A GENRE AND COLLOCATIONAL ANALYSIS OF THE NEAR-SYNONYMS *TEACH, EDUCATE AND INSTRUCT*: A CORPUS-BASED APPROACH

Thana Kruawong[a]; Supakorn Phoocharoensil[b]
([a]thana.kruawong@gmail.com; [b]supakorn.p@litu.tu.ac.th)

Language Institute, Thammasat University Thailand, Khlong Nueng, Khlong Luang District, Pathum Thani 12120, Thailand

Abstract: This study investigated three synonymous verbs, *teach, educate*, and *instruct*, in terms of their collocational patterns and distribution across genres. Data were drawn from the Longman Dictionary of Contemporary English (2021) and the Corpus of Contemporary American English (COCA). The findings from this investigation revealed that from distribution patterns among text types, *teach* is far more widely and commonly used than *educate* and *instruct*, with the highest frequency among eight genres. The frequency data also revealed that all three synonyms are preferred in formal genres than in spoken conversation, as represented by academic literature. The results also revealed that grouping of noun collocates into categories offers more insightful information about each target synonym's co-occurring authentic use. This can be more helpful to EFL students than simply looking up a term in a dictionary or relying on native speakers' intuition, which may not be dependable or trustworthy.

Keywords: COCA, collocations, corpus, noun collocates, synonyms

DOI: http://dx.doi.org/10.15639/teflinjournal.v33i1/75-97

Among several aspects of English as second or foreign language, synonyms can be a barrier to users wishing to become successful in a language (Yeh et al., 2007). Many languages, including English, have synonyms, e.g., *trouble* and *problem*, *achieve* and *accomplish*, *assess*, *measure* and *evaluate*, or *join*, *participate* and *attend*. Synonyms may be something that regularly presents difficulties for learners when they acquire L2 English vocabulary (Jackson & Amvela, 2007). Many L2 learners assume that they may swap a synonym for a word without affecting meaning or structure; however, there appear no words in English that can correctly occur in the same context of usage. Such near-synonym substitution might result in ungrammaticality or unnatural production in L2 English (Thornbury, 2002).

To widen ESL and EFL vocabulary size, words can be taught using rote learning (McKeown et al., 2012) and a Data-Driven Learning approach (Jafarpour et al., 2013). Notwithstanding, raising awareness of interchangeability among these must be encouraged as some words cannot always be substituted for others in all contexts of use. However, there are some factors and contexts that should be taken into consideration. If students are aware of synonyms, it will be beneficial for their learning progression and their ability towards authentic language use. In addition, knowing synonyms will help improve English skills, especially writing abilities (Yeh et al., 2007). Adopting a variety of vocabulary with the same core meaning when writing any kind of text can effectively captivate readers and sustain their attention.

Applying a corpus approach in English language teaching and learning has become more extensive and practical among ELT practitioners, since corpus data are able to provide authentic evidence for users to observe various distinctive patterns between synonymous words (Tsui, 2005). Near-synonyms are considered to be lexical pairs that have very similar literal and denotational meanings; however, some may differ in collocational behavior. Thus, they are not always interchangeable. Hu (2015) mentioned that even though two words may share similarities in terms of cognitive meaning, they may, in specific contexts, reveal different collocational patterns. As stated by Gu (2017), synonymous words should be used carefully as they usually differ in their semantic prosody and collocational behaviors.

There have been a number of studies investigating similarities and differences in usage among English synonyms; however, the synonyms *teach, educate*, and *instruct* are yet to be investigated. These words are categorized as being among the top 2,500 most common words in English (Macmillan Dictionary online, 2021). As can be seen from online forums on English learning websites such as https://wikidiff.com/, https://brainly.in/ and https://hinative.com/en-US, a number of questions have been posted concerning how to differentiate these three synonymous words. Many online forum users have offered answers to the questions, mostly based on their intuitions and with little or no concrete empirical evidence provided as support. Thus, it would be a favorable idea to investigate these three synonyms in a systematic way, using corpus data as the main evidence to support the results.

Near-synonyms

Near-synonyms refer to words in the same language which are close in meaning but cannot be substituted for each other in every context (Harley, 2006).

