Please cite this paper as follows:
Ganji, M., & Derakhshan, A. (2020). Developing a checklist for evaluating research articles in applied linguistics. *Teaching English Language, 14*(2), 239-268. https://doi.org/ 10.22132/TEL.2020.121858

Research Paper

Developing a Checklist for Evaluating Research Articles in Applied Linguistics

Mansoor Ganji
English Department, Chabahar Maritime University, Chabahar, Iran
Ali Derakhshan[1]
Department of English Language and Literature, Faculty of Humanities and Social Sciences, Golestan University, Gorgan, Iran

Abstract

Although researchers, postgraduate students, and university professors are under tremendous pressure to publish or perish, there exist few widely acknowledged rating scales or checklists for evaluating manuscripts submitted to journals in applied linguistics. To top it all, the existent peer-reviewing criteria are typically not publicly shared; thus, peer-reviewing is a closed, occluded genre. To fill this lacuna in applied linguistics, this research critically analyzed the evaluation scales, existing review criteria, and available documents from 18 journals, two publishing institutions, and APA guidelines in order to come up with a comprehensive checklist for evaluating research articles in applied linguistics. The analysis of the data through inductive content analysis revealed 43 categories, which were later classified under six main themes. It was found that researchers must be aware of *the practicalities* of research, include the *essentials* in their manuscript, and appeal to editors by adding elements of *face validity*. Having met these requirements, researchers must enjoy excellent *composition skills* and adhere to the fundamental principles in research *ethics* in order to be able to write a research paper having high *overall value*. The paper concludes with several implications for researchers and opens up some avenues for future research.

---

[1] Corresponding author: a.derakhshan@gu.ac.ir

1. Introduction

Producing quality research articles has become a great obsession for researchers around the globe, determining whether they will publish or perish (Grimes, Bauch, & Ioannidis, 2018). In such a reality, researchers are under tremendous pressure to constantly generate new research products to keep abreast of their field and survive during the scientific journey (Fanelli, 2010). Among them, PhD candidates are pushed to get published to obtain desirable PhD marks, defend dissertations, and find academic positions (Jalilifar & Shahvali, 2013; Paltridge, 2017; Raitskaya & Tikhonova, 2020). As Coriat (2019) maintains, "publications are often considered a hard currency for evaluating PhD students by graduation committees and funders alike" (p. 1007). Similarly, getting published is a necessity for university professors, required to regularly update and enrich their publication records for getting job promotions, receiving salary increase, or even not being fired (Haven, Bouter, Smulders, & Tijdink, 2019; Maniati, Jalilifar, & Hayati, 2015; Raitskaya & Tikhonova, 2020). However, the bitter fact is that despite this essentiality of research literacy for staying in academia, in many cases, graduate programs fail to sufficiently acquaint students with the knowledge of how to conduct objective, quality, scientific research studies. As a result, many postgraduate students have problems with writing research articles and publishing them (Khany & Abol-Nejadian, 2010).

This situation is aggravated by the fact that journals, being one of the main gatekeepers of determining the quality, value, and methodological rigor of proposed research works for potential publication (Ali & Watson, 2016), may also neglect to sufficiently instruct reviewers on rating a proposed piece of research consistently and objectively (Paltridge, 2017). This, in turn, may jeopardize the quality of the journals because, as Saxena, Thawani, Chakrabarty, and Gharpure (2013) hold the "worthiness of any scientific

journal is measured by the quality of the articles published in it" (p. 125). In essence, inadequate guidance on the peer review process hinders authors and reviewers' inclination and ability to be genuinely engaged in the process (Ali & Watson, 2016). In this regard, Shattell, Chinn, Thomas, and Cowling (2010) also evinced that following quality assurance mechanisms is a prerequisite for journals to guarantee publication of excellent scientific outputs, and in this process, the research literacy of reviewers is very important, because their feedback plays an important role in the decision of the editors to reject or accept proposed manuscripts for publication (Ali & Watson, 2016). In a nutshell, the quality of the reviewing process acts as a criterion to judge the quality of both manuscripts and the journals (Hames, 2007).

As Paltridge (2017) expressed, "training in peer review is an important part of researcher development, given the importance peer review has in the academic community as a whole" (p. 145). In response to this concern, rather recently, some journals and publishers (e.g., *British Medical Journal, Elsevier, the Asia TEFL Journal, and Springer*) have initiated offering peer review training to their reviewers (Schroter et al., 2004). Essential to this training process are acquainting reviewers with the standards of good scientific pieces and also specifying valid and thorough article evaluative criteria. Equally important is to publicly announce these criteria to authors as authors' awareness of the details of the peer review facilitates their better understanding of the process, meeting the expectations of the journals, conforming to the standards of excellence, and preparing higher quality manuscripts (Lee, Sugimoto, Zhang, & Cronin, 2013). Therefore, these indicators play a decisive role in the quality of the peer-reviewing process, which in turn is highly influential in the "professional practice, the academic reward system, and the scholarly publication process" (Lee et al., 2013, p. 2). When accurate evaluative principles are present, a single submitted manuscript will be judged

identically by different reviewers, and less degree of bias, imprecision, and inconsistency will be noticed in the peer-reviewing process (Lee et al., 2013).

This concern has been attended to by various journals which employed different yardsticks based on their priorities and preferences. Nevertheless, the problem is that there is no agreed-upon definition of what constitutes good research in a particular field, and journals in a discipline may set diverse criteria that may be biased, fixed, or impartial (Garcia, Rodriguez-Sánchez, & Fdez-Valdivia, 2015). The resulting unfair evaluations and decisions of such journals are voiced in the complaints of authors, PhD students, researchers, or professors who do not know how to write to please and convince the reviewers, editors, and publishers whose decisions can be subjective and unfair in their minds when precise and consistent evaluative indices are not present (Smith, 2006). To cater for this issue, publishers and editors in a particular research area need to jointly work to gather sufficient field information to develop comprehensive, field-specific evaluative standards to be followed by the members of their respective communities of practice (Nedić & Dekanski, 2016). What current research evidence clearly shows is that, in the majority of disciplines, there is no globally-accepted evaluative form to consistently and objectively assess research works (Mårtensson, Fors, Wallin, Zander, & Nilsson, 2016).

