Generic structure and citation functions in introductions of biological science articles in English-medium international journals

Sujata S. Kathpalia & Christopher S. G. Khoo
Nanyang Technological University (Singapore)
MSSKATH@ntu.edu.sg & chriskhoo@pmail.ntu.edu.sg

Abstract

Although studies on the rhetorical structure of research article introductions and citation analysis are independently available in the literature, there are none that focus on the functions of citations in the context of generic moves in these introductions. To address this gap, the present study analyzes the generic structure as well as citation functions in introductions of biological science articles in English-medium international journals. Specifically, the aim is to examine the relationship between generic moves and citation functions in individual moves. The findings on the relationship between moves and citations indicate that there is a preference for certain types of citations in specific parts of article introductions. This study will not only enable biological science graduate students to better understand research article introductions but also facilitate the writing of these in their own field.

Keywords: article introductions, genre analysis, citation analysis.

Resumen

Estructura genérica y funciones de la citación en las introducciones de artículos de Biología en revistas internacionales en inglés

Si bien existen algunos estudios sobre la estructura retórica de las introducciones de los artículos de investigación y sobre la citación, no se tiene constancia de la existencia de trabajos centrados en las funciones de la citación en relación con los movimientos genéricos en tales artículos. Por ello, el presente estudio analiza la estructura genérica y las funciones de la citación en artículos de Biología

publicados en revistas internacionales en inglés. Concretamente, el objetivo es examinar la relación entre los movimientos genéricos y las funciones de la citación en movimientos independientes. Los resultados obtenidos muestran que existe una preferencia por ciertos tipos de citaciones en determinadas partes de las introducciones de los artículos. Este trabajo no solo permitirá a los estudiantes de posgrado de Biología entender mejor las introducciones de los artículos de investigación, sino que también facilitará su redacción en su propio campo.

Palabras clave: introducciones de artículos, análisis de género, análisis de la citación.

1. Introduction

In today's academic world, graduate students need to learn how to write about their research in English-medium international journals, since sharing their research with the international research community is as important as conducting research. By publishing their research in the form of academic articles, they enter an ongoing global conversation with other scholars in their field. As they are new to this academic genre, a good starting point for graduate students is learning to write article introductions, particularly because they are challenging in terms of the roles they serve as opening sections of articles. These introductions not only perform the multiple roles of providing topic information, engaging the readers and raising the credibility of the article through citations from works of other scholars (Swales, 1981, 1990, 2004) but their generic and citation conventions may vary across different disciplines (Hyland, 2000; Samraj, 2005; Charles, 2006; Hu & Wang, 2014), cultures (Belcher, 1994; Bloch & Chi, 1995; Pennycook, 1996; Rowley-Jolivet & Carter-Thomas, 2014; Dontcheva-Navratilova, 2015, 2016), languages (Hirano, 2009; Loi 2010) and contexts (i.e. national or international) of publications (Hewings, Lillis, & Vladimirou, 2010; Lillis et al., 2010). To help graduate students with these aspects of research writing, this article examines the rhetorical structure and citation conventions of article introductions in biological science articles.

Many studies have been conducted on the rhetorical structure of article introductions, different sections of research articles or entire articles in different disciplines. A majority of these studies have used Swales' (1981) move-structure pattern, which was later refined and renamed the *Create a Research Space* (CARS) model (1990, 2004). Using this model as a starting point

for the rhetorical analysis of moves in article introductions, studies have either focused on specific disciplines such as agricultural and forestry engineering/science (Hopkins, 1985; del Saz Rubio, 2011; Joseph, Lim, & Nor, 2014), computer science/software engineering (Hughes, 1989; Anthony, 1999), education psychology (Swales & Najjar, 1987), electronics (Cooper, 1985), and social science disciplines (Lewin, Fine, & Young, 2001; Ozturk, 2007; Jalilifar, 2012) or a range of disciplines/sub-disciplines for a comparative analysis (Swales, 1981; Crookes, 1986; Samraj, 2002, 2005; Habibi, 2008; Kanoksilapatham, 2012). Yet other researchers extended Swales' generic framework to the analysis of entire research articles including biochemistry research (Thompson, 1993; Kanoksilapatham, 2005), computer science (Posteguillo, 1999), linguistics/applied linguistics (Yang & Allison, 2003), engineering (Peng, 1987; Kanoksilapatham, 2015; Maswana, Kanamaru, & Tajino, 2015) and medical science (Nwogu, 1997) articles. The present study develops Kanoksilapatham's (2005) analysis of biochemistry articles by applying it to introductions of articles from various biological science journals.

Another area relevant to this study is the analysis of citations in article introductions. Citations have been analyzed in published and student writing from different perspectives in the fields of gender studies (Petrić, 2007), information science (White, 2004), linguistics/applied linguistics (Swales, 1986, 1990; Hyland, 1999; Thompson, 2000; Dontcheva-Navratilova, 2015, 2016), science (Mansourizadeh & Ahmad, 2011), and sociology of language (Bazerman, 1988), as well as across different fields (Hyland, 1999; Harwood, 2009; Hu & Wang, 2014). Although different methodologies have been used for citation analysis by scholars in these fields, they agree that the role of citations in academic writing is, on the one hand, to acknowledge the work of colleagues, and, on the other, to raise the credibility of their own claims. At a general level, citation analysis can be based on the rhetorical functions of citations in a text (Moravcsik & Murugesan, 1975; Petrić, 2007), formal criteria that focus on the linguistic realization of citations (Swales, 1990) or a combination of both (Thompson & Tribble, 2001). Other distinctions involve analyzing rhetorical functions of citations based on researchers' intuitions (Petrić, 2007) versus citers' motivations (Harwood, 2009), using different methodologies such as providing authors a list of pre-selected citation categories (Brooks, 1985; Shadish et al., 1995) or eliciting citation functions from authors in relation to their published articles (Wang & White, 1999; Harwood, 2009). More specifically, in the field of applied linguistics,

citation practice has been investigated from the perspective of cultural differences (Belcher, 1994; Bloch & Chi, 1995; Pennycook, 1996, Dontcheva-Navratilova, 2015, 2016), disciplinary variation (Hyland, 2000) and differences among novice and expert writers (Mansourizadeh & Ahmad, 2011).

From the review of studies above, it is apparent that there are many studies that focus on the rhetorical structure of research article introductions and citation analysis independently, but there are hardly any that combine these two aspects to examine the functions of citations in the context of generic moves in article introductions. Therefore, the objective of this study is to overcome this research gap by examining both the rhetorical structure and citation functions in introductions of biological science articles, and specifically to determine the relationship between individual moves and citation functions in these introductions. The two research questions investigated are as follows:

- What is the move structure and frequency in the article introductions of biological science articles in this study?
- What are the functional categories of citations in the introductions of biological science articles and their relationship with moves/steps?

