Paraphrase patterns of expert academic writers: implications for writing development, writing pedagogy, and plagiarism policies

Patrones de paráfrasis de escritores académicos expertos: implicancias para el desarrollo de la escritura, pedagogía de la escritura y políticas de plagio

Paul Michiels*, Karyn Kessler** y Paul Rogers***

ABSTRACT

This study offers a linguistic analysis of language borrowing in expertly produced paraphrases. Within the context of higher education, paraphrase writing, an essential skill for source-based writing tasks across the curriculum and a key component in the development of disciplinary expertise, represents a challenge for students and teachers because 1) there is no precise and generally accepted definition of acceptable academic paraphrase, 2) discussions of paraphrase are more often framed in terms of plagiarism than effective writing, and 3) little is known about the linguistic (or rhetorical dimensions) of paraphrases in writing published by experts. The present study analyzes five corpora of paraphrase/source passage pairs \( (n = 233) \) produced by expert writers. Corpora 1 and 2 contain exemplar paraphrase/source passage pairs drawn from writing guides and handbooks designed for college students. The remaining three corpora of paraphrase/source passage pairings were assembled using award-winning published articles in three disciplines. Using Keck’s (2006) taxonomy for classifying paraphrases, the study found that about 27% of the average paraphrase of the expert writers

Keywords: language borrowing, paraphrases, expert academic writing, writing pedagogy, plagiarism policies.

* Estadounidense. PhD. George Washington University. Washington, Estados Unidos. pmsc56@gmail.com
** Estadounidense. PhD. University of California, Santa Barbara. Santa Barbara, Estados Unidos. kkessler@linguistics.ucsb.edu
*** Estadounidense. PhD. University of California, Santa Barbara. Santa Barbara, Estados Unidos. paulrogers@writing.ucsb.edu
analyzed here is made up of language found in the source passage. Paraphrase patterns of expert academic writers point toward a potential continuum of acceptable language borrowing practices likely driven by disciplinary differences. Implications for writing development, writing pedagogy, plagiarism policy, and further research are discussed.

RESUMEN
Este estudio ofrece un análisis lingüístico de préstamos en paráfrasis elaboradas por expertos. En el contexto de la educación superior, la escritura parafraseada, una habilidad esencial para las tareas de escritura basadas en fuentes en todo el plan de estudios y un componente clave en el desarrollo de la experiencia disciplinaria, representa un desafío para los estudiantes y la facultad porque 1) no existe una definición precisa y generalmente aceptada de paráfrasis académica aceptable, 2) las discusiones sobre la paráfrasis se enmarcan más a menudo en términos de plagio que de escritura efectiva, y 3) se sabe poco sobre las dimensiones lingüísticas (o retóricas) de la paráfrasis en escritos publicados por expertos. El presente estudio analiza cinco corpus de pares de pasajes de paráfrasis/fuente \( (n = 233) \) producidos por escritores expertos. Los corpus 1 y 2 contienen ejemplos de pares de pasajes de paráfrasis/fuente extraídos de guías de escritura y manuales diseñados para estudiantes universitarios. Los tres corpus restantes de pares de pasajes de paráfrasis/fuente se ensamblaron utilizando artículos publicados en tres disciplinas. Usando la taxonomía de Keck (2006) para clasificar las paráfrasis, el estudio encontró que alrededor del 27% de la paráfrasis promedio de los escritores expertos analizados aquí se compone del lenguaje que se encuentra en el pasaje fuente. Los patrones de paráfrasis de escritores académicos expertos apuntan a un potencial continuo de prácticas aceptables de préstamos lingüísticos probablemente impulsadas por diferencias disciplinarias. Se discuten las implicaciones para el desarrollo de la escritura, la pedagogía de la escritura, la política de plagio y la investigación futura.
Introduction

A hallmark of academic writing is the integration of new knowledge to a body of relevant, existing knowledge of a field. For expert academic writers, the effective integration of known and emerging knowledge is a commonplace and a critical priority in the writing of research presentations, books, literature reviews, multiple sections of articles (e.g. introductions, methods, discussions), and more. Reflecting what Harris (2006) names the social and responsive nature of such intellectual work, new and experienced scholars alike must consistently demonstrate this ability to effectively “join the conversation” in order to move it in a new direction through writing. In this specific way, the activity of source integration allows the scholar to validate and/or create intertextual relationships (Bazerman, 2003) and, in the process, earn the privileged role of knowledge crafter (Kellogg, 2008) and disciplinary expert (Tardy, 2009).

Perhaps driven in part by the tradition and prevalence of this knowledge production activity among faculty researchers who are engaged in classroom instruction and perhaps also by the commonness of source-based writing in many professional fields beyond the academy (e.g. white papers, blogs, market research, policy papers, memos, briefs), undergraduate and graduate students in U.S. institutions of higher education are often required to complete several source-based assignments or projects (e.g. writing-to-learn tasks, course-based assessments of content knowledge, and both primary and secondary research papers across the disciplines) throughout their academic careers (including masters theses and doctoral dissertations).

Given this critical role of source-based writing in academic research endeavors, and the ubiquity of source-based writing across the disciplines, this paper aims to examine one common feature of source integration writing activity — paraphrase — in order to better understand acceptable and effective approaches to accomplish and teach this intellectual task. Specifically, drawn from a larger corpus study, this paper examines the language borrowing practices of expert academic writers, including a focus on disciplinary differences, with the goal of informing the teaching of source-based writing in higher education.
Literature review

The complexity and importance of source integration activity among expert writers notwithstanding, the pedagogical treatment of source-based composing as presented to university students is narrowly focused on the development of three skills—summary, direct quotation and paraphrase. Of these three skills, summary writing, defined as the distillation of a source text into the most relevant points for a new audience and purpose (Doolan & Fitzsimmons-Doolan, 2016), is conceptualized as a process of deletion, generalization and construction (Kintsch & van Dijk, 1978). Direct quotation, in contrast, is generally understood to be the verbatim representation of the language of a source text. Finally, the skill of paraphrase is commonly presented as the representation of a specific source passage using one’s own words and sentence structures while maintaining the meaning of the original source. The estimated number of transformations (linguistic and otherwise) entailed in paraphrase writing range from 25 (Bhagat & Hovy, 2013) to over 100 (Milićević, 2008).

Among expert writers, knowing how, when, and why to effectively use these three techniques allows for the successful production and transformation of knowledge. Through source integration writing, expert writers extend and apply existing theoretical and empirical frameworks, establish credibility by demonstrating familiarity with the literature (and the source-use conventions within a given discipline), build arguments for gaps in the knowledge of a field, and anticipate potential objections.