To our dismay, neither does such a comprehensive index exist in our respective field of applied linguistics to accurately gauge the quality of scientific products in this domain. Therefore, to fill this research gap, by drawing on the existing evaluative criteria specified in various quality journals in applied linguistics in both local and international journals, the present study endeavors to develop a comprehensive model to establish the definition of the quality concept for research articles in applied linguistics in the context of Iran. Then, to operationalize this model, an evaluative checklist will be designed as

a communication tool to be used by Iranian applied linguistics journals. It is hoped that this clarity in criteria and standards aids the reviewers to avoid bias in their ratings and reviews, and prevent authors from engaging in plagiarism, data fabrication, and low-quality research productions.

2. Review of the Related Literature
2.1 The Genre of Peer Review Reports

Genres are communicative events that have their distinct lexico-grammatical features and communicative functions and are used in certain settings by members of specific communities of practice (Flowerdew, 2013). Each genre involves a particular schematic structure, being realized through a set of sequential stages (Soodmand Afshar, Doosti, & Movassagh, 2018). According to the advocates of the ESP school to genre theory, being competent at performing a certain genre requires knowing both its staging and the specific form-function mappings at each stage (Swales, 1990). For such a knowledge of genre conventions to be developed, individuals need to receive instruction regarding a particular genre and its writing habits and processes and be exposed to text examples within that genre (Jalilifar, 2010). That is why people who are not members of particular discourse communities are easily recognized as outsiders for their little knowledge of a particular genre by discourse community members (Bhatia, 2004). As Flowerdew (2013) mentions, for instance, writing a research article in a specific discipline requires a lot of expert knowledge on the part of the insiders within that professional community.

This also applies to writing reviewers' report, recognized as a distinct genre within the broader domain of academic writing (Swales, 1996). Review reports are written for a specific audience, are organized in certain ways, discuss certain content, make distinct language choices, are realized in a specific context, and have a distinct purpose (Paltridge, 2017). They also reflect

Developing a Checklist …

specific worldviews and the criteria and characteristics that a quality research must meet in a particular discipline, which are desired by that disciplinary community (Devitt, 2015). Researchers must also be informed of reviewing criteria in a certain discipline to assist them to understand whether their articles fit with those criteria, how their works are assessed by peer reviewers, and whether their works meet the requirements for getting published in quality journals (Marsden, 2015). However, the problem is that journals' peer-reviewing criteria are typically not publicly shared; hence, little is divulged about peer reviews and the standards being important to editors, who are the ultimate assessors of the submitted manuscripts (Falkenberg & Soranno, 2018).

This private nature of peer reviews is well evidenced in the concerns of Swales (1996), who states that because peer-reviewing is a closed, occluded genre (i.e., not available to public), there is a relatively poor understanding of what peer reviewing is and how it is produced, especially among early career researchers, lacking sufficient experience of this type of writing. Despite this shortage of knowledge on how reviewers go for the reviewing task and the challenges they encounter for writing reviews (Hames, 2007), peer-reviewing is an activity that most of the researchers undertake throughout their academic lives (Paltridge, 2017). Butler (1990) propounded that as researchers take more experience of writing review reports for peer-reviewed journals, they gain more expertise in how to make review judgments, behave in expected ways, make specific language choices, perform their identity as a peer reviewer, and on the whole, show their degree of membership in the peer reviewers' community of practice (Strauss & Feiz, 2014).

According to Belcher (2007), it is essential to understand how reviewers evaluate proposed manuscripts, make judgments on their quality, and how they cope with challenges when preparing reviewers' reports. In this regard, the UK

House of Commons Science and Technology Committee (2011) suggested peer review training to early researchers. Similarly, the British Medical Association proposed preparing postgraduate students for doing peer review. To support this momentum, Elsevier and Springer have also run peer review training programs by presenting videos, interactive courses as well as online lectures on their websites (Paltridge, 2017). Davidoff (2004) accentuated that for such pieces of training to be effective, they should go beyond the information transmission approach and incorporate a problem-solving, reflective, and hands-on learning approach. A journal in the realm of applied linguistics that provides peer review mentoring is the Asia TEFL Journal, offering a period of apprenticeship to new reviewers through engaging with experienced reviewers to acquire reviewer quality (Adamson & Fujimoto-Adamson, 2016). In an empirical study of Paltridge (2017), focusing on the opinions of 47 reviewers of the English for Specific Purposes journal regarding their experience of the peer review process, the reviewers reported to have learned how to write reviews through taking the reviews of their own works as models and through learning by doing peer reviews. Furthermore, the reviewers mentioned that there was no agreement among the reviewers on the most clear-cut facets of writing reviews.

2.2 Quality Standards for Peer Review of Journal Articles

Journals are one of the central venues for storage and dissemination of academic knowledge in every discipline, the quality of which normally influences the impact and visibility of research products they encompass. At the same time, the quality of journals is determined by the research articles they decide to publish, which are gauged against evaluative criteria set by the journals themselves (Egbert, 2007). Therefore, the quality of journal editors' and reviewers' evaluations plays a crucial part in the publication process, as they act as quality control agents deciding whether submitted manuscripts are

Developing a Checklist …

publishable or not (Hames, 2007). Peer review has been criticized for stifling creativity; being time-consuming; and being open to abuse, unreliable, biased, expensive, and subjective, and poor at identifying fraud and errors in research. They are further called a flawed process as the quality of reviews may differ, and reviewers' views may vary, journals attend to peer review differently, and on the whole, no clear criteria and benchmarks are present for peer review (Garcia et al., 2015; Lee et al., 2013; Smith, 2006). Nevertheless, despite these deficits, peer review is an essential step in the quality assessment of research practice (Hames, 2007).

According to the European Science Foundation (2012), one of the primary issues in most scientific domains is evaluating the excellence of research products. Nevertheless, there seems to be a shortage of a common definition of what makes quality research and a scarcity of universal evaluative benchmarks for research practice (Mårtensson et al., 2016; Paltridge, 2017). In the absence of such a common metric, review criteria that university boards, journal editors, reviewers, and funding agencies employ may be biased and idiosyncratic, which in turn can hinder making fair evaluations (Mårtensson et al., 2016). The question is how to consistently produce and evaluate research works if such generally-acknowledged yardsticks are not present. The suggestion is, if certain evaluative models are to be developed, it is better if they are discipline-specific. As Mårtensson et al. (2016) rightly articulated, a particular set of evaluative standards working well in a certain discipline, like medicine, may not work well in a dissimilar discipline, such as engineering. Langfeldt, Nedeva, Sörlin, and Thomas (2020) approve this argument by declaring that research quality standards are normally relative and are developed and understood differently within disciplines and contexts.