Despite having a similar basic meaning, these words do not always contain all of the same linguistic features, and thus can be distinguished by particular linguistic criteria (Jackson & Amvela, 2007). The significance of synonyms has attracted more attention in collocational-based research (Daskalovska, 2015; Hashemi et al., 2012). Furthermore, Leech (1981) also argued that to identify two words as synonyms, they must share the rule of implication or the rule that determines the possible substitution of each one.

Corpus

A corpus is as a collection of written and spoken texts that contains a variety of linguistic forms and information. Research benefiting from corpora has gained in popularity among linguistic analysts since corpus-based studies provide guidelines to their efficiency and convenience (Leel, 2011). A corpus, as a combination of naturally-occurring texts, can provide data for researchers to investigate the natural use of linguistic features and linguistic variations. Biber et al. (1998) argued that a salient characteristic of corpus linguistics is its combination of quantitative and qualitative perspectives to describe authentic language use. In terms of quantitative analysis, corpus linguistics allows an investigation of the frequency of distribution patterns of, for example, lexical items or phrases, in different genres of texts. The different co-occurrence of lexical items can identify word behavior and usage. These concepts can provide insights into the analysis of synonyms.

Collocation

The term collocation was first introduced in 1957 by J. R. Firth as a technical term used in lexical study. Collocations are words which co-occur naturally with each other and become commonly accepted considering repeated adoption. Timmis (2015) also emphasized that it is a combination of two lexical words that frequently occur together in near positions. This can influence the way people select words in English productive skills (Lindquist, 2009). For example, the synonymous adjectives *eager* and *enthusiastic* are considered near-synonyms, meaning 'having or showing intense enjoyment, interest, or approval'; however, these two synonyms have different collocations. To illustrate, *eager* is collocated with the preposition *for* while *enthusiastic* is collocated with the prepositions *in* and *about*. Liu and Jiang (2009) discovered that a corpus-based approach has a number of advantages in teaching vocabulary, including improvements in lexicogrammar command, awareness of grammatical structures, and discovery learning skills. Meunier and Reppen (2015) made

similar remarks, emphasizing how corpus-informed teaching may highlight the lexicogrammatical and contextual components of language usage rather than considering lexis and grammar as two distinct levels of language.

Degrees of Formality

Degrees of formality delineate the level of formality in which a set of synonymous words is likely to occur. Some vocabulary in a set of synonyms should be utilized in formal contexts, while others tend to be produced in informal conditions (Jackson & Amvela, 2007). For instance, Leech and Svartvik (2003) provide an example of synonyms that differ in degrees of formality: *children* and *kids.* To elaborate, *children* is commonly used in formal contexts while *kids* is commonly used in informal contexts.

Previous Related Studies

To date, there have been a number of studies investigating English synonyms. To begin with, Lapangdee and Phoocharoensil (2017) investigated the possible noun collocates of *student* and *learner,* and found differences between *student* and *learner* in terms of noun-noun collocations in COCA. The result showed that *student* and *learner* share a similar meaning, but differ in terms of collocations. There were less preferred collocations of *learner,* whereas *student* seemed to have more possible collocates. For this reason, the two synonyms cannot be substituted for one another in all contexts. In addition, Petcharat and Phoocharoensil (2017) analyzed three English synonyms, i.e., *appropriate*, *proper*, and *suitable*. In their study, meanings, degrees of formality, collocations, and grammatical patterns were the focus. Data used in this study were drawn from three dictionaries, as well as COCA. Their findings revealed that the three synonyms shared the same core meaning. However, they still differed in terms of precise meanings, degrees of formality, collocations, and grammatical patterns. It can thus be concluded that the three words cannot be used interchangeably in every context. Moreover, Jirananthiporn (2018) investigated two near-synonymous nouns, *problem* and *trouble*, focusing on the differences in their frequencies, distribution patterns across genres, and collocations with verbs and adjectives. The data of this study were drawn from COCA and then analyzed both quantitatively and qualitatively. Based on the overall frequency ranking, it was revealed that the noun *problem* occurred more often in all genres, with spoken and academic discourse ranking highly and fiction ranking the least. In contrast, the noun *trouble* commonly appeared the

most in fiction and the least in the academic genre. The discrepancy in their occurrences in different registers was somewhat in accordance with findings about the verb collocates of the synonym pair. Jarunwaraphan and Mallikamas (2020) investigated the differences and similarities of two synonymous nouns, *chance* and *opportunity*. In this study, data were drawn from COCA and online dictionaries. Throughout five genres, the noun *opportunity* was used most frequently in academic texts but least often in fiction genres. On the other hand, the noun *chance* seemed to occur least in academic genres, but most often in the spoken genres. In addition, a number of academic words from their list of collocates demonstrates that the noun *opportunity* is likely to be used more often in a formal style than is *chance*. The findings also suggest that near-synonyms, despite sharing similar meanings, may be used differently regarding collocation and semantic prosody.