With regard to the teaching of effective source-based writing, the challenges have been well documented (Hirvela & Du, 2013), particularly since tasks such as paraphrase and summary require learners to simultaneously execute a number of complex cognitive and linguistic activities (Flowerdew & Li, 2007; Kirkland & Saunders, 1991). Research relevant to the prerequisites for effective paraphrase alone reveal that writers need to possess a minimal level of reading comprehension (Roig, 2001; Wette, 2010), a lexicon large enough to produce adequate synonyms (Bhagat & Hovy, 2013), familiarity with clause and phrase-level linguistic transformations (Milićević, 2008), knowledge of the genre of the source text (Bloch, 2012; Jameson, 1993; Ji, 2012), logical inference (Yamada, 2003), content knowledge (including discipline-
specific terminology), and rhetorical knowledge that is sophisticated enough to discern how a single synonymic substitution might subtly impact tone or ethos or readability.

Beyond the high cognitive load associated with paraphrasing, the teaching and learning of paraphrase as a skill for source-based writing is often overshadowed by university-level (and, therefore, classroom-based) concerns with plagiarism; for, though all three skill areas — summary, direct quotation, and paraphrase — can be associated with inappropriate textual borrowing, paraphrase seems to occupy a special place relative to plagiarism as the terms “unsuccessful paraphrase” and “plagiarism” are nearly synonymous. A mooring to plagiarism has resulted in a source-based writing pedagogy that is highly concerned with warnings and admonitions regarding what not to do when paraphrasing (i.e., plagiarizing) rather than what to do. Reflecting this conflation, paraphrase research (in composition, second language writing, and computational linguistics) along with the majority of existing published paraphrase pedagogies have contributed to the current relationship between paraphrase and plagiarism with titles like, “Preventing unintentional plagiarism: A method for strengthening paraphrase skills” (Walker, 2008) and “Plagiarism meets paraphrasing: Insights for the next generation in automatic plagiarism detection” (Barrón-Cedeño et al., 2013).

Similarly, writing handbooks designed to guide less-expert writers toward the acceptable standards of writing in their professions/fields commonly anticipate and answer the student question, “Why do I need to paraphrase?” with, “First, and most importantly, you must avoid plagiarism, or copying words and ideas from sources into your writing without using quotation marks or giving credit to the author” (Fitzpatrick, 2011, p. 198).

Perhaps as a result of this framing of paraphrase in a context of plagiarism, multiple calls for effective (and empirically verifiable) paraphrase pedagogies persist across composition and L2 writing studies (Choi, 2012; Currie, 1998; Howard et al., 2010; Hu, 2015; Jackson, 2006; Keck, 2006; 2010; McInnis, 2009; Pecorari, 2015; Petrić, 2007; Polio & Shi, 2012; Schwabl et al., 2014; Storch, 2012; Stratakis-Allen, 2008).

In spite of these calls, “proven” paraphrase pedagogies grounded in empirical and theoretical work remain scarce. Significantly challenging
the research into effective and acceptable paraphrase writing practices (and pedagogies) is the fact that a precise and widely agreed upon definition of acceptable academic paraphrase does not exist (Dras, 1999; Eberle, 2013; Fisk & Hurst, 2003; Pecorari, 2013; Polio & Shi, 2012; Roig 2001; Rus et al., 2011; Schwabl et al., 2014; Shi, 2010; Uemlianin, 2000; Vila et al., 2014; Yamada, 2003). This lack of definitional consensus among researchers appears most often due to uncertainty regarding what constitutes appropriate levels of language borrowing by the paraphraser from the source passage (Eberle, 2013; Schwabl et al., 2014), and is particularly apparent in research on faculty perceptions of acceptable source use, such as Pecorari and Shaw’s (2012) study in which eight faculty participants reached unanimous consensus on only one of 40 requested judgements of acceptable levels of language borrowing.

Beyond these definitional or threshold challenges for researchers interested in better understanding paraphrase writing, most paraphrase research has narrowly focused on non-expert writers —students—and/or classroom-based interventions. Such studies include examinations of the linguistic dimensions of student paraphrase (Keck, 2006; 2010; 2014; McInnis, 2009; Setoodeh, 2015; Shi, 2004; Sun, 2012), the technologies that can assist students with linguistic transformations inherent in paraphrase writing (Miličević, 2008), and rhetorical theories such as Burke’s pentad (Arrington, 1988) and imitatio within the classical rhetorical curricula (Kalbfleisch, 2016; Terrill, 2011).

While paraphrase research focused on student populations and classroom technologies shed light on early stages in the development of paraphrase proficiency and the interventions that researchers, teachers, and experts in computer assisted language learning technologies are already producing, what is needed is a solid research foundation into the linguistic and rhetorical dimensions of expertly produced paraphrases (ideally drawn from published disciplinary writing). Without an empirically-grounded understanding of what expert writers across the disciplines do with paraphrase in the activity of weaving knowledge together, it has become difficult to find consensus for key, relevant terms among researchers and to scaffold/teach necessary aspects of paraphrase writing. Questions and issues left unanswerable at this time include how many words (overall and/or in a row) can be carried over from the source text to the resulting paraphrase, how frequently para-
phrases should/could be incorporated in a new, “original” text, how long a resulting paraphrase should/can be, what role paraphrase plays in relation to the other available source-based writing techniques. Recognizing these gaps, researchers have called for published text analyses of source-based writing (Polio & Shi, 2012), corpus studies that can be used to inform paraphrase pedagogies (Hu, 2015) and examinations of paraphrases produced in real-operating conditions (Stratakis-Allen, 2008). Such examinations of expert paraphrase practice could lead to discipline-specific working definitions of paraphrase and could clarify the confusion surrounding what counts as an acceptable paraphrase and what doesn’t (Bloch, 2012).

Thus, rather than looking to the paraphrase practices of less experienced student writers, the present study turns to the writing of successful, expert academics — and, specifically, particular aspects of their language borrowing practices reflected in writing handbooks and published texts across the disciplines — in order to empirically determine what appear to be acceptable and effective features of paraphrase writing.

Drawn from a larger corpus-based study (Michiels, 2019) that examines exemplar paraphrases and their corresponding source passages in writing handbooks, as well as paraphrases and source texts located in published and award-winning academic articles from three disciplines, the present study aims to address two research questions with potential pedagogical implications: 1) What can a linguistic analysis of language borrowing patterns tell us, empirically, about the paraphrase practices of expert writers? 2) What can this same linguistic analysis tell us about disciplinary differences in paraphrase practices?