In an attempt to make a stride toward developing a common language for research articles evaluation in a specific discipline, through reviewing the

previous literature and guidelines, Rostami, Khadjooi, Abasaeed-Elhag, and Ishaq (2011) developed a systematic evaluative tool for assessing the scientific value of research papers in biomedical journals. In another study, Lee et al. (2013) were concerned with the issue of bias in peer-reviewing process. In their review study, the authors examined the history, function, and scope of peer review, unraveling and criticizing the claims of bias. On the other hand, peer-reviewing impartiality is realized when the same evaluative criteria are applied and interpreted similarly and consistently for gauging a manuscript by different reviewers. The authors concluded that as the issues of impartiality and sociality are inevitable in the peer-reviewing process, it is questionable whether impartiality should or can be regarded as the aim in the peer review.

Similarly, Tight (2003), considering the peer-reviewing process an important yet a covert and unexplored activity, examined the quality and quantity of the reviews he had received on his submitted manuscripts and book proposals. He also examined the different reviews on the same papers for exploring the points of convergence and divergence in the reviews, and also how the decisions of the editors might be related to the reviews. Overall, what he concluded was that "reviewers, particularly good reviewers, are hard to find" (p. 302). This point was also reached by Karney (1998), who stated that disappointment and misunderstanding might happen with the reviewing process, as unqualified works might get published while quality papers may erroneously get rejected.

Furthermore, in a systematic review study, Belcher, Rasmussen, Kemshaw, and Zornes (2016) attempted to design a protocol including quality criteria for a comprehensive, systematic, and rigorous review of research practice in a transdisciplinary context. Meanwhile, they reported the rarity of generally-acknowledged transdisciplinary evaluative benchmarks and criteria because of the difficulty of setting standards that hold true across disciplines. Despite the

hardships, Belcher et al. (2016) developed a transdisciplinary quality assessment framework encompassing a set of criteria, principles, and definitions. Various sub-criteria were introduced under the four main criteria of relevance, credibility, legitimacy, and effectiveness. Research has evinced that the need for developing a general evaluative model is also felt in applied linguistics as each journal in this area employs its own set of criteria for reviewing manuscripts. Lazaraton (2003) investigated the diverse set of criteria being proposed for assessing qualitative articles in applied linguistics. One primary finding of this study was that the employed criteria in this field typically privilege certain types of qualitative research, namely conversation analysis and ethnography, at the expense of excluding others. What the author of this research argued was that as evaluative criteria determine to a large degree the quality and characteristics of research articles getting published, there is a strong link between the criteria that the professional journals establish and validate, and the choices that the researchers make during their research production, including the choice of research method.

Paltridge (2017) dedicated a book to examining the genre of peer review in applied linguistics. For this aim, he collected 97 reviewers' reports to submissions to the journal of English for Specific Purposes and 45 reviewers' responses to a questionnaire, asking about the reviewers' experiences and challenges about reviewing and how they learned to write review reports. Among the findings of this book was that as the reviewers' experience of taking on the role of peer-reviewing for journals increases, they become more competent at how to make judgments in a specific context and write review reports in expected ways to show their membership in the peer reviewers' community of practice in applied linguistics. The results of all these studies attest to the concern that commonly-agreed evaluative criteria are lacking in many disciplines.

To sum it up, most of the research studies conducted in the area of evaluating research articles have investigated the problems researchers have while writing papers (Khany & Abol-Nejadian, 2010), the genres and moves of research papers or its different sections (Adel & Ghorbani Moghadam, 2015; Alinasab, Gholami, & Mohamadnia, 2021; Mehrabi, Jalilifar, Hayati, & White, 2018; Paltridge, 2017), stance and personal identity construction (Babaii, Atai, & Mohammadi, 2015; Rahimi, Yousofi, & Moradkhani, 2019), and phrases and lexical bundles in academic writing (Nasseri, 2021; Shirazizadeh & Amirfazlian, 2021). The second group of studies has focused on the reviewing process (Tight, 2003), reviewers training (Davidoff, 2004), transdisciplinary quality assessment (Belcher et al., 2016), reviewing the existing scales (Lazaraton, 2003), and review scales in biomedical journals (Rostami et al., 2011). As an attempt to fill this research gap, in the present study, we endeavored to design a common metric to be used in the peer-reviewing process of research articles submitted to Iranian journals of applied linguistics.

3. Methodology
3.1 Design of the Study
Descriptive design is a research design in which various aspects of a social phenomenon are discussed, the relationship among its variables are described, and the results are presented using descriptive statistics (Riazi, 2016). This study employed a descriptive design in order to describe and criticize the existing evaluation scales, if any, which are used for reviewing, evaluating, and judging the quality of manuscripts submitted to the Iranian and foreign journals in the field of applied linguistics. The present research is exploratory since, to the best of the researchers' knowledge, this study addresses a research gap to which scant attention has been paid in the Iranian context (Rostami et al., 2011). However, the main aim of this research was to develop a

comprehensive checklist for evaluating the quality of research papers in the field of applied linguistics.

3.2 Data Collection Procedures

In order to gather the needed data used by different journals, editors, and reviewers and develop a comprehensive checklist for evaluating the papers submitted to journals in applied linguistics, the researchers searched the websites of all the Iranian journals which publish papers in applied linguistics and some high-ranking foreign journals according to their impact factor (IF) index. This was done in order to find the rating scales employed by the reviewers and editors of these journals. However, none of the target journals publicly announced or posted any criteria, rating scale, or checklist for evaluating the papers. Thus, the researchers contacted the managing editors, editors, editors-in-chief, and reviewers of these journals for the required data. The researchers emailed at least three authorities in each journal to obtain the existing data. Since most of the editors of the contacted high-ranking journals did not respond to the emails or stated that they did not employ any objective scale and relied on the subjective and qualitative judgments of the reviewers, the researchers had to contact other journals in the list which were lower in ranking. All in all, the researchers contacted more than 38 journals in this field, eight of which did not respond to the emails. Another group of journals' authorities (12 journals) responded that they did not employ any objective and fixed criteria for evaluating the articles, and they mostly relied on the expert but subjective comments of the reviewers. Finally, the rating scales of 18 journals were collected, put together, and analyzed. Besides these journals, the researchers collected data from two main publishers of journals and well-known associations in publishing academic research (Wiley, Springer, and American Psychological Association). The data collected from all these sources were typed, put together, and categorized according to the sources

from which they had been compiled. Our initial corpus resulted in a bank of data, which was 14 pages (4646 words).