The findings on the relationship between moves and citation types will reveal whether there is a preference for certain types of citations in specific parts of article introductions. Pedagogically, this study will specifically benefit graduate students in biological science by familiarizing them with the typical generic and citation conventions of research articles in their field, and also contribute towards the field of English for Specific Academic Purposes in general.

2. Methodology

The data for this study comprise 50 articles extracted from 10 journals in the field of biological science. The journals were selected from the top ten journals with the highest impact factors listed under three categories in Thomson Reuters' Journal Citation Reports: *Biochemistry and Molecular Biology, Cell Biology,* and *Biotechnology and Applied Microbiology*. These three subfields of biological sciences are closely related and as such there is substantial overlap in the journal listing under the three categories. For this study, ten specific

journals were selected from each of the three categories and five research articles were extracted from each of these journals for analysis. The journals included: *Cell, Genome Research, Molecular Psychiatry, Molecular Systems Biology, Nature Medicine, Nature Biotechnology, Metabolic Engineering, Genome Biology, Cancer Cell,* and *Cell Stem Cell*. Survey journals (often with titles beginning with "Annual Review of", "Advances in", "Reviews in" and "Trends in") and survey articles were excluded from this study. Care was also taken to include some journals from publishers other than Nature Publishing Group and Cell Press, two very prominent publishers in biological sciences.

The overall rhetorical structure of the research articles was dependent on the journals, with articles in two of the journals following the conventional IMRAD (Introduction, Methods, Results and Discussion) structure and in 8 of the journals following a modified IRAD-M (Introduction, Results, Discussion and Method) structure. The article introductions in the corpus varied in length ranging from 191 words to 1535 words. The articles were numbered and analysed from two perspectives, genre analysis and citation analysis, in order to determine the relationship between moves/steps and citation functions in different parts of the article introductions. The examples used in this article are numbered (e.g. B1) and their sources are provided in Appendix A. The conceptual framework employed for these is presented in the sub-sections below.

2.1. Framework for move analysis

Swales' (1990, 2004) CARS model has been extensively used by researchers to account for the rhetorical structure of article introductions in various fields. According to him, the 3-move model is motivated by the three main functions of article introductions: to convince readers about the significance of the research topic (Move 1), to indicate that there is a niche or gap in past work that needs to be filled (Move 2) and to show how that niche will be occupied through the present study (Move 3). This model is presented in Table 1:

Moves	Steps
Move 1 Establishing a territory	Step 1 Claiming centrality and/or Step 2 Making topic generalization(s) and/or Step 3 Reviewing items of previous research
Move 2 Establishing a niche	Step 1A Counter-claiming or Step 1B Indicating a gap or Step 1C Question-raising or Step 1D Continuing a tradition
Move 3 Occupying the niche	Step 1A Outlining purposes or Step 1B Announcing present research Step 2 Announcing principal findings Step 3 Indicating RA structure

Table 1. Swales' (1990) CARS model for article introductions.

This model has been refined further by several discourse analysts based on the analysis of article introductions in different disciplines. The adaptation that is particularly applicable to the present study is Kanoksilapatham's (2005) model as it was based on the analysis of 60 biochemistry research articles. As biochemistry articles belong to the bioscience discipline, the model was applied to this study's corpus of biological science articles. Although it was used as a starting point for this study, a few modifications were made to represent the rhetorical structure of the biological science articles in the present corpus after an initial analysis of 20 articles. A comparison between Kanoksilapatham's model and the modifications made to it in the present study is presented in Table 2 below and followed by a justification for these changes:

Kanoksilapatham's model	Modifications
Move 1 Announcing the importance of the field Step 1 Claiming the centrality of the topic Step 2 Making topic generalizations Step 3 Reviewing previous research	Move 1 Introducing the field Step 1 Claiming centrality of the topic Step 2 Elaborating on the topic
Move 2 Preparing for the present study Step 1 Indicating a gap Step 2 Raising a question	Move 2 Preparing for the present study Step 1 Indicating a gap Step 2 Raising a question/issue Step 3 Extending the research
Move 3 Introducing the present study Step 1 Stating purpose(s) Step 2 Describing procedures Step 3 Presenting findings	Move 3 Introducing the present study Step 1 Stating purpose(s)/hypotheses Step 2 Describing procedures Step 3 Presenting findings Step 4 Highlighting the value of the study

Table 2. Modifications to Kanoksilapatham's model.

As can be seen from the above tables, Kanoksilapatham modified Swales' (1990) model by renaming the moves and steps, deleting some of the steps (Move 2 -Steps 1A and 1D; Move 3 – Step 1B and Step 3) and adding steps (Move 3- Step 2 Describing procedures). Similarly, Kanoksilapatham's model was altered further to better account for the corpus in this study. The name of Move 1 was changed to 'Introducing the field' as shown in Example 1, to better reflect the dual purpose of the opening move in article introductions – that of persuading readers of the importance of the field (Move1-Step 1) as well as providing them with relevant information on the topic of research (Move1-Step 2):

(1) Move 1 Introducing the field:
Step1**:** Heart failure (HF) is a devastating disease and a major cause of morbidity and mortality worldwide. *Step 2*: HF often follows myocardial infarction (MI) that is usually accompanied by a massive loss of cardiomyocytes (CMs). These CMs cannot be regenerated by the adult mammalian heart and cannot yet be replaced and/or regenerated via cell-based therapies. Unfortunately, transplanting CMs into an infarcted heart yields only transient and marginal benefits (Burridge et al., 2012). (B8)

In addition, Steps 2 and 3 were combined and renamed 'Elaborating on the topic', as the main goal seems to be to provide background information on the topic of the article through both cited and non-cited information, as is evident in (1). Another reason for renaming Step 3 'Reviewing previous research' is that it is misleading as citations are not only found in Move 1 but are spread throughout the introduction (Swales 1990: 140) and this is reflected in the modified version of the CARS model (Swales, 2004).