Methods

Creating corpora of paraphrases and corresponding source passages

One approach to empirically determine acceptable language borrowing practices among expert writers is to linguistically/textually analyze sets of resulting paraphrases against their corresponding source passages. Such comparisons allow researchers to analyze linguistic dimensions such as levels of language borrowing, types of lexical and
syntactic transformation, changes in length from the source passage to the paraphrase, the maintenance or substitution of key words, and the use of direct quotation within the paraphrase. Further, on the rhetorical side, analyses of the content of paraphrases against source texts (particularly in the context of published writing) allows researchers to examine acceptable patterns in this source-based writing activity, including semantic changes, additions, omissions, and other stylistic alterations. This paper focuses solely on the patterned, linguistic aspects of expertly produced paraphrases identified through the larger study’s corpus of paraphrase/source passage pairs as described below.

The first two corpora comprise the exemplar paraphrases and their corresponding source passages drawn from writing handbooks in their treatment of paraphrase as a skill for source-based writing as presented to, generally, undergraduate and graduate university students. Exemplar paraphrases are typically presented in writing handbooks using a short source passage and one or two example paraphrases, often with the purpose of distinguishing acceptable and unacceptable attempts (see Figure 1 for an example from Troyka & Hesse, 2007). For the purpose of this study, data collection from handbooks involved gathering source passages and their corresponding exemplar paraphrases (i.e., paraphrases marked or indicated to be acceptable).
**Corpus 1: L1 Handbooks**

The first corpus was built using the exemplar paraphrases and accompanying source passages drawn from a collection of writing handbooks and guides designed for English L1 writers (the designation L1 refers to students who use English as a first or dominant language) in higher education at the writing center of a large research 1 (R1) university located in the mid-Atlantic United States. The writing center was chosen as a location for this material because its library of writing guides and handbooks (several hundred) far surpassed the number of resources on this topic contained in the university library and in the libraries of the surrounding universities. Handbooks were chosen because they are a resource that teachers frequently use or encourage students to
use in courses where source-based writing is a learning outcome (Stratakis-Allen, 2008,). The 128 handbooks yielded 107 exemplar paraphrase/source passage pairs. These handbook samples will be referred to as the L1 Handbooks.

**Corpus 2: L2 Handbooks**

The second corpus was built using exemplar paraphrases and source passages from a collection of writing handbooks and guides found in the instructional library of the same university’s international student pathway program. Resources within this library were primarily designed for students using English as a second or additional language (referred to in this study as L2). Nineteen of these handbooks contained treatments of paraphrase which provided 32 exemplar/source passage pairs. Corpus 2 will be referred to as the L2 Handbooks.

**Paraphrase/source passage pairings from published journal articles**

The remaining three corpora were constructed using award-winning articles published in peer-reviewed journals: *College Composition and Communication (CCC)*, the *Journal of Second Language Writing (JSLW)*, and a number of different journals published under the Institute of Electrical and Electronics Engineers (IEEE). These three publications were chosen because they represent different disciplines, different citation styles (*CCC* uses MLA; *JSLW* uses APA, and IEEE articles use IEEE), and broadly different approaches to knowledge-making and source-based writing.

The ten articles that were used to build the *CCC* corpus had all won the Richard Braddock Award, which is granted annually to “the outstanding article on writing or the teaching of writing” (National Council of Teachers of English, n.d.), and spanned 2009-2018. The ten articles used to build the *JSLW* corpus had all won the “Best Article Award” and spanned 2006-2017. Finally, the ten articles from IEEE had all won the Donald G. Fink Award for “the most outstanding survey, review, or tutorial paper” (Institute of Electrical and Electronics Engineers [IEEE], 2019). Recipients of the Donald G. Fink Award, which was discontinued in 2016, spanned engineering sub-disciplines (and journals), including nanomagnetics, wireless communications technologies, and imaging sensor technologies. Articles from IEEE spanned 2007-2016.
Corpora 3, 4, and 5: Published academic writing across the disciplines

Whereas handbooks generally include original source passages in their explanations and examples of paraphrase so that the writer can follow the transformation of the text, the published writing of expert academics will point to a given source, but not always to the specific passage within that source. For instance, MLA citation style (present in the CCC corpus) requires an author to provide a page number or local indicator whenever a paraphrase is employed. APA citation style, used by authors in the JSLW corpus, does not require authors to provide a page number or local indicator while paraphrasing. Similarly, IEEE citation style does not require authors to provide a page number while paraphrasing. As a result, finding instances of paraphrase in published academic writing, and then finding the precise source passage upon which a given paraphrase was based (particularly in articles whose citation style does not mandate the use of page numbers with paraphrase) can be considered as a high-inference activity; that said, this approach to research into paraphrase practices is not without precedent and was accomplished by Howard et al. (2010) whose attempt to locate paraphrases and the corresponding source passages was successful in “almost every case in all 18 papers” coded for that study (p. 183) and by Pecorari (2013) whose analysis of dissertations led to the location of 352 of the 464 source passages.

To find potential paraphrases within published journal articles in this study, the articles in the three disciplinary corpora were coded by the first author for all instances of integrated and non-integrated single-source citations. The sources referenced within these citations were then gathered and reviewed to locate the potential source passage. A set of inclusion and exclusion criteria were then applied to select paraphrases linked to specific, locatable passages.

Inclusion criteria for paraphrase/source passage pairings in Corpora 3, 4, and 5:

1. Paraphrase corresponds to the content of a specific source passage.

2. Paraphrase refers to a single author but draws information from up to three different, clearly discernible locations within that
author’s work. In other words, the paraphrase points to various content within the same paragraph, the same page, or elsewhere in the same original source text.

3. Paraphrase contains all or most of the idea units in a specific source passage, even when other source passages contain a smaller number of idea units. To illustrate, a sample paraphrase with source passages is provided. In the example, the resulting paraphrase contains the most ideational overlap with Potential Source Passage 1, and therefore it was selected as the basis for comparison (though another section of the source text, Potential Source Passage 2, also contained some of the idea units found in the paraphrase).

Example of Inclusion Criterion #3

Resulting paraphrase:

**Following Storch’s (2002) account of interaction patterns with respect to “equality” and “mutuality”, Li and Zhu (2013) examined** wiki Discussion records supplemented with wiki Page and History records, and derived three distinct patterns of interaction in EFL group wiki writing: collectively contributing/mutually supportive, authoritative/responsive, and dominant/withdrawn. (Li & Kim, 2016, p. 26).

Potential Source Passage 1:

Analysis of data from the wiki “Discussion”, “Page”, and “History” modules on each group tab revealed that the three small groups displayed three distinct patterns of online interaction: collectively contributing/mutually supportive, authoritative/responsive, and dominant/withdrawn. (Li & Zhu, 2013, p. 61).