3.3 Data Analysis Procedure

The texts of the criteria, scales, and email responses were read three times by the researchers. Using the Track Changes and Comment function in Microsoft Word, the data were categorized, and the categories and themes were labeled. In fact, an inductive content analysis approach was followed, and the researchers read and coded the data without using any previously used scale or checklist. In order to ensure inter-rater reliability in coding, the researchers did the first reading together and coded the first five pages of the text separately. The inter-rater reliability was .89. Then, they read and coded the rest of the text and reported the results. Reading and analyzing the data for the first time, the researchers could only find those categories which were obviously noticed, shown by headings, or were easy to find. These categories were named according to the phrases used in the text (around 40%). For example, some of the names which were directly taken from the words used by the journals were significance, contribution, originality, so on.

After reading the texts for the second time, some other parts of the text which were similar to the existing categories in nature but used a different wording were discovered and added to the available categories (almost 25%). For example, novelty was added to originality. Finally, the researchers went through the remainder of the data, looking for new, insignificant, irrelevant, or repetitive categories. This time, the researchers had to create some new names and labels (e.g., argumentations), modify the previous categories (e.g., gap in knowledge), making them broader to include these small themes (e.g., significance of the study), or dividing them into two categories (e.g., review divided into up to review and references) because of finding more differences in the instances. This stage resulted in the naming of 35% of the categories.

4. Results and Findings

First of all, the texts were checked for repetitive and irrelevant information, which resulted in the reduction of the text from 4646 words to 3705 words. In doing so, the texts were converted into separate lines or paragraphs, each one focusing on one topic. The names for categories and themes were made consistent in spelling, and each category was given a code. This time the number of words were reduced to 3537 words and typed in 37 pages. All in all, 43 categories were found, having at least one to 33 subcategories or wordings. To be more exact, the final analysis and checking of the categories resulted in the following findings.

Table 1

Categories Used for Assessing the Quality of Papers in Applied Linguistics Journals

Number	Name of the Category	Repetitions in the Bank of Items
1.	Abstract	11 times
2.	Argumentation	6 times
3.	Citations	1 time
4.	Clear aims and questions	12 times
5.	Conclusion	10 times
6.	Contextualization	4 times
7.	Contribution	11 times
8.	Data analysis	8 times
9.	Data collection	3 times
10.	Discussion	18 times
11.	Ethics	12 times
12.	Findings	6 times
13.	Following APA	5 times
14.	Future research	1 time
15.	Implications	2 times
16.	Instrument	4 times
17.	Introduction	8 times
18.	Keywords	1 time
19.	Length	2 times
20.	Limitations	5 times
21.	Main elements	4 times
22.	Methodology	33 times

23.	Organization	8 times
24.	Originality	18 times
25.	Overall quality	17 times
26.	Participants	5 times
27.	Plagiarism	2 times
28.	Rationale	4 times
29.	References	27 times
30.	Results	8 times
31.	Review of literature	10 times
32.	Sampling	3 times
33.	Scope	9 times
34.	Search Engine Optimization	2 times
35.	Significance of the study	9 times
36.	Statement of the problem	3 times
37.	Statistics	3 times
38.	Suggestions	3 times
39.	Tables and figures	8 times
40.	Theoretical framework	1 times
41.	Theory and practice	1 times
42.	Title	10 times
43.	Writing quality	22 times

As the names of categories show, the evaluation of papers in the journals is based on many factors. The first theme can be entitled *Essentials*, which includes four categories of scope, length of the article (word limits), plagiarism, and main elements. This is called *Essentials* since those papers which exceed the word limit, use plagiarized materials more than a certain amount specified on the journal's website, do not fit the scope, or do not contain one of the main and essential elements of an academic paper will be desk-rejected. However, it must be mentioned that some of these decisions in this stage might not be objective; as a result, there are authors who complain of being rejected because of scope, for example. Finally, one category which was not present in the gathered data but can be added to this list is the format of the papers. Some journals desk-reject those papers which do not follow a certain format specified by the journal's editorial board.

The second theme, which was not limited to a specific part of the document, was concerned with the overall face and mechanics of the study, and might be neglected by some journals, was named *Face Validity*. This theme included three categories of search engine optimization, following APA rules and regulations, and in-text citations. These categories can be checked by the editor-in-chief or managing editor by having a quick look at the manuscript. Therefore, neglecting these basic skills in academic writing, which does not require much time or expertise, might lead to the rejection of the manuscript by the journal in a matter of a few days or sometimes a few months.

The third theme, which was related to specific parts, sections, or even subsections of an academic paper and was the most important theme in most journals, can be called *The Practicalities*. This theme included the highest number of categories (26 categories out of 43), including the category with the highest repetition (methodology 33 repetitions), and most probably is the most decisive factor while deciding for the rejection or acceptance of manuscripts by the reviewers. This is the area the evaluation of which takes more time, patience, meticulous reading, searching, and energy. Thus, the editors or managing editors assign two to three reviewers to do this. However, as mentioned above, each and every journal or reviewer has its own criteria or subjective judgment in this regard. The fact that there are so many categories in this regard is due to the fact that we have compiled all the factors and scales available in 21 sources. This theme includes the following categories: abstract, clear aims and questions, conclusion, data analysis, data collection, discussion, findings, future research, implications, instrument, introduction, keywords, methodology, participants, references, results, review of literature, sampling, significance of the study, statement of the problem, statistics, suggestions, tables and figures, theoretical framework, limitations, and title.

The next theme was exactly the opposite of the previous one in two aspects. Firstly, it was not clear how the reviewers should measure this aspect of the manuscript; put it simply, it was not operationalized. Secondly, these categories were implemented or could be scattered throughout the text; thus, the reviewer had to read the paper completely in order to make a fair decision in this regard. This theme was entitled *Overall Value*, which consisted of five categories. Categories such as contribution, originality, overall quality, rationale, and theory and practice. The paper was expected to contribute to the field meaningfully, be original, present a good academic work with useful results for the community, be based on a strong reason and address an important problem, and address both the theory and practice of language teaching.