Proceeding to Move 2-Step 1 'Indicating a gap', its function is to create a research space for the current study in both frameworks as illustrated in (2):

(2) Move 2-Step 1: Indicating a gap
Cardiovascular progenitor cells (CPCs) may offer a promising avenue for cardiac-regenerative therapy ... Unfortunately, although these committed CPCs might be more suitable for cardiac cell therapy in vivo, they have yet to be extensively expanded, thus significantly limiting their applications. (B8)

However, Move 2-Step 3 'Extending the research' was reintroduced from Swales' model as a few distinct examples of this step were identified in the present corpus. Example 3 illustrates how the authors mention their own preliminary research first and the current one as an extension of it:

(3) Move 2-Step 3: Extending the research
Interestingly, our preliminary analysis of GABA+ [gamma-butyric acid] aging trend using the traditional MEGA-PRESS [MEGA-Point REsolved Spectroscopy Sequence] method, showed a similar trend of age-related reduction of GABA levels in patients with schizophrenia.[18] Motivated by these prior observations, our study included an adequately powered sample to examine the schizophrenia diagnosis by age cohort effect on GABA concentration, and utilized a GABA assessment technique that removes the macromolecule contamination. (B27)

Some of the names of the steps in Moves 2 and 3 were also expanded to account for alternative realizations of the steps in the form of questions and/or issues (Examples 4a and 4b, Move 2-Step 2 Raising a question/issue) and statement of purpose and/or hypothesis (Example 5, Move 3-Step 1 Stating purpose(s)/hypotheses):

(4) Move 2-Step 2 Raising a question/issue
4a. Raising a question: However, this raises the question of whether microRNA and protein cell fate determinants act independently or coordinate with each other to determine cell fate. (B6)
4b. Raising an issue: Although several fluorescent probes have been developed to study inflammasome activation *in vitro*[10-12], their use for *in vivo* applications has remained challenging, lacking translation to single-cell resolution at the organ level[10,12,13]. (B44)

(5) Move 3-Step 1 Stating purpose(s)/hypotheses
To address this, we employed MEGA-PRESS with frequency selective pulses at 1.9 and 1.5 p.p.m.[24] to assess macromolecule suppressed GABA in a cohort of participants with schizophrenia and psychiatrically healthy controls. We tested the hypothesis that older participants with schizophrenia would have lower anterior cingulate GABA levels compared with older control participants. We predicted that there would be no significant difference between younger participants with schizophrenia and controls. (B27)

While the remaining steps in Move 3 (Examples 6 and 7) are the same as in Kanoksilapatham's model, Step 4 'Highlighting the value of the study' (Example 8) was added due to the moderately high percentage of this step (46%) in the corpus:

(6) Move 3-Step 2 Describing procedures
We developed and utilized INDIGO [INferring Drug Interactions using

chemo-Genomics and Orthology] to analyze a large compendium of publicly available chemogenomic data in *Escherichia coli* (Nichols et al, 2011) to identify predictive genetic features of antibiotic synergy and antagonism, and subsequently infer novel drug interactions. (B33)

(7) Move 3-Step 3 Presenting findings
By finding orthologs of genes identified by INDIGO to be predictive of drug interactions in *E. coli*, we successfully predicted drug synergy and antagonism in the bacterial pathogens *Mycobacterium tuberculosis* and *Staphylococcus aureus*. (B33)

(8) Move 3-Step 4: Highlighting the value of the study
Hence, the cell fate determinants provide a safeguard mechanism against excessive stem cell proliferation when normal homeostasis is disrupted by inflammation or oncogenic mutation. (B6)

An important point to note about steps is that they can be realized independently or be conflated. In Examples 6 and 7, Steps 2 and 3 in Move 3 are realized independently but they could be conflated, as shown in (9):

(9) Move 3-Steps 2 and 3: Combination of computational analysis and quantitative experiments revealed that the unique regulatory kinetics among miR-34a, Numb, and Notch enables a robust binary switch so that Notch level is steady and insensitive to precise miR-34a level except for a sharp transition region. (B6)

In the quantitative analysis of the corpus, both conflated and independent steps were included in the frequency count of the steps for all three moves. In other words, parts of sentences that served particular functions were counted as steps, even though they were not expressed in the form of independent sentences.

An inter-coder reliability test was carried out by the second author on a sample of 15 articles. The percentage agreement and Cohen's kappa value for each step are given in Table B1 in Appendix B. Cohen's kappa measures the proportion of agreement after removing chance agreement between two coders. Following Cohen (1960), Landis and Koch (1977), and Viera and Garrett (2005), values of 0.21 to 0.40 were considered to represent fair agreement, 0.41 to 0.60 moderate agreement, and 0.61 to 0.80 substantial agreement. While most of the steps have a moderate to substantial agreement, Move 2-Step 2 Raising a question/issue and Move 3-Step 2 Describing procedures show fair agreement. As the differences were due to conflation of steps within and across moves, the issue was resolved by

considering these as embedded steps within sentences/clauses in the coding.

2.2. Framework for citation analysis

The citation analysis in this study is based on functional criteria although some contextual information is also provided on citation types using formal criteria. A combination of resources was employed to arrive at a comprehensive inventory of citations with specific functions. As a starting point, 19 functional types were collated, 8 from Petrić's (2007) text-based citation analysis of high- and low-rated master's thesis of second language writers on gender studies and 11 from Harwood's (2009) interview-based study of computer scientists and sociologists on citation use in their own articles. After taking into consideration the similar, different and overlapping categories, a list of categories was prepared and tested on 20 article introductions in the corpus. The term 'Attribution' is not used as a category label in the present study, as attributing information for the sake of collegiality or to protect oneself is considered to be an overall function of all citations. The citation categories included in the present study and their definitions are provided below:

1. Applying: This citation function is defined according to Petrić (2007) as making connections between the cited authors' and the writers' works in terms of concepts, terminology and methodology. In the present study, this definition is extended to include both adoption of other authors' works as well as adaptation of previous authors' works.
2. Associating: This citation function is the same as Petrić's (2007) 'Establishing links between sources' and it highlights both similarities and differences between or among different citations.
3. Building: The term is borrowed from Harwood (2009) and used in the present study for all citations that contribute towards building knowledge on the topic of the article whether in terms of topic generalizations, elaborations or exemplifications. This term also encompasses what Petrić (2007) refers to as 'Attribution' and 'Exemplification' in her study, with the former defined as information related to terminology, propositions or research activity and the latter as an example to illustrate the writer's statement.
4. Evaluating: This refers to Petrić's (2007) citation function of 'Evaluation' that includes both negative and positive evaluation of other authors' works. Harwood (2009) uses the terms 'Credit' (along with the use of evaluative adjectives) and 'Engaging' to refer to these citations as they appear to involve both praise and criticism of others' works.

5. Signposting: This functional category is named after Harwood's (2009) category 'Signposting' and Petrić's (2007) 'Further reference', both of which serve the function of referring readers to other sources for further information on the topic, equations/algorithms or methodology. Signposting citations serve three specific functions of helping less informed readers, keeping the argument on track and saving space (Harwood, 2009).
6. Supporting: This function is the same as Harwood's (2009) 'Supporting' function in that information is cited from other sources to support and justify the writers' topic of research, method/methodology and claims, especially when reader/reviewer objections are anticipated.

Examples from the corpus to illustrate each of these categories/sub-categories along with article numbers (in round brackets) are provided in Table 3. 'None' refers to sub-categories that were not found in the introductions of the present corpus.