Potential Source Passage 2:

Analysis of data indicated that the three groups demonstrated three distinct patterns of interaction: collectively contributing/mutually supportive, authoritative/responsive, and dominant/withdrawn. (Li & Zhu, 2013, p. 68).

4. Paraphrase of the original source text is clear enough for isolation even if the writer’s ideas or representation of the source text is drawn from more than one location.
Exclusion criteria for paraphrase/source passage pairings in Corpora 3, 4, and 5. In other words, paraphrases that did the following were not included in the analysis:

1. Paraphrase includes a direct quotation of four or more words.
2. Paraphrase draws primarily or solely from the title of the original source text.
3. Paraphrase refers to a single keyword, key term, or key concept.
4. Paraphrase could have been written based on multiple passages in the source.
5. Paraphrase so significantly distills or generalizes the content of the source passage as to appear to be a summary.
6. Paraphrase appears to cobble together information from four or more distinct sections of the source passage. Note: paraphrases that are constructed from two, and sometimes three, locations were, however, included in the corpus, as was common practice among handbooks paraphrasing longer source passages.
7. Paraphrase semantically diverges from the source passage or appears to make significant inferences.
8. Paraphrase has initial and terminal boundaries that are ambiguous.

The IEEE corpus contained 44 paraphrase/source passage pairs. The CCC corpus contained 15 paraphrase/source passage pairs. The JSLW corpus contained 40 paraphrase/source passage pairs.

Identifying language borrowing patterns

Unique Links. Across all corpora, an analysis of language borrowing was conducted using an adaptation of the taxonomy developed by Keck (2006). Keck’s taxonomy begins with a source passage and a corresponding paraphrase. Before calculations of borrowing are performed, the paraphrase is stripped of any reporting verbs or phrases (for example, “According to Thompson” or “Klein argues”) so as to establish the cleanest possible textual boundaries for the analysis of borrowed content from each source passage to each paraphrase. Two additional concepts are then used to calculate degrees of language borrowing:
unique links and general links. Unique links, according to Keck, are defined as “individual lexical words (i.e., nouns, verbs, adjectives, or adverbs, or exactly copied strings of words used in the paraphrase that (a) also occurred in the original excerpt but, (b) occurred in no other place in the original text” (p. 266). Unique links are counted as instances of borrowing from the source passage.

Keck (2006) provides the following example in which unique links are bolded:

**Original**

Women have less work experience, less seniority, a lower rate of unionization and so on (Samuelson, paragraph 7).

**Attempted Paraphrase**

**Women have less** job experience, **less seniority**, and a low **rate of unionization**. (p. 266).

According to Keck’s (2006) taxonomy, the bolded language represents unique links — content words or combinations of content and function words that appear in the source text and also in the paraphrase. Notice that the function word “of” is counted as part of a unique link because it is part of the larger phrase “rate of unionization,” a phrase that occurs in the source passage. In contrast, according to Keck, “general links” are defined as “lexical words used in the paraphrase that occurred in the original excerpt but that also occurred elsewhere in the original text” and are not counted as instances of borrowing (p. 267). It appears that Keck created the “general link” category to account for a writer’s need to recycle key words and key concepts upon which the base meaning of a paraphrase is built. For example, if a short source passage about grandparents uses the term “grandparents” three or four times, this term would be considered a general link and would not count toward language borrowing if a writer were to incorporate the word “grandparents” into a paraphrase of that passage.

Keck (2006) then created four paraphrase classifications determined by the percentage of unique links present in a given paraphrase. Our study uses these classifications, which are structured as follows: For a “Near Copy,” 50% or more of the resulting paraphrase language
is contained in unique links. For a “Minimal Revision” 20-49% of the language of the resulting paraphrase is contained in unique links. For a “Moderate Revision” 1-19% of the language of the resulting paraphrase is contained in unique links, and for a “Substantial Revision” none of the language of the resulting paraphrase is contained in unique links.

Slight adaptations were made to Keck’s (2006) taxonomy for the purposes of this study. First, because many of the sources used in the present study were full articles or books, general links were limited to the confines of the source passage within the longer text. Second, groups of three or more function words that appeared to be reproduced from the source passage were counted as unique links. Third, unique links that were modified from singular to plural (or vice versa) in the paraphrase were counted as unique links, as were unique links that were made into hyphenate constructions in the paraphrase. The number of paraphrase/source passage pairings across corpora can be found in Table 1.

Table 1
Number of Paraphrase/Source Passage Pairings Across Corpora

<table>
<thead>
<tr>
<th>Corpora</th>
<th>Number of paraphrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 Handbook corpus</td>
<td>107</td>
</tr>
<tr>
<td>L2 Handbook corpus</td>
<td>27</td>
</tr>
<tr>
<td>CCC corpus</td>
<td>15</td>
</tr>
<tr>
<td>IEEE corpus</td>
<td>44</td>
</tr>
<tr>
<td>JSLW corpus</td>
<td>40</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>233</strong></td>
</tr>
</tbody>
</table>

Note. L1 = English as the first language, L2 = English as a second or additional language, CCC = College Composition and Communication, IEEE = Institute of Electrical and Electronics Engineers, JSLW = Journal of Second Language Writing.

Source: Own elaboration.

Word strings

Finally, paraphrases within the five corpora were analyzed for the presence of strings of four or more unquoted words found identically in a row in the source passage. The use of word strings as an added layer of linguistic analysis stems from Roig’s foundational work assessing the paraphrases of students and university instructors.
In the second of three experiments, Roig (2001) asked 109 professors (from various disciplines) to paraphrase a (decontextualized) two-sentence passage from an academic journal article from the field of psychology with a Flesch-Kincaid readability level of 15.6, which, Roig noted, was 2.5 grade levels above the optimal level for undergraduates. Roig coded these paraphrases for strings of five, six, seven, or eight consecutive words (strings of four or fewer copied words were not accounted for). Results of the coding indicated that 30% of the paraphrases contained five-word strings, 22% contained six-word string, 18% contained seven-word strings, and 9% contained eight-word strings (p. 315). In addition, 24% of professors were found to have distorted the meaning of the source text (though, according to Roig, these distortions were often minor). In the third experiment, 43 psychology professors paraphrased the complex passage from the psychology journal, and 64 psychology professors paraphrased a less complex passage on astrology. Twenty-six percent of professors who paraphrased the complex passage appropriated strings of five words, and 9% copied strings of eight or more words. Results were considerably different for the easier passage: no professors borrowed strings of seven or eight words, and only 3% borrowed strings of five words (p. 319).