Another theme that was general in nature but was mostly related to the writing ability and skills of the researchers was entitled "*Composition Skills*" and consisted of four categories of argumentation, organization, writing quality, and contextualization of the topic. This theme is so important since even an excellent piece of research that is presented in a weak form and language is bound to be rejected. In other words, this is the packaging and presentation of the ideas, which makes the first and most significant impression both on the editor and reviewers since it can even conceal other weaknesses.

Last but by no means least, the other category, which is an exceptionally important aspect of conducting research, is usually neglected by some researchers and does not have a specific heading in the manuscripts, in our view, is called *"Ethics"* of research. The researchers intentionally separated this single category and considered it alone as a theme since ethics is the single most important factor in conducting any research. Figure 1 below displays the information on the themes found in this study.
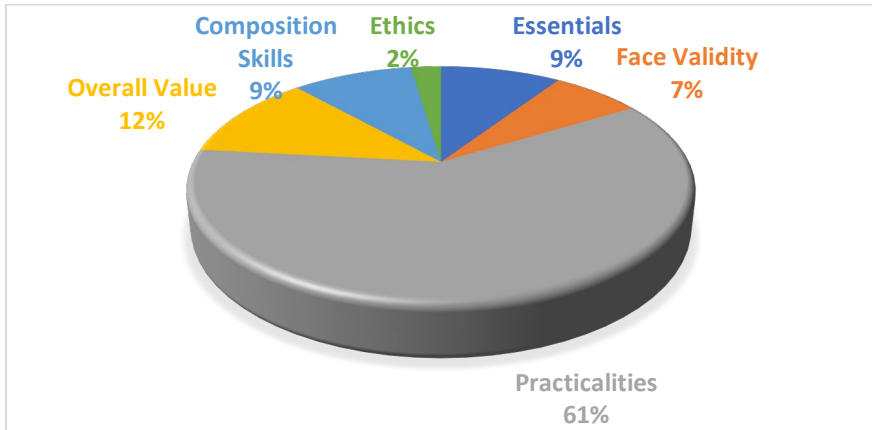
Developing a Checklist …

Figure 1. The Six Themes Involved in Evaluating the Manuscripts

Based on the data gathered, categories found, and the frequency and weight of each category, the researchers came up with a comprehensive checklist that can be used by journals, editors, or reviewers while deciding about the quality of a manuscript. This checklist is detailed enough and includes almost all the items mentioned by the said sources. It is ordered based on the number of times each item appeared in the data in order to show how important each item is. Then, each item is followed by the main elements or aspects which were highlighted by different journals (Table 2).

Table 2
The Final Checklist for Evaluating Manuscripts in Applied Linguistics

Number	Item and Frequency	Phrases used by the journals
1.	Methodology (33 times)	Be appropriate, sound, and valid for answering the questions Be clearly articulated in enough details Include design, participants, data collection, and data analysis Use ppropriate statistical techniques
2.	References (27 times)	Be current, comprehensive, and relevant Be accurate and according to the suggested format Include domestic and foreign sources

		Balance between books, journal articles, and conferences
3.	Writing quality (22 times)	Follow spelling, punctuation, and grammar rules
		Write in academic style and register
		Be clear and easy to follow
		Contain enough cohesion and coherence
4.	Originality (18 times)	Contain an innovative aspect
		Present new content
		Address a recent issue
		Enjoy originality and novelty
		Address a recent topic of interest
5.	Discussion (18 times)	Discuss the findings
		Relate them to previous studies
		Provide reasons
		Go beyond data and results
		Discuss overall threats to internal validity
		Report only the most important data
		Mention both for and against arguments
6.	Overall quality (17 times)	Discuss the overall strengths
		Mention the main weaknesses
		Make a judgment about the overall evaluation
		Provide main comments for the editor
7.	Clear aims and questions (12 times)	Contain clear research questions
		Present unambiguous research hypotheses
8.	Ethics (12 times)	Mention the confidentiality of the data
		Code the participants
		Ask the participants to sign the consent forms
		Give the participants enough information about research
9.	Abstract(11 times)	Summarizing the content accurately
		Avoid unnecessary information
		Stand alone
		Follow IMRD model
10.	Contribution (11 times)	Fill in a research gap
		Contribute to the field significantly
		Resolve conflicts in the field
11.	Conclusion (10 times)	Draw sound conclusions
		Present relevant implications
		Be only based on and supported by the data
12.	Review of the literature (10 times)	Contain relevant, up to date, and sufficient information
13.	Title (10 times)	Capture readers' attention
		Reflect the content
		Be clear

Developing a Checklist …

14.	Significance of the study (9 times)	Address an important issue Explicitly present the significance Have a clear rationale
15.	Scope (9 times)	Be interesting for the readers Be relevant to the aims Look interesting for the researchers
16.	Tables and figures (8 times)	Be necessary Be arranged and formatted well Avoid duplication of data presented within the text Be followed by detailed and clear legends
17.	Results (8 times)	Be according to the research questions Answer the research questions Do not contain unnecessary data Do not include interpretations
18.	Organization (8 times)	Enjoy high readability Use cohesive devices Use enough and clear headings Present the materials logically
19.	Introduction (8 times)	Present the background Engage the readers Present the problems in a context Include the rationale for conducting the research End in aims of the study
20.	Data analysis (8 times)	Use appropriate statistical tests Interpret the results
21.	Argumentation (6 times)	Use clear argumentation Present the arguments logically
22.	Findings (6 times)	Present the findings clearly Contain adequate explanations Present the main findings only
23.	Following APA (5 times)	Follow APA in in-text citations Follow APA in the reference list
24.	Limitations (5 times)	Acknowledge the limitations Suggest specific areas for future research
25.	Participants (5 times)	Describe the participants/corpus clearly Mention the number of participants State the sampling procedure Show the time and place of the study
26.	Contextualization (4 times)	Contextualize the research problem in the introduction Present the theoretical framework of the study
27.	Rationale (4 times)	Mention the reason for conducting the research