Category/Sub-category	Example
1. Applying: Adopting/Adapting other authors' works	
- Theory	None (usually prevalent in the methods section of articles)
- Terms/definitions	None (usually prevalent in the methods section of articles)
- Method	To address this, we employed MEGA-PRESS with frequency selective pulses at 1.9 and 1.5 p.p.m.[24] to assess macromolecule suppressed GABA in a cohort of participants with schizophrenia and psychiatrically healthy controls. (B27)
2. Associating: Establishing links between sources	
- Similarities	Interestingly, our preliminary analysis of GABA+ aging trend using the traditional MEGA-PRESS method, showed a similar trend of age-related reduction of GABA levels in patients with schizophrenia.[18] (B27)
- Differences	The inconsistency is reminiscent of earlier MRS [magnetic resonance sprectroscopy] findings in glutamate, although subsequent evidence suggests that glutamate was elevated in younger schizophrenia patients,[20-22] but reduced in chronic, older patients.[18,23] (B27)
3. Building: Building topic knowledge by citing other authors	
- Topic generalisations	Gamma-butyric acid (GABA) is the primary inhibitory neurotransmitter in the mammalian brain.[1] (B27)
- Topic elaborations	GABA dysfunction has been implicated in the pathophysiology and cognitive deficits of schizophrenia, mainly based on post-mortem and preclinical research. [2,3] (B27)
4. Evaluating: Evaluating other authors' works	
- Positively	Such use of levothyroxine (L-T4) at supraphysiologic doses has offered promise in several open-label studies, including rapid cycling,[6] prophylaxis-resistant bipolar patients[7] and for patients with refractory uni- or bipolar depression.[8] (B29)
- Negatively	The overall evidence from treatment trials in bipolar depression is sparse, and the role of antidepressants and their efficacy remains controversial.[2-5] (B29)
5. Signposting: Directing readers to other sources for more information	
- Topic	N-terminal fragments of mutant HTT protein interfere with gene transcription in the early stages of the disease process (reviewed in ref. 16). (B41)
- Equations	As a consequence, the translation efficiency (λ_i) of each transcript (m_i) can be described by a constant (*i.e.* leading to the protein abundance: $P_i = \frac{m_i}{\mu}\lambda_i$ with μ being the rate of growth in h^{-1} (Klumpp *et al*, 2009)). (B32)

- Method	One common approach to this is to rank stretches of sequence by their phylogenetic conservation [7]. (B15)
- Results	By improving the sequencing and data analysis processes, we were able to sequence the cfDNA [cell-free DNA] of the 15 CRPC [castration-resistant] patients that were previously excluded due to low yield of DNA in their plasma samples, to validate all of the mutations reported in our previous work [25] and to identify new candidate mutations. (B14)
6. Supporting: Justifying by citing other authors	
- Topic significance	Human induced pluripotent stem cells (iPSCs) constitute a promising tool for investigating the underlying pathophysiology of traditionally challenging neurodevelopmental disorders.[13] (B26)
- Method	We have demonstrated the utility of iPSCs to investigate the functional consequences of *MECP2* [methyl-CpG-binding ptotein-2] loss of function.[14,15] (B26)
- Key findings	Several studies in transgenic mice, overexpressing *MECP2*, have shown a progressive neurological phenotype, including tremor, gait ataxia, seizures, anxiety, abnormal learning and behavior, and early death, recapitulating some aspects of the human disease.[6-9] (B26)
- Explaining findings	And cyanobacteria which was chosen as a chassis in our study can convert CO_2 and solar energy into organic compounds through oxygenic photosynthesis with an efficiency higher than that of terrestrial plants (Zhou et al., 2014b). (B22)

Table 3. Examples of citations based on their functions.

The subcategories of 'Building' were later collapsed and referred to as 'Topic elaborations' as it was difficult to distinguish between general claims on a topic and their elaborations. Moreover, the authors' claims were not always extracted from other sources as shown in (10):

(10) Topic generalization: Adult tissues are maintained and constantly regenerated by the activity of tissue-specific multipotent stem cells. Topic elaboration: This reservoir is periodically activated and compensates for cell loss and tissue damage, while preserving the stem cell pool (Barker et al., 2010) (B7)

For the sake of economy, two of the sub-categories of 'Supporting' - key findings and explaining findings - were also combined as they are both related to the results of a study. As for the sub-categories of 'Applying-Theory and Terms/definitions', they were not present in the corpus since they usually appear in other sections of a research article such as the methods section (Petrić, 2007). The functional category 'Tying-Similarities/Differences' is not included in this study of article introductions because it is more relevant for the results and discussion section of articles when comparing the findings of one's own study to that of other studies (Petrić, 2007; Mansourizadeh & Ahmad, 2011).

All the citation functions obtained substantial inter-coder agreements of over 80% while 'Evaluating' obtained moderate agreement at 67%. As

previously mentioned, values of 0.61 to 0.80 represent substantial agreement and 0.41 to 0.60 represent moderate agreement. Please refer to Table B2 in Appendix B for the percentage agreement and Cohen's Kappa value for the analysis of citation functions. The discrepancies in coding the negative and positive evaluation functions in citations were resolved by re-examining the identification criteria comprising a set of linguistic markers denoting praise and criticism (for example, positive evaluation is signalled by the word 'promise' and negative evaluation by the words 'sparse' and 'controversial' in Table 3, item #4).

In this study, the citation functions identified are deduced through a text analysis rather than interviews with authors to account for their motivations/intentions for using particular citation types. Although the latter method is insightful, it could not be adopted here because the overall aim was to identify trends in move and citation practice as well as move-citation relationship in articles introductions. To achieve this aim, it was important to extract a reasonably large number of articles from authentic and well-established journals in the field of bioscience.

The final stage of the analysis involved investigating the interaction between moves/steps and citation types most frequently found in the different parts of article introductions. The citation types used with most frequency in the three moves and the reasons for their frequency were identified. The analysis showed that there is in fact a match between the rhetorical structure of article introductions and citation functions in the specific moves/steps. The findings are presented and discussed at length in the next section.

3. Results and discussion

The findings of the genre analysis and citation analysis of bioscience article introductions are reported in this section along with a discussion about the frequency of citation types across different moves/steps.

3.1. Genre analysis of article introductions

The role of introductions in any research article, according to Swales (1990, 2004), is to re-establish the significance of the study, place it in the context of past research and show how it will overcome existing knowledge gaps. The move/step analysis of the 50 articles in the corpus showed that all three

moves are included in the introductions of all the articles irrespective of their length. This rhetorical analysis conforms to Swales' analysis of article introductions with regard to the presence and sequence of moves. However, the frequency of the steps is varied (see Table 4).

