability is 10.0 points including 4.0 points for content, 2.0 points for organization, 2.0 points for language use, and 2.0 points for mechanics use. The score is converted into 4 levels. Excellent to good level includes Level A and level B+. Level A is in the score range of 9-10 points. Level B+ is in the score range of 8.0-8.9 points. Fair level includes level B and level C+. Level B is in the score range of 7.0-7.9 points. Level C+ is in the score range of 6.5-6.9 points. Average include level C and level D+. Level C is in the score range of 5.5-6.4 points. Level D+ is in the score range of 4.0-5.4 points. Poor level includes level D and level F. Level D is in the score range of 4.0-4.9 points. Level F is below point 4.0. Detailed scoring rubric of assessing research paper writing ability is presented in Table below.

### Table 3.1
Scoring rubrics of assessing research paper writing

<table>
<thead>
<tr>
<th>Writing components</th>
<th>Score value</th>
<th>Level/ Description of criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-10 (A)</td>
<td></td>
<td>EXCELENT TO GOOD: Excellent to very good treatment of the subject • Content relevant to the topic • Balanced presentation of relevant and legitimate information that clearly supports a central purpose or argument and shows a thoughtful, in-depth analysis of a significant topic • Readers gain important insights.</td>
</tr>
<tr>
<td>8.0-8.9 (B+)</td>
<td></td>
<td>FAIR: Adequate treatment of the topic • Most content relevant to the topic • Information provides reasonable support for a central purpose or argument and displays evidence of a basic analysis of a significant topic • Readers gain some insights.</td>
</tr>
<tr>
<td>7.0-7.9 (B)</td>
<td></td>
<td>AVERAGE: Treatment of the topic is hardly adequate • Some irrelevant content, relevant to the topic • Analysis is basic or general • Readers gain few insights.</td>
</tr>
<tr>
<td>6.5-6.9 (C+)</td>
<td></td>
<td>POOR: Inadequate treatment of the topic • almost no useful detail • Reader is confused or may be misinformed • Central purpose or argument is not clearly identified • Analysis is vague or not evident.</td>
</tr>
<tr>
<td>5.5-6.4 (C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.0-5.4 (D+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.0-4.9 (D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;4 (F)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Organization</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-10 (A)</td>
<td></td>
<td>EXCELENT TO GOOD: Fluent expression • ideas clearly stated/supported • Appropriately organized paragraphs or sections • Logical sequenced (coherence) • Connectives appropriately used (cohesion).</td>
</tr>
<tr>
<td>8.0-8.9 (B+)</td>
<td></td>
<td>FAIR: Uneven expression • but main ideas stand out • Paraphrasing or section organization evident • Logical sequenced (coherence) • Connectives appropriately used (cohesion).</td>
</tr>
<tr>
<td>7.0-7.9 (B)</td>
<td></td>
<td>AVERAGE: Very uneven expression • Ideas difficult to follow stand out • Paraphrasing or section organization does not help the reader, logical</td>
</tr>
<tr>
<td>6.5-6.9 (C+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.5-6.4 (C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.0-5.4 (D+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;4 (F)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language use (Vocabulary+ grammar)</td>
<td>EXCELENT TO GOOD: Accurate word and usage• Appropriate selection to match register• Confident handing of appropriate structures • Hardly any errors of agreement, tense, number, word order/function, articles, pronouns, prepositions, meaning never obscured.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAIR:• Occasional mistakes in word choice and usage• Register not always appropriate• acceptable grammar- but problems with more complex structures•Mostly appropriate structures• some errors of agreement, tense, number, word order/function, articles, pronouns, prepositions, meaning some times obscured</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AVARAGE: A noticeable number of mistakes in word choice and usage• Register not always appropriate• insufficient range of structures with control only shown in simple constructions• Mostly appropriate structures•Frequent appropriate structures• Some errors of agreement, tense, number, word order/function, articles, pronouns, prepositions, meaning some times obscured</td>
<td></td>
</tr>
<tr>
<td></td>
<td>POOR: Uncomfortably frequent mistakes in word/idiom choice and usage• Register , no appropriate sense of register• Major problems with structures- even simple ones• frequent errors of negation agreement, tense, number, word order/function, articles, pronouns, prepositions, meaning often obscured</td>
<td></td>
</tr>
</tbody>
</table>