28.	Main elements (4 times)	Contain abstract, keywords, introduction, review, method, results, discussion, conclusion, and references
29.	Instrument (4 times)	Have a piloting stage to ensure reliability Validate the instrument Acknowledge the sources of the instruments
30.	Data collection (3 times)	Explain the procedure in detail
31.	Sampling (3 times)	Use a clear sampling method Present the reason for using this method
32.	Statement of the problem (3 times)	State the problem clearly
33.	Suggestions (3 times)	Make suggestions based on the findings
34.	Statistics (1 time)	Use appropriate statistical test Meet their assumptions
35.	Implications (2 times)	End in several relevant implications for the field
36.	Length (2 times)	Do not exceed the word limit
37.	Plagiarism (2 times)	Use available plagiarism software
38.	Search Engine Optimization (2 times)	Optimize the title, abstract, and keywords for search purposes
39.	Future research (1 time)	Make several suggestions for future research
40.	Keywords (1 time)	Include five to seven keywords
41.	Theoretical framework (1 time)	Be based on an appropriate theoretical framework
42.	Theory and practice (1 time)	Address both theory and practice
43.	Citations (1 time)	Support every claim or argument by citing

5. Discussion

The analysis of the data revealed that there existed six main themes in the evaluation of articles in applied linguistics. The findings and results of the study were similar to the elements highlighted by Aspinwall, Simkins, Wilkinson, and McAuley's (1992) model, Rostami et al.'s (2011) checklist, Belcher et al.'s (2016) evaluative framework, and Mårtensson et al.'s (2016) concept model of research and concept hierarchy of research quality. The first important factor in evaluating the papers was following a series of rules and regulations called *Essentials*. In fact, the papers were first checked for scope,

Developing a Checklist …

length (word limits), plagiarism, and main elements. Tight (2003), Mårtensson et al. (2016), and Belcher et al. (2016) also emphasized that one of the essential features of a good paper is whether it is relevant to the scope of the journal or not. As the name of the theme reveals, almost all journals specify some certain limitations on these factors, and these are the essential features of every manuscript submitted to journals. A quick look at the homepages of almost all journals and, more specifically, the guide for authors sections, illustrates that these features are required in all journals. However, since most of the journals are published online these days, the limitations on the word limit are not as strict as before. As regards the finding on plagiarism, other researchers also considered plagiarism as one of the serious problems in academic writing and a reason for the rejection of papers (Baždarić, Bilić-Zulle, Brumini, Petrovečki, 2012; Butler, 2010; Zhang, 2010). Containing the main elements of a paper was previously pointed out by Rostami et al. (2011). They argue that a high-quality manuscript should include abstract, objectives and hypotheses, background, participants, sample size, recruitment, baseline data, randomization, statistical methods, methods, results, outcomes, all of which must be judged objectively by the reviewers.

The next finding was related to a set of factors that had to do with search engine optimization, following APA, and in-text citations. This comes as no surprise since a large number of journals in the field of applied linguistics require the authors to follow APA format in their submissions. The next important factor which was pointed out by some journals was that the abstract and keywords must be optimized for searching. That is to say, they ask the authors to use suitable titles and keywords in order to help future researchers find their articles more easily, hence increasing the chances of referencing and improving the journal's ranking. This finding is in line with Mårtensson et al.'s

(2016) assertions, who postulated that a high-quality paper should be searchable, consumable, and accessible.

Furthermore, another finding was related to the issues which seem to be the most important factors in the evaluation of a manuscript since this finding included the highest number of categories (26 categories out of 43), and even the category with the highest repetition, methodology with 33 repetitions), was in this category. Evaluating this aspect of research was found to be the main job of reviewers. The results showed that enjoying clear aims and research questions, following a certain methodology, basing the arguments on previous studies, and being useful for the society and the academic community were among the important features of a research paper. All these areas were included in the model presented by Mårtensson et al. (2016). They developed a new conceptual model of research and a concept hierarchy of research quality. Belcher et al. (2016) also argued that a high-quality research should be relevant and credible, which means they should start with obvious problems and come up with practical solutions to these problems through following a systematic and detailed plan, basing their judgments on the reliable data. With regard to the findings on the introduction and its main elements, this study confirms the findings by Ebrahimi and Weisi (2019). On the other hand, the results revealed that following a specific and detailed plan for conducting the research, utilizing the right instruments, using the right statistical procedures, choosing representative participants through correct sampling procedure, and interpreting the results fairly were other important factors in evaluating an article, and reviewers' dissatisfaction with each of these elements will lead to the rejection of the manuscript. These findings are similar to the previous checklists, identified problems in writing research papers, and models put forward by previous researchers (Belcher et al., 2016; Khany & Abol-

Nejadian, 2010; Mårtensson et al., 2016; Rostami et al.,2011; Schou, Høstrup, Lyngsø, Larsen & Poulsen, 2011).

   The analysis of the data also revealed that, besides meticulously reading and reviewing each and every section of the article, journal editors and reviewers need to make sound decisions regarding the overall value and worth of a manuscript, depending on the reason the research was conducted, originality of the research, and the extent to which it contributes to the theory and practice in the said area. These findings are in line with those of Mårtensson et al. (2016) and Schou et al. (2011), who concluded that any quality research study is a conscious action, a response on the part of the researcher to the problems noticed in the context. They found the originality and contribution of the research so important that one of the four main concepts in their conceptual model of good research was labeled contributory. Belcher et al. (2016) also underlined the contribution that research papers make as an essential element and included this in the relevance category defined in their quality assessment framework. Most of the respondents asserted that originality, contribution, and theory and practice link are among the most important factors when they want to choose an academic journal (Egbert, 2007).

   However, the composition skills of the researchers can be considered as the single most important feature of a manuscript, which can greatly impress the editors and reviewers. It seems that if the researchers present their content in a language that is accurate, flawless, and organized, they might manage to persuade the editors successfully. Besides writing quality, the argumentation and organizing skills of the writer as well as contextualizing the topic are also needed in creating a good impression on the reviewers. These findings are in line with the previous literature revealing that efficient research papers must employ enough argumentation skills (Beck, 1990; Jalilifar, 2011; Schou et al.,

2011; Somashekhar, 2020), be presented in flawless language (Egbert, 2007; Maniati, Jalilifar, & Hayati, 2015; Somashekhar, 2020), be organized well (Belcher et al., 2016), and be communicable (Mårtensson et al., 2016).