Move	Frequency		Steps	Frequency	
Move 1 Introducing the field	50	100%	Step 1 – Claiming centrality of the topic	44	88%
			Step 2 – Elaborating on the topic	50	100%
Move 2 Preparing for the present study	50	100%	Step 1 – Indicating a gap	29	58%
			Step 2 – Raising a question/issue	1	2%
			Step 3 – Extending the research	4	8%
			Combining steps	16	32%
Move 3 Introducing the present study	50	100%	Step 1 – Stating purpose(s)/hypotheses	50	100%
			Step 2 – Describing procedures	40	80%
			Step 3 – Presenting findings	32	64%
			Step 4 – Highlighting the value of the study	23	46%

Table 4. Frequency of moves and steps.

Move 1 serves the dual function of promoting the topic of research (Step 1) and providing adequate information on it (Step 2). As both functions are equally important, the frequency of both steps is very high, at 88% and 100% respectively. Not only do authors have to compete with other researchers for readership but they also have to ensure that, once they have convinced readers of the significance of their research, they need to keep their attention by giving them all the necessary information related to their topic. In accordance with Swales' (2004) modified version of the CARS model, Step 3 'Reviewing items of previous research' was omitted in our analysis as it was observed that citations are not concentrated in one place but are spread in different parts of the introduction. In comparison to Kanoksilapatham's (2005) analysis of biochemistry research articles, the 100% frequency count of Moves 1 and 3 of the introduction is similar, but in her analysis Move 2 only has a frequency of 67%. This shows that there could be variations in the rhetorical structure of bioscience articles depending upon their sub-discipline.

In Move 2, the main function is to convince readers of the motivation for the research. There are three specific ways of achieving this – by indicating that there is a gap in past research (Step 1), raising questions or issues related to past research (Step 2) and/or by stating that past research needs to be extended in some way (Step 3). In terms of frequency, Step 1 seems to be

the preferred way of achieving this in the corpus (58%), followed by combining the steps (32%), whereas the individual realization of the other steps (Steps 2 and 3) is least preferred (10%). Similar to Kanoksilapatham's (2005) findings, it was observed that this move is most frequently realized in biological science articles through Step 1 rather than through the other alternatives.

As for Move 3, the main function seems to be to occupy the niche that has been created in the previous move. Therefore, the step that is most crucial at this stage is stating the aim of the study in the form of purpose statements, research questions or hypotheses (Step 1). Among the four steps, Step 1 has the highest frequency at 100% with Step 2 'Describing procedures' (80%), Step 3 'Presenting findings' (64%), and Step 4 'Highlighting the value of the study' (46%) appearing in declining frequencies. Step 2 probably has a high frequency of occurrence to give readers key information on methodology at the outset, as the method section is usually postponed to the end in biological science articles rather than following the conventional IMRAD sequence (Introduction, Methods, Results and Discussion). In the sample, articles extracted from 8 of the 10 journals have a delayed methods section. However, even the 10 journal articles that follow the conventional IMRAD structure seem to have this step in their introductions, following the conventional structure of introductions in the discipline. Another interesting observation was that 5 out of the 10 article introductions that omit this step are from the same journal (i.e. *Cancer Cell*), indicating that authors could be influenced by journal conventions based on their reading of sample articles. As for Step 3, it appears that it is more likely to be present in hard science articles such as physics (Swales & Najjar, 1987) and biochemistry (Kanoksilapatham, 2005), than in social science articles. According to these authors (Swales & Najjar, 1987; Kanoksilapatham, 2005), only the principal findings are announced in the introduction as a teaser to the more detailed information in the results section. In biological science articles, it appears that the authors attempt to provide a preview of not only the results but of the entire article by means of Move 3 in their introductions. To determine whether the presence or absence of moves/steps in the introduction could be influenced by the instructions to authors, a review was conducted of the 10 journal websites from which data were extracted for this study. It was found that these documents only provide general guidelines on the structure of articles, without elaborating on the rhetorical structure of article introductions.

3.2. Citation analysis in article introductions

The citations in this study were analyzed according to functional categories and according to their interaction with moves and steps in article introductions. The total number of citations in the article introductions of 50 articles were 699 (in 3,338 words), with the largest number of 481 in Move 1, followed by 186 in Move 2 and the smallest number of 32 in Move 3. Past research (Thompson, 2005; Mansourizadeh & Ahmad, 2011) has shown that citation density is high in the introduction of articles when compared to the other sections, since writers need to position their articles in the related research context, review relevant past literature and highlight the gap in past research.

As for citation types, it has been documented that non-integral citations are predominantly used in hard disciplines because writers place more emphasis on the message rather than on authors to show objectivity and to avoid interruptions in the flow of their arguments (Hyland, 1999, 2000; Thompson, 2005; Hewings et al., 2010). Although biological science articles in this study tend to use both the name-year (60%) and numbered (40%) style format depending upon journal preferences, the citation analysis showed that only 2% of the citations are integral citations, whereas the remaining 98% are non-integral. This confirms Hyland's (1999, 2000) findings that non-integral citations are prominently used in the hard sciences because the focus is on conveying ideas objectively as opposed to singling out authors who propose them. Apart from objectivity, Hewings et al. (2010) point out that non-integral citations enable writers to provide an uninterrupted flow in their arguments. The propensity of non-integral citations could also be due to the choice of journals, most of which fall under the molecular biology category. Comparison with articles from other sub-disciplines of Biology may reveal differences that are worth exploring.

Furthermore, a comparison between citation forms revealed that there is a propensity for paraphrasing in biological science articles, especially in molecular biology articles. With the exception of one direct quote citation, all the other citations in the present corpus were paraphrases of information from other sources. In Dubois' study (1988), biomedical scientists confirmed that extended quotations are rarely used in biomedical science articles and are only restricted to phrases and terms that belong to the public domain of science.

The functional analysis of citations shows that their concentration and functions vary across moves in article introductions. In the present corpus,

Move 1 has the largest number of citations (68%), with less than half this amount in Move 2 (27%) and the least in Move 3 (6%). The distribution of citations also varies according to functions, although with some degree of overlap (see Table 8). Tables 5-7 present the distribution frequency of citations according to their functions in the three moves.

The density of citations in Move 1 is the highest at 68% and correspondingly, it has a spread of citations with 5 different functions, as shown in Table 5.

Citation Function	Step 1: Claiming centrality of the topic	Step 2: Elaborating on the topic	Total	%
Associating	–	18	18	4%
Building	–	251	251	52%
Evaluating	1	41	42	9%
Signposting	–	21	21	4%
Supporting	45	104	149	31%
Total	46	435	481	100%

Table 5. Functions of citations in Move 1.