In short, Roig found that expert writers paraphrasing texts taken from their field tended to borrow slightly less than professors paraphrasing unfamiliar source material. As readability went up, professors tended to borrow much less language in the paraphrases. Though Roig chose to ignore strings of four or fewer words, the present study identified strings of four words, given that writing handbooks tend to mandate the use of quotation marks for strings of more than two or three words.

Results
Paraphrase data according to Keck’s (2006) taxonomy and word string borrowing are provided in Table 2.
Table 2

Characteristics of paraphrases across corpora

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>L1 (107)</th>
<th>L2 (27)</th>
<th>CCC (15)</th>
<th>IEEE (44)</th>
<th>JSLW (40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substantial revision (% of corpus)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7.5</td>
</tr>
<tr>
<td>Moderate revision</td>
<td>46.7</td>
<td>29.6</td>
<td>66.7</td>
<td>25</td>
<td>22.5</td>
</tr>
<tr>
<td>Minimal revision</td>
<td>49.5</td>
<td>55.6</td>
<td>33.3</td>
<td>45.5</td>
<td>47.5</td>
</tr>
<tr>
<td>Near copy</td>
<td>3.7</td>
<td>14.8</td>
<td>0</td>
<td>29.6</td>
<td>22.5</td>
</tr>
<tr>
<td>Overall average (unique links)</td>
<td>23.3</td>
<td>27.8</td>
<td>19.1</td>
<td>38.0</td>
<td>31.9</td>
</tr>
<tr>
<td>Percentage of paraphrases containing a string of four or more consecutively copied words without quotation marks</td>
<td>29.0</td>
<td>25.93</td>
<td>6.7</td>
<td>50</td>
<td>37.5</td>
</tr>
</tbody>
</table>

Note. L1 = English as the first or dominant language; L2 = English as a second or additional language; CCC = College Composition and Communication; IEEE = Institute of Electrical and Electronics Engineers; JSLW = Journal of Second Language Writing.

Source: Own elaboration.

Corpus 1: L1 Handbook paraphrases

The language borrowing mean for exemplar paraphrases within the L1 Handbook corpus was 23.3%, meaning, on average, that less than a quarter of the overall language of modeled paraphrases consisted of unique links found within the source passage. That said, none of the paraphrases within the L1 Handbook corpus were classified as substantially revised (a substantially revised paraphrase would contain no unique links). In other words, every paraphrase within the L1 Handbook corpus borrowed at least one unique link from the source passage (in addition to general links). The bulk of the paraphrases fell fairly evenly between moderately revised (46.7%) and minimally revised (49.5%). A small number of paraphrases were classified as near copy (3.7%).

Table 3 provides an illustration of this degree (23.5%) of overall language borrowing and the representation of unique links in both the original source text and the resulting paraphrase provided to student readers in such cases. Underlined words and phrases in the column on the right represent the unique links borrowed from the source passage.
Table 3
Representative example of language borrowing patterns in L1 Handbooks

<table>
<thead>
<tr>
<th>Source passage</th>
<th>Paraphrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual-reality systems can use <em>cyberspace</em> to represent physical <em>space</em>, even to the point that we feel telepresent in a transmitted scene, whether Mars or the deep ocean.</td>
<td>We can achieve the illusion of being present in remote locations, for example the planet <em>Mars</em> or <em>deep parts of the ocean</em>, by using <em>virtual-reality</em> equipment that creates a <em>cyberspace</em> representation of real-world <em>space</em> (Heim 80).</td>
</tr>
</tbody>
</table>


Corpus 2: L2 Handbook paraphrases

The language borrowing mean across paraphrases in the L2 Handbook corpus was 27.8%, meaning slightly more than one quarter of the overall language in a model paraphrase provided to L2 writers was made up of unique links. As was the case with L1 Handbook data, none of the paraphrases within the L2 Handbook corpus were classified as instances of substantial revision. Most paraphrases fell within the minimally revised category (55.6%) while moderately revised paraphrases comprised 29.6% of the corpus, and near copy paraphrases represented 14.8% of the L2 Handbook exemplars. A representative example (27.1% unique link borrowing) from the L2 Handbook corpus can be found in Table 4.

Table 4
Representative example of language borrowing patterns in L2 Handbooks

<table>
<thead>
<tr>
<th>Source passage</th>
<th>Paraphrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is quite obvious that healthy people are happier than unhealthy people. What is now becoming increasingly evident through study is that the reverse is also true: happy people are healthier than unhappy people. It appears that happiness, which simply means having happy thoughts most of the time, causes biochemical changes in the brain that in turn have profoundly beneficial effects on the body’s physiology (p. 208).</td>
<td>Deepak Chopra wrote that we all recognize that people with good health are happier than those whose health is not as good. Research is revealing that the opposite is true as well: Happy individuals enjoy better health than unhappy individuals. Happiness, or the quality of thinking positively most of the time, probably stimulates biochemical reactions in the brain that protect one’s physical health (p. 208).</td>
</tr>
</tbody>
</table>

Note. The bolded language in the paraphrase was not included in the total word count. Source: Fitzpatrick (2011, p. 208).
Corpus 3: CCC paraphrases

The overall mean of language borrowed for paraphrases within the CCC corpus was 19.1%. There were no instances of substantial revision or near copy within the CCC corpus. Most paraphrases fell within the moderate revision classification (66.7%), with the remaining 33.3% of the paraphrases being classified as minimal revisions. The example paraphrase in Table 5 contains 19.1% unique links, which is identical to the overall borrowing average for this corpus.

Table 5
Representative example of language borrowing patterns in CCC corpus

<table>
<thead>
<tr>
<th>Potential source passage</th>
<th>Potential paraphrase</th>
</tr>
</thead>
</table>

“Chapters” are the local units of SVA and are based at universities and post-secondary institutions. With 950 chapters, SVA has a presence at most schools in the nation and even several nations abroad. In 2014—SVA’s seventh year of existence—officials from the organization hope to not only prove that the Post-9/11 GI Bill was worth it and worth expanding further. They also hope to grow their organization even further to accommodate the student veterans that are returning and looking to use their education benefits. SVA has nearly doubled its number of chapters since its conference last year.

Reflecting the rapid growth of veterans on college campuses, the national Student Veterans Association (SVA) has seen the number of its campus chapters multiply exponentially, from fewer than 50 before 2010 to 950 in fall 2014, with membership doubling from 2013-2014 alone (Romney, 347).

Source: Goldblatt (2017, p. 347); Romney (2014, n.p.).