The last finding, which is, in the researchers' point of view, the most important aspect of a manuscript and research project, is the inclusion of some information on ethics. This category was not present in most of the checklists prepared and used for reviewing the research articles. Corroborating this finding, several scholars considered ethics an essential feature of every research study. Pickersgill (2012) and Schou et al. (2011) rightly argued that science is an ethical business by nature, and science and ethics are in many ways co-produced. Besides, Belcher et al. (2016) and Mårtensson et al. (2016) included ethics as one of the components of legitimacy and conforming categories in their quality assessment framework.

6. Conclusion

Having analyzed the rating scales, evaluation checklists, and documents available on the websites of 18 journals in applied linguistics, Wiley and Springer guides for authors, and American Psychological Association notes on academic papers, the researchers found 43 categories, which were later categorized under six main themes. These were the main criteria that were employed in the evaluation of the submitted manuscripts. The researchers finally came up with a comprehensive checklist that can be used by novice and experienced researchers who do not know exactly on what criteria the editors and reviewers of the journals in applied linguistics base their decision. This checklist can also be used by supervisors, postgraduate students, and inexperienced reviewers who are going to submit a high-quality paper or have just started their job as a reviewer.

This checklist not only included the general and common elements of these models, frameworks, and checklists but also added some new dimensions to

Developing a Checklist …

these sources and made them more filed specific. This checklist is different from the previous ones, which presented a limited list of general items with no indication of how these elements could be evaluated or what elements they should include. To be more exact, this checklist has operationalized the previously presented abstract concepts and has considered all the elements from A to Z. Furthermore, it can expand our knowledge in this regard since some of the previous studies in this area based their model and framework on the data gathered from different fields which might differ considerably in the way their manuscripts are written and evaluated.

This study, like any other research, suffers from a number of shortcomings. Although the researchers emailed several journal editors, editors-in-chief, and managing editors, only a limited number responded to the emails or sent the evaluation rubric in response. Some other journals did not employ any rating scale for evaluating the manuscripts and relied solely on the subjective judgment of the reviewers; thus, the number of scales received was not as expected. Another limitation of the study was that most of the scales collected and analyzed in this study have come from the Iranian journals in applied linguistics, so the results should be generalized cautiously. Still another limitation was that there existed little research on the development of a rating scale, questionnaire, or checklist specifically designed for evaluating papers in applied linguistics.

Considering the above-mentioned limitations, future researchers are suggested to prepare a checklist or questionnaire based on the comments of the reviewers and editors because there is enough data in this regard, and these comments can help us design another checklist or scale for evaluating research manuscripts. Future studies can also validate this checklist or questionnaires developed in this area. Finally, another line of research can investigate the

influence of using checklists and evaluation scales on the objectivity of the evaluation process and the final opinion of the reviewers.
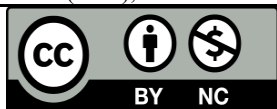
References

Adamson, J., & Fujimoto-Adamson, N. (2016). Sustaining reviewing quality: Induction, mentoring, and community. *Journal of English Scholars Beyond Borders, 2*, 29-57.

Adel, M. R., & Ghorbani Moghadam, R. (2015). A comparison of moves in conclusion sections of research articles in psychology, Persian literature and applied linguistics. *Teaching English Language, 9*(2), 167-191.

Ali, P. A., & Watson, R. (2016). Peer review and the publication process. *Nursing Open, 3*(4), 193-202.

Alinasab, M., Gholami, J., & Mohammadnia, Z. (2021). Genre-based revising strategies of graduate students in applied linguistics: Insights from term papers. *Journal of English for Academic Purposes, 49,* https://www.sciencedirect.com/science/article/pii/S1475158520304756

Aspinwall, K., Simkins, T., Wilkinson, J. F., & McAuley, M. G. (1992). *Managing evaluation in education.* London, UK: Routledge.

Babaii, E., Atai, M. R., & Mohammadi, V. (2015). Stance in English Research articles: Two disciplines of the same science, *Teaching English Language, 9*(1), 1-27.

Baždarić, K., Bilić-Zulle, L., Brumini, G., Petrovečki, M. (2012). Prevalence of plagiarism in recent submissions to the Croatian Medical Journal. *Science and Engineering Ethics, 18,* 223–239.

Beck, C. T. (1990). The research critique: General criteria for evaluating a research report. *Journal of Obstetric Genecologic, and Neonatal Nursing, 19*(1), 18-22.

Belcher, B. M., Rasmussen, K. E., Kemshaw, M. R., & Zornes, D. A. (2016). Defining and assessing research quality in a transdisciplinary context. *Research Evaluation, 25*(1), 1-17.

Belcher, D. D. (2007). Seeking acceptance in an English-only research world. *Journal of Second Language Writing, 16*(1), 1-22.

Bhatia. V. K. (2004). *Worlds of written discourse: A genre-based view.* London, UK: Continuum International.

Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity.* New York, NY: Routledge.

Butler, D. (2010). Journals step up plagiarism policing. *Nature, 466*(7303), 167.

Coriat, A. (2019). PhD merit needs to be defined by more than just publications. *Nature Human Behaviors, 3*(10), 1007.

Davidoff, F. (2004). Improving peer reviews: Who's responsible? *BMJ, 328*, 658-659.

Devitt, A. (2015). Genre performances: Genre analysis and rhetorical-linguistic genre studies. *Journal of English for Academic Purposes, 19*, 44-51.

Ebrahimi, B., & Weisi, H. (2019). Genre variation in the introduction of scientific papers in Iranian and international computer science journals. *Issues in Language Teaching*, *8*(2), 51-82.

Egbert, J. (2007). Quality analysis of journals in TESOL and Applied Linguistics. *TESOL Quarterly, 41*(1), 157-171.

European Science Foundation, (2012). Evaluation in research and research funding organizations: European practices. In: *A report by the ESF Member Organization Forum on Evaluation of Publicly Funded Research.*, ISBN978-2-918428-83-1.

Falkenberg, L. J., & Soranno, P. A. (2018). Reviewing reviews: An evaluation of peer reviews of journal article submissions. *Limnology and Oceanography Bulletin, 27*(1), 1-5.

Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States data, *PLOS ONE, 5*(4), 1-7.

Flowerdew, J. (2013). *Discourse in English language education*. London, UK: Routledge.

Garcia, J. A., Rodriguez-Sánchez, R., & Fdez-Valdivia, J. (2015). Bias and effort in peer review. *Journal of the American Society for Information Science and Technology, 66*(10), 2010-2030.