Recently, Phoocharoensil (2021) examined genres and collocation patterns, in which three synonyms *consequence, result,* and *outcome* usually occur. The data were drawn from COCA. Of all the eight genres in COCA, the three synonyms appeared with the highest frequency in academic genres, and with the lowest frequency in informal genres, e.g., in TV and movie subtitles and in fiction. Concerning pedagogical implications, the fact that common verb and adjective collocates repeatedly co-occurred with synonymous nouns should be taken into consideration. Based on the frequency and the MI value (≥3), *consequence* is often used with verb and adjective collocates to convey a negative sense. In addition, the typical collocates of *result* are clearly associated with research-oriented contexts.

This study aims to distinguish between the synonyms *teach, educate,* and *instruct* by implementing a corpus free software to analyze the authentic use of the three words from different genres of texts. Phoocharoensil (2010) stated that there are several aspects of synonym analysis, such as, definitions, collocations, colligation, connotations, and styles. In this research, the two distinct features, namely distribution across genres and collocations, were selected and analyzed thoroughly. The three verbs appear to share some core meaning. According to the Longman Dictionary of Contemporary English Online (2021), the definitions of each verb are as follows:

Teach

1. to give lessons in a school, college or university or to help someone learn about something by giving them information
2. to show someone how to do something
3. to show or tell someone how they should behave or what they should think

Educate

1. to teach a child at a school, college or university
2. to give someone information about a particular subject or to show them a better way to do something

Instruct

1. to officially tell someone what to do
2. to teach someone something, or show them how to do something
3. to officially tell someone about something
4. to employ a lawyer to represent you in court

(Longman Dictionary of Contemporary English Online, 2021)

From the above definitions, *teach* seems to be more commonly known and used to provide the definition for the word *educate* and *instruct*. Based on the researchers' teaching experience, a number of questions about how to differentiate between *teach, educate* and *instruct* have been raised during class time. Only the researchers' intuition was relied upon for answers; no solid evidence was provided for the students to show how these words can be distinguished in real use. Several studies have confirmed the effectiveness of using corpus linguistics as an approach to distinguishing synonyms (e.g., Jarunwaraphan & Mallikamas; 2020; Jirananthiporn, 2018; Phoocharoensil 2021). It is expected that this study will provide a clear and reliable explanation as to how these synonyms can be distinguished systematically and authentically. Therefore, this study aims to distinguish these three words by implementing a corpus free program to scrutinize the authentic use of the three words from various genres of texts.

Research Questions

(1) What is the distinction between *teach, educate and instruct* in terms of frequencies and patterns of distribution across different text types?
(2) What is the distinction between *teach, educate and instruct* in terms of typical collocations with nouns?

METHOD

This study aimed to analyze three target synonymous words: *teach, educate* and *instruct*. The corpus used in this study was the Corpus of Contemporary American English or COCA. Having been created by Mark Davies, COCA is one of the largest and the most widely-known corpus of English that language

researchers can access without charge. Not only is it considered to be the largest corpus of American English, but it also collects a lot of authentic data representing real English language use. The latest version of COCA consists of more than 580 million words or texts, including eight genres: newspapers, magazines, fiction, and academic texts, spoken, TV/movies subtitles, blogs and other web pages. The corpus is also being regularly updated (Davies, 2020). Hence, this diverse, high quantity, and up-to-date information leads to reliability and validity attested by the number of international users (e.g., Liu, 2010; Yoo & Shin, 2020).

The researchers also considered the importance of the setting on the function tools in COCA to assure that the results obtained are consistent. In order to control the lemmatization of the three verbs, brackets were put before searching the words as well as "verb.ALL" in the search boxes (see Figure 1). The noun collocates set were then controlled in terms of the object noun, appearing four orders after verbs.