Corpus 4: IEEE paraphrases

The overall mean for language borrowing among the IEEE corpus was 38.0%. No instances of substantial revision were found in this corpus. Twenty-five percent of the paraphrases in this corpus fell into the moderate revision classification; 45.5% were deemed minimally revised, and 29.5% were classified as near copy. In the example in Table 6, 41% of the language of the paraphrase is contained in unique links (this example was closest to the overall corpus mean of 38%).
Table 6
Representative example of language borrowing patterns in IEEE corpus

<table>
<thead>
<tr>
<th>Potential source passage</th>
<th>Potential paraphrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Jordan et al., 1999, p. 413)</td>
<td>(Krishnan, 2010, p. 2540)</td>
</tr>
<tr>
<td>Heating of certain organs or tissues to temperatures between 41°C and 46°C preferentially for cancer therapy is called 'Hyperthermia'. Higher temperatures up to 56°C, which yield widespread necrosis, coagulation or carbonization (depending on temperature) is called 'thermo-ablation' (413).</td>
<td>Note that hyperthermia differs significantly from thermoablation, which employs higher temperatures, up to 56°C, to crudely destroy cells leading to necrosis, coagulation and carbonization (188) (2540).</td>
</tr>
</tbody>
</table>

Source: Jordan et al. (1999, p. 413); Krishnan (2010, p. 2540).

Corpus 5: JSLW paraphrases

On average, 31.9% of the language in the JSLW paraphrases was made up of unique links. The JSLW corpus was the only one of the five corpora which contained instances of substantial revision (7.5% of paraphrases fell under this category). The remaining paraphrases were categorized as follows: 22.5% were deemed moderate revision; 47.5% were minimal revisions, and 22.5% were classified as near copy. In the example in Table 7, 31.6% of the language is contained in unique links, which is nearly identical to the 31.9% overall average in this corpus.

Table 7
Representative example of language borrowing patterns in the JSLW corpus

<table>
<thead>
<tr>
<th>Potential source text</th>
<th>Potential paraphrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Otsuji, 2010, p. 6)</td>
<td>(Belcher, 2014, p. 62)</td>
</tr>
<tr>
<td>Nonetheless, it is ironic that I am categorising and fixing the students’ cultural and ethnic backgrounds in a context where we are aiming to deconstruct Japanese cultural practices. These exchanges disclose my unconscious assumptions that the students have a monolithic fixed origin. Moreover, it indicates not only how invalid it is to think that people belong to one particular culture and language group but also how easy it is to fall into the trap of essentializing cultural background and origin (p. 6).</td>
<td>Interestingly, despite her postmodern self-awareness, Otsuji (2010) admits to sometimes falling into the essentializing trap of categorizing her own Japanese-language students in Australia by their apparent ethnicity (62).</td>
</tr>
</tbody>
</table>

Note. The bolded language in the paraphrase was not included in the total word count.
Source: Belcher (2014, p. 62); Otsuji (2010, p. 6).
Discussion

Language borrowing patterns of expert writers

This study set out, first, to empirically investigate the language borrowing patterns among expert writers in order to better understand paraphrase as a technique for source integration activity. When analyzed for two specific language borrowing patterns—unique links and word strings—data from the full corpus analysis suggest that 27.7% of expertly produced paraphrases contain unique links and 37.3% contain unquoted word strings of four or more words identical to those found in the source passage.

A representative example of an expert-produced paraphrase that most closely aligns with this overall finding (27.2% average unique links) from the full corpus can be seen here from the IEEE corpus:

Source Passage

As concerns the compass, Wehner and coworkers have shown in a series of ingenious experiments that *Cataglyphis*, like honey bees, relies primarily on the polarization (POL) pattern of the sky as a compass reference (Lebhardt et al., 2012, p. 526).

Paraphrase

In nature, the desert ant *Cataglyphis fortis* uses this polarization pattern of the sky to aid its navigation to and from home [25] (York et al., 2016, p. 1453).

Applying Keck’s (2006) classification system for determining the level of a writer’s revision when producing a paraphrase, data from this study show that an overwhelming majority of expert writers do not produce substantially revised paraphrases (i.e. paraphrases that borrow no language from the source text other than general links). Across the handbook corpora (Corpus 1 and 2) which contained 134 instances of modeled (i.e. “acceptable”) paraphrase, not a single example provided to student writers in this context offered a paraphrase that could be classified as a substantial revision. This pattern held true for expert academic writing from across the disciplines (Corpus 3, 4, and 5) where, of the 233 instances of paraphrase analyzed, only three instances could be classified as substantial revisions. The presence of
unique links and word strings in these data suggest that expert writers, in the activity of source-based writing, both produce and likely find it acceptable for other expert writers to produce paraphrases that borrow language from source texts.

When handbook paraphrase guidelines for student writers such as the ones included in Table 9 are compared with the language borrowing patterns found in the paraphrase examples of expert writers, the data suggest that handbooks offering categorical statements prohibiting any amount of language borrowing (i.e., calls for students to replace “all the words” or to put the source passage “entirely” in their own words) or prohibiting the use of word strings (i.e. deeming any use of word strings as plagiarism) do not reflect the acceptable paraphrase practices of expert writers and that this is the case across several disciplines. In fact, the contradiction between what student writers are told to do and what expert writers appear to accept and themselves do in their own academic writing extends to the language borrowing practices of the handbook authors themselves.

Table 8

<table>
<thead>
<tr>
<th>Handbooks’ treatment of acceptable language borrowing in paraphrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prohibiting language borrowing</td>
</tr>
<tr>
<td>“Note that you replace all the words, not just some of them, for a paraphrase” (Alfano &amp; O’Brien, 2011, p. 125).</td>
</tr>
<tr>
<td>“An acceptable paraphrase needs to be entirely in your own words” (Ramage et al., 2012, pp. 342-343).</td>
</tr>
<tr>
<td>“When you paraphrase a passage, start by identifying key words and phrases and substituting synonyms for them” (Greene &amp; Lidinsky, 2008, p. 126).</td>
</tr>
</tbody>
</table>

*Note. Italicics are added for emphasis.*


Beyond potential implications for handbook guidelines in their treatment of language borrowing in paraphrase writing, the lack of exemplar paraphrases across the corpora that could be classified as
substantially revised appear to support Keck’s (2006) decision not to count repeated keywords (or general links) in assessments of overall language borrowing. Though additional research is needed to more fully understand what happens to keywords as they are linguistically and rhetorically repurposed in paraphrased writing, data from this study suggest that expert writers tend to retain rather than replace or avoid keywords from original source texts as they paraphrase. Larger corpus-based examinations of the language borrowing patterns of expert academic writers are needed to verify and/or challenge the lack of substantially revised paraphrases found across all corpora in this study.