Grimes, D. R., Bauch, C. T., & Ioannidis, J. P. A. (2018). Modelling science trustworthiness under publish or perish pressure. *Royal Society Open Science, 5*(1), 1-14.

Hames, I. (2007). *Peer review and manuscript management in scientific journals: Guidelines for good practice*. Malden, MA: Blackwell.

Haven, T. L., Bouter, L. M., Smulders, Y. M., & Tijdink, J. K. (2019). Perceived publication pressure in Amsterdam: Survey of all disciplinary fields and academic ranks. *PLOS ONE, 14*(6), 1-12.

Jalilifar, A. (2010). Thematization in EFL students' composition writing and its relation to academic experience. *RELC Journal, 41*(1), 31-45.

Jalilifar, A. (2011). World of attitudes in research article discussion sections: A cross-linguistic perspective. *Journal of Technology and Education, 5*(3), 177-186.

Jalilifar, A., & Shahvali, P. (2013). Writing *suggestions for further research* in Iranian applied linguistics theses: A generic and metadiscoursal investigation. *Iranian Journal of Applied Linguistics (IJAL), 16*(1), 79-105.

Karney, B. W. (Guest Editorial) (1998). Reviewing the review process. *Journal of Hydraulic Engineering, 124*(6), 561-562.

Khany, R., & Abol-Nejadian, R. (2010). Iranian post-graduate students' problems in writing academic research articles. *Teaching English Language, 4*(2), 1-26.

Langfeldt, L., Nedeva, M., Sörlin, S., & Thomas, D. A. (2020). Co-existing notions of research quality: A framework to study context-specific understandings of good research. *Minerva, 58*, 115-137.

Lazaraton, A. (2003). Evaluative criteria for qualitative research in applied linguistics: Whose criteria and whose research? *The Modern Language Journal, 87*(1), 1-12.

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology 64*(1), 2-17.

Maniati, M., Jalilifar, A., & Hayati, A. (2015). Iranian scholars' revision of their submitted manuscripts: Signaling impersonality in text. *Journal of Research in Applied Language Studies, 6*(1), 118-140.

Marsden, J. R. (2015). How could my paper have gotten rejected? In J. Liebowitz (Ed.), *A guide to publishing for academics. Inside the publish or perish phenomenon* (pp. 67-78). Boca Raton, FL: CRC Press.

Mårtensson, P., Fors, U., Wallin, S.-B., Zander, U., & Nilsson, G. H. (2016). Evaluating research: A multidisciplinary approach to assessing research practice and quality. *Research Policy, 45*(3), 593-603.

Mehrabi, K., Jalilifar, A., Hayati, A., & White, P. (2018). Investigating action nominalization in the introduction sections of research articles: A cross-disciplinary study of hard and soft sciences. *Teaching English Language, 12*(2), 1-26.

Nasseri, M. (2021). Is postgraduate English academic writing more clausal or phrasal? Syntactic complexification at the crossroads of genre, proficiency, and statistical modelling. *Journal of English for Academic Purposes, 49,* https://www.sciencedirect.com/science/article/pii/S1475158520304744

Nedić, O., & Dekanski, A. (2016). Priority criteria in peer review of scientific articles. *Scientometrics, 107*, 15-26.

Paltridge, B. (2017). *The discourse of peer review: Reviewing submissions to academic journals.* London, UK: Palgrave Macmillan: London.

Pickersgill, M. (2012). The co-production of science, ethics and emotion. *Science, Technology, and Human Values, 37*, 579–603.

Rahimi, M., Yousofi, N., & Moradkhani, S. (2019). Exploring the professional identity construction of ELT researchers in higher education. *Teaching English Language, 13*(1), 85-117.

Developing a Checklist …

Raitskaya, L. & Tikhonova, E. (2020). Pressure to publish internationally: Scholarly writing coming to the fore. *Journal of Language and Education, 6*(1), 4-7.

Riazi, M. A. (2016). *The Routledge encyclopedia of research methods in applied linguistics: Quantitative, qualitative and mixed-methods,* Routledge. Basingstoke: Taylor & Francis Ltd.

Rostami, K., Khadjooi, K., Abasaeed-Elhag, R., & Ishaq, S. (2011). How to evaluate a manuscript for publication? *Gastroenterology and Hepatology: From Bed to Bench, 4*(2), 58-62.

Saxena, A., Thawani, V., Chakrabarty, M., & Gharpure, K. (2013). Scientific evaluation of the scholarly publications. *Journal of Pharmacology and Pharmacotherapeutics, 4*(2), 125-129.

Schou, L., Høstrup, H., Lyngsø, E.E., Larsen, S. & Poulsen, I. (2011). Validation of a new assessment tool for qualitative research articles. *Journal of Advanced Nursing, 68*(9), 2086-2094.

Schroter, S., Black, N., Evans, S., Carpenter, J., Godlee, F., & Smith, R. (2004). Effects of training on quality of peer review: Randomised controlled trial. *BMJ, 328*, 328-673.

Shattell, M. M., Chinn, P., Thomas, S. P., & Cowling, W. R. (2010). Authors' and editors' perspectives on peer review quality in three scholarly nursing journals. *Journal of Nursing Scholarship, 42*(1), 58-65.

Shirazizadeh, M., & Amirfazlian, R. (2021). Lexical bundles in theses, articles and textbooks of applied linguistics: Investigating intradisciplinary uniformity and variation. *Journal of English for Academic Purposes, 49*, https://doi.org/10.1016/j.jeap.2020.100946

Smith, R. (2006). Peer review: A flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine, 99*(4), 178-182.

Somashekhar, S. P. (2020). Art of scientific writing. *Indian Journal of Gynaecologic Oncology, 18*(2). https://doi.org/10.1007/s40944-020-00382-y

Soodmand Afshar, H., Doosti, M., & Movassagh, H. (2018). A genre analysis of the introduction section of applied linguistics and chemistry research articles. *Iranian Journal of Applied Linguistics, 21(*1), 163-214.

Strauss, S., & Feiz, P. (2014). *Discourse analysis. Putting our worlds into words*. New York, NY: Routledge.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge, UK: Cambridge University Press.

Tight, M. (2003). Reviewing the reviewers. *Quality in Higher Education, 9*(3), 295-303.

Zhang, Y. (2010). Chinese journal finds 31% of submissions plagiarized. *Nature, 467*(7312), 153.