The top 2 citation functions of supporting (31%) and building (52%) in Move 1 correlate with the rhetorical functions of Step 1 (Claiming the centrality of the topic) and Step 2 (Elaborating on the topic). The supporting citations are mainly the 'Topic Significance' sub-type in Step 1 but 'Key Findings' in Step 2. As for the building citations, their main function is to offer information from various sources to build on the readers' topic knowledge and provide them sufficient background knowledge to contextualize and understand the article.

In Move 2, the predominant citation functions are those of supporting (33%), evaluating (32%) and building (25%), as can be seen in Table 6. In this move, evaluating citations usually appear after the building (i.e. information elaboration) and supporting (i.e. key findings of other studies) citations, because authors usually review past literature on the topic before evaluating it negatively in support of the move's function of highlighting a gap in past research, as in (11):

(11) Building: Thus, the term MDS/AML [myelodysplastic syndromes/acute myeloid leukemia] is often used to describe a biological continuum in leukemia pathogenesis (Cazzola et al., 2013; Kreso and Dick, 2014). Evaluating (Negative): However, the molecular basis of MDS to AML

> transition remains poorly understood (Bejar and Steensma, 2014; Larsson et al., 2013; Zhang et al., 2015), underscoring the need to better understand the specific pathways responsible for disease initiation and progression. (B1)

The analysis revealed that the majority (83%) of the evaluation citations (i.e. 49 out of 59) are indeed negative in Move 2 of the introduction. Although there are some positive evaluations (17%), they usually precede the negative evaluations, either as a buffer or as contrast in praise-criticism pairs (Example 12), a juxtapositioning strategy that is also mentioned in Petrić (2007):

> (12) Evaluating (Positive): Interestingly, the presence of TILs [tumor-infiltrating lymphocytes] has been previously shown to correlate with better patient outcomes during various antitumor therapies in multitude of cancers (Galon et al., 2006; Hwang et al., 2012; Mahmoud et al., 2011). Evaluating (Negative): However, it is commonly known that the tumor microenvironment often inhibits activated T cells from entering tumor tissues or prevents effective T cell priming for tumor control through various pathways (Gajewski et al., 2013). (B5)

As for the match between move and citation functions, this strategy of critiquing previous work via citations is consistent with the move function of preparing the ground for the present study.

Citation Function	Step 1: Indicating a gap	Step 2: Raising a question/issue	Step 3: Extending the research	Total	%
Associating	10	1	–	11	6%
Building	40	3	4	47	25%
Evaluating	54	–	5	59	32%
Signposting	5	–	2	7	4%
Supporting	29	3	30	62	33%
Total	138	7	41	186	100%

Table 6. Functions of citations in Move 2.

Move 3 has very few citations (6%) because in this part of the introduction the authors usually introduce their own work in response to the gap that has been created in the previous move. They provide a preview of their article by elaborating on the purpose, method, results and significance of their study. As in other moves, the citations are specifically related to these rhetorical functions (see Table 7), with signposting citations mainly justifying

the methods used (Example 13, Step 2) and supporting citations highlighting or explaining key findings of other studies (Example 14, Step 3):

(13) Move 3 Introducing the Present Study-Step 2 Describing Procedures
Signposting: For the protein domain division, we annotate each gene's protein domains based on the Conserved Domain Database (CDD) [6], a collection of conserved domain sequences. (B15)

(14) Move 3 Introducing the Present Study-Step 3 Presenting Findings
Supporting: Based on the strong evidence supporting GABAergic function in working memory,[2,25] we predicted that GABA level would specifically relate to working memory compared with other clinical, cognitive and functional measures. (B27)

Citation Function	Step 1: Stating purpose(s) /hypotheses	Step 2: Describing procedures	Step 3: Presenting findings	Step 4: Highlighting the value of the study	Total	%
Applying	–	3	–	–	3	9%
Signposting	–	12	1	3	16	50%
Supporting	1	4	4	–	9	28%
Others	1	3	–	–	4	13%
Total	2	22	5	3	32	100%

Table 7. Functions of citations in Move 3.

The citation functions and their frequency across different moves in the introductions are presented in Table 8.

Citation Function	Move 1	Move 2	Move 3	Total
Building	84%	16%	–	42.6%
Supporting	68%	28%	4%	31.5%
Evaluating	42%	58%	–	14.4%
Signposting	46%	16%	38%	6.3%
Others	50%	31%	19%	5.2%
Total	67.7%	26.6%	5.7%	100%

Table 8. Citation functions across moves/steps.

The overall findings reveal that the most predominant citations in article introductions are building (43%) and supporting (32%) citations in keeping with the main role of these introductions, which is to prepare readers for the article and to raise its credibility. Similarly, the distribution of citations across moves shows that building (84%) and supporting (68%) citations are also

most predominant in Move 1 to support the move's functions of boosting topic centrality (Step 1) and topic knowledge (Step 2). As for evaluating citations, the trend seems to be to evaluate other studies positively in Move 1 (78% of the positive evaluations across moves) but negatively in Move 2 (88% of the negative evaluations across moves), supporting the opposing functions of the two moves – those of building topic knowledge and showing a gap in past research, respectively. Signposting citations are important in Move 1-Step 2 (46%) and Move 3-Step 3 (38%) in order to economize on length of topic elaborations in the former move and methodological information in the latter. The distribution of the citations with different functions across moves provides concrete evidence that there is a close relationship between rhetorical functions of moves and citation functions in introductions of biological science articles in the corpus.

4. Conclusion

This study was motivated by our work with science PhD students in a graduate English course, in response to questions related to research paper writing in their disciplines. The focus of this study was on examining the relationship between rhetorical moves and citation functions in biological science article introductions, as there is a dearth of studies that specifically examine the interaction between these two aspects of writing in academic articles. The findings show that there is indeed a one-to-one relationship between moves/steps and citations, although some functions cut across moves and steps. The study has direct implications for designing English for Specific Academic Purposes courses, especially for graduate students. Not only are these students required to write dissertations but they are also expected to publish journal articles in their fields during their candidature. In fact, most universities nowadays provide these students an option of submitting dissertations that are a compilation of published articles. Therefore, raising their awareness of both rhetorical structure and citation use in article introductions, as well as across different sections of articles would be beneficial for writing dissertations and journal articles. As the focus of this article is only on article introductions of biological science articles, the interaction between moves and citations needs to be examined in other sections of the article and extended across other disciplines, because there could be disciplinary variations across fields. Future work could also build a database of typical moves, citations and linguistic features for novice

researchers in different fields. Other aspects that could be explored include citation density, citation integration and sources of citations in research articles within and across disciplines. For a holistic account of move and citation practice, it is also recommended that the text analysis carried out in this study be supplemented by a questionnaire survey or interview of authors to gain insight into their move and citation practice motivations.