**Pedagogical implications for developing writers**

The findings of this study point towards several important pedagogical lines of inquiry. First, a fuller empirical description of the language borrowing patterns present in the source integration activity of expert writers can contribute to a developmental model and approach to learning how to effectively and acceptably work with source material. A data-informed, developmental model that takes the real patterns of language borrowing practices among expert writers as the end point would allow for a less contradictory set of paraphrase guidelines for developing writers and would support a pedagogical approach that could more accurately anticipate the role of source integration work in the activity of learning and of becoming an expert.

A developmental approach would support existing research that has tracked student paraphrase performance over time and indicated a developmental arc associated with paraphrase writing that is likely marked by dips and recoveries in performance (i.e., U-shaped learning). Such a pedagogy would also align with Rogers’ (2008) longitudinal analysis of 200 source-based papers from 40 students across five years which showed that while writers’ ability to effectively work with source material increased across their undergraduate years and into their first year of graduate school or professional life, that growth was uneven and nonlinear (as was found to be the case with other dimensions of participants’ writing abilities).

A developmental understanding of this complex academic writing activity might also account for an increase of inaccurate citations which have been shown to appear as students learn to balance the demands
of the linguistic and rhetorical complexity of paraphrase, particularly in the context of actual writing assignments, as opposed to decontextualized paraphrase practice (Wette, 2010). Further, it may account for novice paraphrasers who may need to pass through a developmental stage in which misrepresentations of the meaning of the source become more frequent as they rely less heavily on the language of the source text (Storch, 2012; Wette, 2010). In terms of language borrowing, it has been hypothesized that the complexity of paraphrase writing may require students to pass through stages in which inappropriate borrowing practices—such as Howard’s (1995) patchwriting—may be “inevitable” (Hayes & Introna, 2005). McDonough et al. (2014) even argued that the linguistic dimension of paraphrase development may require students to move from borrowing fewer long strings of words from the source text to more short strings, resulting in a temporary increase in the total amount of borrowed language. In other words, these studies suggest that students may need to temporarily borrow more language as they improve at paraphrase writing. A comprehensive and developmental perspective on paraphrase use as a component of source-based writing and which is empirically informed by the acceptable patterns and practices of expert writers has important implications for college-level pedagogies.

The impact of disciplinarity on word-string borrowing represents a key complexity when attempting to situate paraphrase within a developmental continuum. For instance, Roig (2015) points out that for journals in some branches of the sciences (such biomedical engineering), directly quoting is not a practice common to the genre, yet the language of such disciplines, particularly word strings used for single concepts, can be “very difficult to paraphrase” (p. 13). In a guide to ethical writing produced for the Office of Research Integrity in the U.S. Department of Health and Human Services, Roig provides an example source passage and exemplar paraphrase drawn from the biomedical journal *Science*. He points out that some language contained in the source passage (for instance, terms contained in word strings such as “mammalian histone lysine methyltransferase”) cannot be easily paraphrased without potentially altering the meaning of the source. In a somewhat comforting turn of events, Roig, a key figure in research on academic paraphrase, admits that the original paraphrase produced for an earlier edition of the guide contained a misrepresentation of the source passage. Thus,
an exemplar paraphrase written by the author of several landmark studies on academic paraphrase and designed to show how much more language needs to be borrowed in some instances still misrepresented the source passage, and likely because the paraphrase had not borrowed enough language from the source. The recognition that even expert writers may need to borrow more language from an original text when paraphrasing should serve as a point of encouragement for anyone learning to paraphrase.

Data from this study indicate that the kind of proficiency that paraphrase pedagogies (and pedagogical research) is oriented toward—i.e., assumed increases in instances of substantial revision as a goal or common arc of development—may not align with the proficiency and practice of expert writers when they paraphrase. This potential mismatch is most apparent in pedagogical paraphrase research that uses Keck’s (2006) taxonomy. For example, in Storch’s (2012) study of 26 English-L2 students in Australia, 31% of the students’ preintervention paraphrases were classified as substantially revised and an astonishing 45% of post-intervention paraphrases were substantially revised. Keck’s (2006) study found that substantially revised paraphrases were present in 10% of the summaries written by undergraduate study participants. Similarly, in McInnis’s (2009) study, which analyzed 35 paraphrases from a group of four English L1 college students and five English L2 international students, three of the English L2 participants (out of four) produced substantially revised paraphrases (the English L2 learners in McInnis’s study were the only participants to produce substantially revised paraphrases). Importantly, McInnis found that paraphrases which borrowed the least language (in terms of unique links) from the source passage also tended to be evaluated as least appropriate or acceptable when considered against other criteria (such as maintaining key ideas, avoiding summary, and including attribution). These studies and the data here suggest that source-based writing pedagogies and pedagogical research using interventions which are designed to push students toward producing more substantially revised paraphrases may need further consideration as the underlying assumptions about the patterns of acceptable paraphrase embedded in such interventions may be steering students away from necessary and normal development and may be out of alignment with empirical evidence of expert practice.
Language borrowing across the disciplinary corpora

In addition to empirically examining the collective, overall patterns of language borrowing present in the paraphrase practices of expert academic writers, this study set out to better understand any potential discipline-based differences in such patterns. Though the corpora for this study were fairly small, data indicate that discipline-driven variations may exist. For instance, with regard to unique links, the degree of overall language borrowing was shown to vary among the disciplines represented in this study. Specifically, expert writers representing the humanities demonstrated less willingness to produce or accept unique links in the language of paraphrases within academic publications (on average, 19.8% of resulting paraphrase language in CCC was contained in unique links) than was the case among expert writers representing a social science orientation (on average, 31.9% of JSLW paraphrase language was contained in unique links). Further, expert writers representing the sciences in this study demonstrated the highest level of acceptability for unique links to appear in the paraphrased language of published scholarship (on average, 38% of IEEE paraphrase language was contained in unique links). Taken together, this range of acceptability with regard to the presence of borrowed language from source passages in resulting paraphrases among expert academic writers from different disciplines might be seen to reflect a broader continuum of acceptability that reflects epistemologies and knowledge production systems inherent in different fields of study and reflected in source-based writing such as paraphrase.

Disciplinary differences between the humanities oriented corpus (CCC) and the science oriented corpus (IEEE) were particularly apparent in two areas of analysis: the percentage of paraphrases classified as “Near Copy” and the percentage of paraphrases borrowing (or appearing to borrow) strings of four or more words without quotation. The CCC corpus did not contain any paraphrases classified as Near Copy (i.e., paraphrases in which 50% or more of the language was contained in unique links or borrowed language). A little under one-third (29.6%) of the paraphrases in the IEEE corpus were classified as Near Copy. Moreover, only 6.7% of the paraphrases in the CCC corpus contained strings of four or more words copied (it seems) from the source passage without quotation. Exactly one half of the IEEE paraphrases contained
such word strings. These findings further point to a discipline-driven continuum of acceptability for borrowed language as expert writers within engineering disciplines covered by the IEEE umbrella appear to be more willing to produce and accept borrowed language in the form of word strings than is the case among expert academic writers in the humanities.