| 4.0-4.9 (D) <4 (F) | sequence difficult to follow (coherence)•Connectives largely absent (cohesion). |
|语言使用 | **EXCELENT TO GOOD:** 准确的词汇和用词• 合适的选择以匹配语体• 自信的处理合适的结构• 几乎没有同意、时态、数量、词序/功能、文章、代词、介词的错误。 |
| | **FAIR:** 偶尔的词汇选择和用词错误• 语体不总是合适的• 可接受的语法，但更复杂结构有错误• 大部分合适的结构• 一些同意、时态、数量、词序/功能、文章、代词、介词的错误偶尔被遮蔽。 |
| | **AVERAGE:** 显著的词汇和用词错误• 语体不总是合适的• 结构范围不足，只在简单结构中显示控制• 大部分合适的结构• 常见的合适的结构• 一些同意、时态、数量、词序/功能、文章、代词、介词的错误一些时候被遮蔽。 |
| | **POOR:** 不舒适频繁的词汇选择和用词错误• 语体，没有合适的语体感受• 重大结构问题，即使是简单的• 经常的否定同意、时态、数量、词序/功能、文章、代词、介词的错误经常被遮蔽。 |

| 4.0-4.9 (D) <4 (F) | 意义难以理解（连贯性）• 连接词几乎不使用（连贯性）。 |
| **POOR:** 缺乏流畅的表达，想法非常难以理解• 几乎没有归纳/组织的概念• 没有逻辑序列（连贯性）• 连接词不被使用（连贯性）。 |

| 9-10 (A) 8.0-8.9 (B+) | 准确的词汇和用词• 合适的选择以匹配语体• 自信的处理合适的结构• 几乎没有同意、时态、数量、词序/功能、文章、代词、介词的错误。 |
| 7.0-7.9 (B) 6.5-6.9 (C+) | 偶尔的词汇选择和用词错误• 语体不总是合适的• 可接受的语法，但更复杂结构有错误• 大部分合适的结构• 一些同意、时态、数量、词序/功能、文章、代词、介词的错误偶尔被遮蔽。 |
| 5.5-6.4 (C) 5.0-5.4 (D+) | 显著的词汇和用词错误• 语体不总是合适的• 结构范围不足，只在简单结构中显示控制• 大部分合适的结构• 常见的合适的结构• 一些同意、时态、数量、词序/功能、文章、代词、介词的错误一些时候被遮蔽。 |

| 4.0-4.9 (D) <4 (F) | 意义难以理解（连贯性）• 连接词几乎不使用（连贯性）。 |
| **POOR:** 缺乏流畅的表达，想法非常难以理解• 几乎没有归纳/组织的概念• 没有逻辑序列（连贯性）• 连接词不被使用（连贯性）。 |

| 9-10 (A) 8.0-8.9 (B+) | 准确的词汇和用词• 合适的选择以匹配语体• 自信的处理合适的结构• 几乎没有同意、时态、数量、词序/功能、文章、代词、介词的错误。 |
| 7.0-7.9 (B) 6.5-6.9 (C+) | 偶尔的词汇选择和用词错误• 语体不总是合适的• 可接受的语法，但更复杂结构有错误• 大部分合适的结构• 一些同意、时态、数量、词序/功能、文章、代词、介词的错误偶尔被遮蔽。 |
| 5.5-6.4 (C) 5.0-5.4 (D+) | 显著的词汇和用词错误• 语体不总是合适的• 结构范围不足，只在简单结构中显示控制• 大部分合适的结构• 常见的合适的结构• 一些同意、时态、数量、词序/功能、文章、代词、介词的错误一些时候被遮蔽。 |

| 4.0-4.9 (D) <4 (F) | 意义难以理解（连贯性）• 连接词几乎不使用（连贯性）。 |
| **POOR:** 缺乏流畅的表达，想法非常难以理解• 几乎没有归纳/组织的概念• 没有逻辑序列（连贯性）• 连接词不被使用（连贯性）。 |
3.4.5 Test raters

One of the most effective ways of dealing with inconsistency in marking is through the selection of raters. Heaton (1974) stated that if an assessment is based on several judgments, the result is far more reliable than a mark based on a single paper judgment. For this research, raters for grading the participant’s tests on writing will be 2 well-trained and qualified lecturers of TESOL in Can Tho University. However; in some cases, where there is a wide discrepancy in the marks allocated, the script will be examined one again by all three makers, and each mark will be discussed until an agreement is reached (Heaton, 1974).