Acknowledgements

This article is an outcome of a project entitled 'Evidence-Based Teaching of Literature Review Writing', funded by the Singapore Ministry of Education, Tertiary Education Research Fund (Grant no. MOE2015-1-TR05).

Article history:
Received 07 January 2019
Received in revised form 25 March 2020
Accepted 30 April 2020

References

Anthony, L. (1999). "Writing research article introductions in software engineering: How accurate is a standard model?" *IEEE Transactions on Professional Communication* 42(1): 38-46.

Bazerman, C. (1988). *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science*. Madison: University of Wisconsin Press.

Belcher, D. (1994). "The apprenticeship approach to advanced academic literacy: Graduate students and their mentors". *English for Specific Purposes* 13(1): 23-34.

Bloch, J. & L. Chi (1995). "A comparison of the use of citations in Chinese and English academic discourse" in D. Belcher & G. Braine (eds.), *Academic Writing in a Second Language: Essays on Research and Pedagogy*, 231-274. Norwood, NJ: Ablex.

Charles, M. (2006). "Phraseological patterns in reporting clauses used in citation: A corpus based study of theses in two disciplines". *English for Specific Purposes* 25(3): 310-331.

Cohen, J. (1960). "A coefficient of agreement for nominal scales". *Educational and Psychological Measurement* 20(1): 37-46.

Cooper, C. (1985). *Aspects of Article Introductions in IEEE Publications*. Unpublished master's dissertation, University of Aston, Birmingham, UK.

Crookes, G. (1986). "Towards a validated analysis of scientific text structure". *Applied Linguistics* 7(1): 57-70.

Del Saz Rubio, M.M. (2011). "A pragmatic approach to the macro-structure and metadiscoursal features of research article introductions in the field of agricultural sciences". *English for Specific Purposes* 30(4): 258-271.

Dontcheva-Navratilova, O. (2015). "Cross-cultural variation in citation practices: A comparative analysis of citations in Czech English-medium and international English-medium linguistics journals" in R. Plo-Alastrue & C. Pérez-Llantada (eds.), *English as a Scientific and Research Language*: *Debates and Discourses*, 185-205. Berlin/Boston: De Gruyter Mouton.

Dontcheva-Navratilova, O. (2016). "Rhetorical functions of citations in linguistics research articles: A contrastive (English-Czech) study". *Discourse and Interaction* 9(2): 51-74.

Dubois, B. L. (1988). "Citation in biomedical journal articles". *English for Specific Purposes* 7(3): 181-193.

Habibi, P. (2008). "Genre analysis of research article introductions across ESP, psycholinguistics, and sociolinguistics". *Iranian Journal of Applied Linguistics* 11(2): 87-114.

Harwood, N. (2009). "An interview-based study of the functions of citations in academic writing across two disciplines". *Journal of Pragmatics* 41(3): 497-518.

Hewings, A., T. Lillis & D. Vladimirou (2010). "Who's citing whose writings? A corpus based study of citations as interpersonal resource in English medium national and English medium international journals". *Journal of English for Academic Purposes* 9(2): 102-115.

Hirano, E. (2009). "Research article introductions in English for specific purposes: A comparison between Brazilian Portuguese and English". *English for Specific Purposes* 28(4): 240-250.

Hopkins, A. (1985). *An Investigation into the Organizing and Organizational Features of Published Conference Papers*. Unpublished master's dissertation, University of Birmingham, UK.

Hu, G. & G. Wang (2014). "Disciplinary and ethnolinguistic influences on citation in research articles". *Journal of English for Academic Purposes* 14(1): 14-28.

Hughes, G. (1989). *Article Introductions in Computer Journals*. Unpublished master's. dissertation, University of Birmingham, UK.

Hyland, K. (1999). "Academic attribution: Citation and the construction of disciplinary knowledge". *Applied Linguistics* 20(3): 341-367.

Hyland, K. (2000). *Disciplinary Discourses: Social Interactions in Academic Writing*. Harlow/New York: Longman.

Jalilifar, A. R. (2012). "Research article introductions: Sub-disciplinary variations in applied linguistics". *Journal of Teaching Language Skills* 29(2): 29-55.

Joseph, R., J.M.H. Lim & N.A.M. Nor (2014). "Communicative moves in forestry research introductions: Implications for the design of learning materials". *Procedia-Social and Behavioral Sciences* 134: 53-69.

Kanoksilapatham, B. (2005). "Rhetorical structure of biochemistry research articles". *English for Specific Purposes* 24(3): 269-292.

Kanoksilapatham, B. (2012). "Research article structure of research article introductions in three engineering subdisciplines". *IEEE Transactions on Professional Communication* 55(4): 294-309.

Kanoksilapatham, B. (2015). "Distinguishing textual features characterizing structural variation in research articles across three engineering sub-discipline corpora". *English for Specific Purposes* 37(1): 74-86.

Landis, J.R. & G.G. Koch (1977). "The measurement of observer agreement for categorical data". *Biometrics* 33(1) 159-174.

Lewin, B.A., J. Fine & L. Young (2001). *Expository Discourse: A Genre-based Approach to Social Science Research Texts*. New York: Continuum.

Lillis, T., A. Hewings, D. Vladimirou & M. J. Curry (2010). "The geolinguistics of English as an academic lingua franca: Citation practices across English-medium national and English-medium international journals". *International Journal of Applied Linguistics* 20(1): 111-135.

Loi, C.K. (2010). "Research article introductions in Chinese and English: A comparative genre-based study." *Journal of English for Academic Purposes* 9(4): 267-279.

Mansourizadeh, K. & U.K. Ahmad (2011). "Citation practices among non-native expert and novice scientific writers". *Journal of English for Academic Purposes* 10(3): 152-161.

Maswana, S., T. Kanamaru & A. Tajino (2015). "Move analysis of research articles across five engineering fields: What they share and what they do not". *Ampersand* 2: 1-11.

Moravcsik, M.J. & P. Murugesan (1975). "Some results on the function and quality of citations". *Social Studies of Science* 5(1): 86-92.

Nwogu, K.N. (1997). "The medical research paper: Structure and functions". *English for Specific Purposes* 16(2): 119-138.

Ozturk, I. (2007). "The textual organisation of research article introductions in applied linguistics: Variability within a single discipline". *English for Specific Purposes* 26(1): 25-38.

Peng, P. (1987). "Organizational features in chemical engineering research articles": ELR Journal 1(1):79-116.

Pennycook, A. (1996). "Borrowing others' words: Text, ownership, memory and plagiarism". *TESOL Quarterly* 30(2): 201-230.