On a related note, overall, data from this study show varying degrees of paraphrase presence in the writing of expert academic writers across those disciplines included. For example, paraphrasing seemed to be used more frequently as a technique for source integration work in the IEEE articles (44 instances were found) than in the CCC articles (only 15 instances were found). A key question becomes: How might disciplinary differences in source-based writing reconcile with often unrecognized disciplinary values (such as the preference for summary over paraphrase, the preference for more or less language borrowing in paraphrase, and the preference for sparse use of paraphrase) and how might faculty account for those differences in writing instruction? Larger corpus studies of source-based writing patterns can bring into sharper focus this potential continuum of acceptability within and across disciplines.

**Plagiarism policies and paraphrase**

In addition to these (and other) important pedagogical misalignments between practice and prescriptions are the serious consequences of misalignments between expert practice, pedagogical prescription and plagiarism policies. For example, what becomes of policies for first-year university writers faced with the task of producing source-based writing using a variety of challenging sources and genres, some of which (academic journal articles, edited collections, etc.) are being encountered for the first time?

If the production of unsuccessful paraphrases, such as those deemed instances of “patchwriting” reflects a student’s presence in a given developmental stage (Liu et al., 2018), plagiarism policies oriented toward zero tolerance may stand in the way of student skill development. Detect, deter, and discipline policy orientations may be problematic for several reasons: they may assume that students can immediately accomplish paraphrases on par with expert practice; they may inadvertently close
out the possibility of failure at paraphrase as a viable step in the learning process (Abasi & Akbari, 2008), and they may build fear and anxiety, two emotions that likely hinder both intellectual curiosity (Eyler, 2018) and creativity (Bloch, 2012). It is thus of little surprise that students in perception studies have reported not wanting to attempt paraphrase in source-based writing projects because it is not worth the risk (Hirvela & Du, 2013). Finally, beyond the negative potential impact on individual writers, plagiarism policies may set and enforce a standard for acceptable language borrowing that is not empirically grounded in the actual paraphrase practices of expert writers in various disciplines.

The potential discrepancy between plagiarism policy and expert practice is further complicated by emerging research into text recycling (Anson & Moskovitz, 2021; Haapanen & Perrin, 2020; Hall et al., 2021; Moskovitz, 2021). Text recycling refers to an author’s reuse of material (either verbatim or “substantively equivalent in form and content”) taken from their own previously published work (and used without quotation). A study of 400 published STEM articles from four disciplines found that around three sentences per analyzed paper contained material recycled from an author’s previous publication and that authors frequently alter recycled text in ways that strongly resemble patchwriting (Anson & Moskovitz, 2021). Determining the lines between minimally altered text recycling (which seems to be a common practice in some disciplines, and is both accepted and required in many workplaces) and paraphrase is yet another area in need of further research (Moskovitz, 2021). Concurrent or retrospective think-aloud protocols could shed light on how expert writers conceptualize the difference between recycling and paraphrasing.

This is by no means to say that existing plagiarism policies are misguided or unnecessary, but there appears to be a need to start investigating how plagiarism policies might be preventing or hindering students from passing through certain stages of source-based writing development, particularly with respect to paraphrase.

The development of paraphrase pedagogies capable of furnishing students with accurate, discipline specific, and more easily operationalized conceptualizations of paraphrase will require a robust, multidimensional, and multidisciplinary research agenda. Such a research agenda could use further analyses of student paraphrase practices (and
perceptions), plagiarism policies, existing pedagogies, and expert practice (and perceptions) as a means to triangulate a deeper understanding of this concept, and one that could directly inform classroom pedagogy. It seems that a research stream dedicated to expert paraphrase practice would be an ideal starting point because of this applicability to classroom pedagogy. The fields of writing and rhetoric could provide discourse-based interview and think-aloud studies of experts writing for publication, longitudinal studies of paraphrase development, and analysis of the rhetorical functions and dimensions of paraphrase in published writing (which have received very little treatment). Researchers in sub-disciplines within the field of linguistics could work to create larger paraphrase corpora, additional methods for identifying paraphrases within published writing, and additional analysis (and tools for the analysis) of how key constructions within academic writing, such as complex noun phrases (Biber & Gray, 2010), function and transform as authors move from a source passage to a paraphrase. A more secure basis of knowledge about expert practice, and one informed by disciplinary difference, could then open the door to better informed experimental studies rooted in comparative pedagogical interventions.

The present study has several limitations. The corpus of expert writers’ paraphrase/source passage pairs was fairly small (n = 233). This was particularly true for the disciplinary paraphrases. The CCC corpus, for example, had only 15 pairs. The work of isolating paraphrases and source passages from disciplinary sources was undertaken by the first author only, which adds a level of uncertainty and imprecision to the already uncertain and imprecise process of identifying potential paraphrases. Moreover, though Keck’s (2006) taxonomy proved to be a useful tool for analyzing aspects of language borrowing and though Keck’s approach to excluding general links seems to have been born out in this data, larger scale analyses may benefit from a different, and perhaps simplified, way of conceptualizing these and other language borrowing practices in published paraphrases among expert writers.

Conclusion
The goal of the current project was not to generate generalizable or even precise propositions about paraphrase, disciplinary or otherwise. The project knowingly sacrificed precision to what Moffett (1983) would
call a “strategic gain in concept”. We have attempted to show that the goal of ensuring that paraphrase pedagogies are setting students in the right direction (i.e., toward expert practice rather than arbitrary proscription and untested rules of thumb) and ensuring that plagiarism policies (and the cultures that arise around them) are not standing in the way of key stages of paraphrase development is perhaps best achieved by starting with the experts and working our way back. This is likely the best way of creating inclusive developmental paraphrase pedagogies that permit space for failure, account for the high cognitive load associated with this essential source-based writing task, and orient practice opportunities toward the development of disciplinary expertise.

References


Institute of Electrical and Electronics Engineers (2019). IEEE Donald G. Fink Prize. IEEE. https://www.ieee.org/about/awards/bios/fink-recipient.html

Jackson, P. A. (2006). Plagiarism instruction online: Assessing undergraduate students’ ability to avoid plagiarism. College and


Teaching” (pp. 153-167). Royal Belgium Academy of Science and Arts. https://pdfs.semanticscholar.org/04b7/1f066a5c4bc9adad28361ebbba1bc7c91710.pdf