3. Findings and Discussions

Students’ research paper writing ability in terms of content, organization, vocabulary, language (grammar)

The results from the Independent Samples T-Test showed that research paper writing ability in terms of content, organization, language and mechanics of the group treated with the process genre instruction significantly outperformed that of the control group. There was no significant difference in the means of pretest scores on content \((p = .254)\), organization \((p = .838)\), language \((p = .795)\), and mechanics \((p = .385)\) between the two groups. Meanwhile, there was a statistically significant difference in the means of posttest scores on content \((p = .000)\), organization \((p = .002)\), language \((p = .002)\), and mechanics \((p = .000)\) between the two groups. The results are presented in Table 4.1.2 and Table 4.1.3.

Table 4.1.2
Descriptive Statistics and Independent Sample T-Test Statistics of Pretest Scores on Content, Organization, Vocabulary and Grammar of the Two Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Content</th>
<th>Organization</th>
<th>Language</th>
<th>Mechanics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>P</td>
<td>M</td>
<td>P</td>
</tr>
<tr>
<td>Con</td>
<td>4.750</td>
<td>.254</td>
<td>4.050</td>
<td>.838</td>
</tr>
<tr>
<td>Ex</td>
<td>4.866</td>
<td>.066</td>
<td>7.066</td>
<td>3.977</td>
</tr>
</tbody>
</table>
The results from pretest and posttest on research paper writing in English indicated that the process genre approach to research paper writing instruction significantly improved university students’ research paper writing ability in terms of content, organization, language and mechanics. These findings are consistent with earlier findings of Badger and White’s (2000), Kim and Kim’s (2005); Chow’s (2007), Anik and Nihayah’s (2009); Babalola’s (2012) studies that the implementation of the process genre approach in teaching writing positively effected students’ overall writing ability. Moreover, Pujianto (2014) mentioned that PGA helps students develop writing skills of report text. It is clear that in the experimental group of this study, students were aware of the purpose and audience of the text, identify certain conventions of writing each sub-components of research papers including schematic structures and language.

The finding is also in line with the study of Handayani and Siregar (2013) that through the process genre approach in the classroom, the students understand the content of the text. They are also active and enthusiastic during the teaching learning process. The result of the research showed that the application of the process genre approach significantly improved students’ achievement in writing descriptive text. The finding is agreed with Agibuay (2016) that there is a marked improvement in the descriptive writing compositions of students through the process genre approach. Further, the results from the paired t-test showed a significant difference between the pretest and the posttest.

The finding is also consistent with Shariatmadari (2016) that the participants in the experimental group outperformed the control group and in line with Agesta and Cahyono (2017) that there is a significant effect of the process genre approach on the students’ writing achievement in terms of organization, vocabulary, grammar, and mechanics, but no significant difference in the writing ability of the extrovert and introvert students who were taught based on the process genre approach. Noticeably, the results of this study are not consistent with those of Herawati & YayukWidyastuti (2015) that there was no significant difference between the experimental and the control group in writing an argumentative essay.

5. Limitations, Recommendations, implications and Conclusions

From the results obtained from Delphi technique, this study is limited to propose the scoring rubric perceived as essential for EFL/ESL lecturers and examiners to evaluate research paper writing ability of EFL/ESL undergraduate students, who major in English and study four skills of English: listening, speaking, reading, writing. Especially, they have already learnt basic academic writing skills and academic writing language in the previous courses. This scoring rubric is, therefore, not generalized to the assessment of research paper writing ability of EFL/ESL undergraduate students in all fields. However, it is believed that this paper will significantly contribute to the practical applications for examiners, lecturers,
undergraduate and postgraduate students in EFL/ESL contexts to meet the requirements of their own educational institutions. Noticeably, it is suggested that Higher Education institutions in general and lecturers in specific should adjust or modify the description of each criteria of the scoring rubric and adapt a model of research paper writing or conduct further research with different research method and instruments to determine more positive and negative points of the process genre approach.

Acknowledgements

I am grateful to my main supervisor and co-supervisors for their research knowledge support, and the sponsors of the Mekong 1000 Project for their financial support.

References