Petrić, B. (2007). "Rhetorical functions of citations in high- and low-rated master's theses". *Journal of English for Academic Purposes* 6(3): 238-253.

Posteguillo, S. (1999). "The schematic structure of computer science research articles". *English for Specific Purposes* 18(2): 139-160.

Rowley-Jolivet, E. & S. Carter-Thomas (2014).

"Citation practices of expert French writers of English: Issues of attribution and stance" in A. Łyda & K. Warchał (eds.), *Occupying Niches: Interculturality, Cross-culturality and Aculturality in Academic Research,* 17-34. New York: Springer.

Samraj, B. (2002). "Introductions in research articles: Variation across disciplines". *English for Specific Purposes* 21(1): 1-17.

Samraj, B. (2005). "An exploration of a genre set: Research article abstracts and introductions in two disciplines". *English for Specific Purposes* 24(2): 141-156.

Shadish, W.R., D. Tolliver, M. Gray & S.K. Sen Gupta (1995). "Author judgments about works they cite: Three studies from psychology journals". *Social Studies of Science* 25(3): 477-498.

Swales, J. (1981). *Aspects of Article introductions*. Birmingham, England: University of Aston.

Swales, J. (1986). "Citation analysis and discourse analysis". *Applied Linguistics* 7(1): 39-56.

Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge: CUP.

Swales, J. (2004). *Research Genres: Explorations and Applications*. Cambridge: CUP.

Swales, J. & H. Najjar (1987). "The writing of research article introductions". *Written Communication* 4(2): 175-191.

Thompson, C. (2005). "'Authority is everything': A study of the politics of textual ownership". *International Journal of Educational Integrity* 1(1): 1-12.

Thompson, D.K. (1993). "Arguing for experimental "facts" in science: A study of research article results sections in biochemistry". *Written Communication* 10(1): 106-128.

Thompson, P. (2000). "Citation practices in PhD theses" in L. Burnard & T. McEnery (eds.), *Rethinking Language Pedagogy from a Corpus Perspective*, 91-101. Bern: Peter Lang.

Thompson, P. & C. Tribble (2001). "Looking at citations: Using corpora in English for academic purposes". *Language Learning & Technology* 5(3): 91-105.

Viera, A.J. & J.M. Garrett (2005). "Understanding interobserver agreement: The kappa statistic". *Family Medicine* 37(5): 360-363.

Wang, P. & M.D. White (1999). "A cognitive model of document use during a research project. Study II. Decisions at the reading and citing stages". *Journal of the American Society of Information Science* 50(2): 98-114.

White, H.D. (2004). "Citation analysis and discourse analysis revisited". *Applied Linguistics* 25(1): 89-116.

Yang, R. & D. Allison (2003). "Research articles in applied linguistics: Moving from results to conclusions". *English for Specific Purposes* 22(4): 365-385.

Dr Sujata Kathpalia is a Senior Lecturer at the Language and Communication Centre (LCC), Nanyang Technological University, Singapore. She has designed many academic writing courses and she teaches these courses to undergraduates and postgraduates. Her research interests include genre analysis, academic writing, code switching and English language teaching.

Dr Christopher Khoo is an Associate Professor in the Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore. He teaches courses in knowledge organization, information behavior, and data mining. His research interests are in text mining, medical data mining, knowledge organization, and information behavior on social media.

Appendix A

B1: Guezguez, B. et al. (2016). *Cancer Cell* 29: 61-74. URL: http://doi.org/10.1016/j.ccell.2015.11.012

B5: Tang, H. et al. (2016). *Cancer Cell* 29: 285-296. URL: http://doi.org/10.1016/j.ccell.2016.02.004

B6: Bu, P. et al. (2016). *Cell Stem Cell* 18: 189-202. URL: http://doi.org/10.1016/j.stem.2016.01.006

B7: Chiacchiera, F. et al. (2016). *Cell Stem Cell* 18: 91-103. URL: http://doi.org/10.1016/j.stem.2015.09.019

B8: Zhang, Y. et al. (2016). *Cell Stem Cell* 18: 368-381. URL: http://doi.org/10.1016/j.stem.2016.02.001

B14: Lallous, N. et al. (2016). *Genome Biology* 17: 10. url: http://doi.org/10.1186/s13059-015-0864-1

B15: Gussow, A. B. et al. (2016). *Genome Biology* 17: 9. url: http://doi.org/10.1186/s13059-016-0869-4

B22: Wang, Y. et al. (2016). *Metabolic Engineering* 34: 60-70. URL: http://doi.org/10.1016/j.ymben.2015.10.008

B26: Nageshappa, S. et al. (2016). *Molecular Psychiatry* 21: 178-188. URL: http://doi.org/10.1038/mp.2015.128

B27: Rowland, L. M. et al. (2016). *Molecular Psychiatry* 21: 198-204. URL: http://doi.org/10.1038/mp.2015.34

B29: Bauer, M. et al. (2016). *Molecular Psychiatry* 21: 229-236. URL: http://doi.org/10.1038/mp.2014.186

B32: Borkowski, O. et al. (2016). *Molecular Systems Biology* 12: 870. url: http://doi.org/10.15252/msb.20156608

B33: Chandrasekaran, S. et al. (2016). *Molecular Systems Biology* 12: 872. url: http://doi.org/10.15252/msb.20156777

B41: Dickey, A. S., et al. (2016). *Nature Medicine* 22,1: 37-45. URL: http://doi.org/10.1038/nm.4003

B44: Sagoo, P. et al. (2016). *Nature Medicine* 22,1: 64-71. URL: http://doi.org/10.1038/nm.4016

Appendix B

Move and steps	Percentage agreement	Cohen's Kappa
Move 1 Introducing the field		
Step 1 Claiming centrality of the topic	0.87	0.44
Step 2 Elaborating on the topic	1.00	-
Move 2 Preparing for the present study		
Step 1 Indicating a gap	0.80	0.55
Step 2 Raising a question/issue	0.40	0.12
Step 3 Extending the research	0.80	0.33
Move 3 Introducing the present study		
Step 1 Stating purpose(s)/hypotheses	0.93	- *
Step 2 Describing procedures	0.67	0.40
Step 3 Presenting the findings	0.93	0.86
Step 4 Highlighting the value of the study	0.93	0.86

**This step is almost always present.*

Table B1. Results of inter-coder reliability test for step analysis.

Citation function	Percentage agreement	Cohen's Kappa
Applying	– *	–
Associating	0.80	0.47
Building	0.87	0.44
Evaluating	0.67	0.33
Signposting	0.93	0.63
Supporting	0.80	– **
Tying	– *	–

**This citation function was not found in any Introduction section in the sample.*
***The second coder identified this citation function in every article.*

Table B2. Results of inter-coder reliability test for analysis of citation functions.