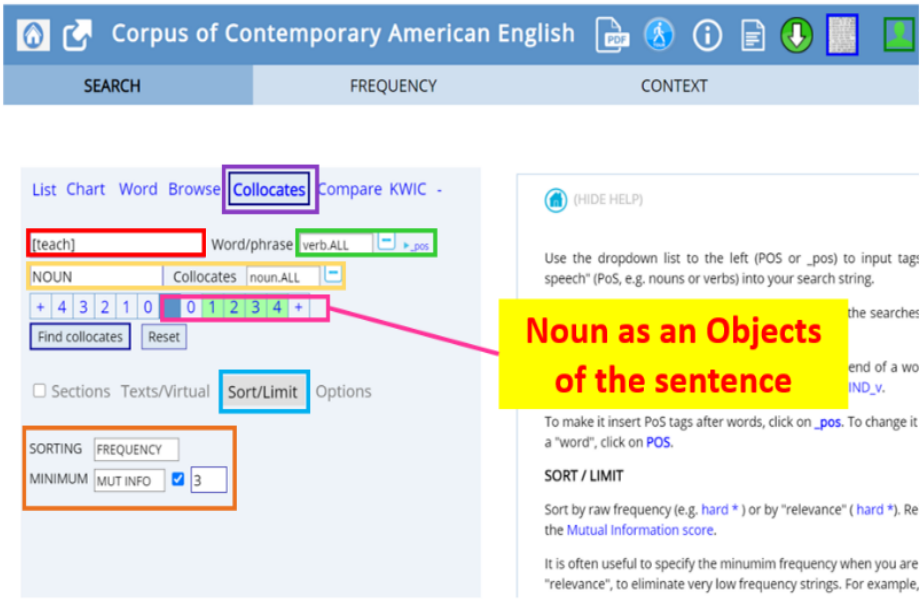

Figure 1. The setting before investigating the collocates in COCA

In order to answer the first research question, COCA was consulted for frequencies and distribution across eight genres of the three synonyms, *teach*,

educate, and *instruct*. To address the second research question, frequent collocations of the three target verbs were scrutinized. As the target synonyms are verbs, the focus was on object noun collocates that strongly co-occurred with them. Both frequency and the Mutual Information (MI) scores or values were considered when analyzing collocational strength. Gablasova et al. (2017) argued that the MI-score adopts a logarithmic scale to designate the ratio between the frequency of the collocation and the frequency of random co-occurrence of the two words in the blending. The MI score is often used to regulate a strong collocation between two component words. However, it should be noted that some collocations identified by very high MI scores could actually be low in frequency, and hence would not reasonably serve as representative examples. Therefore, Schmitt (2010) stated that the MI score should be used carefully in combination with a minimum frequency threshold. In this research, noun collocates in the top-forty of a frequency list were investigated with MI scores ≥ 3, which can be considered as the significance level for collocational association (Cheng, 2011).

After obtaining the data from COCA, an analysis of the three synonyms was conducted manually. There were four steps to the analysis of the collocation in order to answer the second research question. First, the top 50 noun collocates of three synonyms were scrutinized. Some repetitive words were eliminated in terms of lemmatization. In this step, it is important to carefully and manually recheck the information retrieved as merely relying on the automatic process can be misleading regarding the inaccuracy of the findings obtained. Next, all noun collocates were ranked by their frequency of occurrence. The final step was classifying the noun collocates into categories in order to illustrate the types of nouns co-occurring with the three synonyms more distinctly. Because all near-synonyms do not share the same collocational patterns, this process can help explain the lack of strict or absolute synonyms in English (Crawford & Csomay, 2016).

FINDINGS AND DISCUSSION

To answer the two research questions, the findings with regard to the overall frequency of *teach, educate* and *instruct* in eight different genres are presented and discussed followed by the collocation frequency of each of the synonyms.

Table 1. Distribution of the synonyms *teach, educate* and *instruct* across eight genres

	Teach			***Educate***			***Instruct***	
	Frequency	**Per million**		**Frequency**	**Per million**		**Frequency**	**Per million**
Academic	42,089	351.36	**Academic**	5,046	42.12	**Academic**	3,157	26.35
Web-GENL	22,529	181.31	**Blog**	5,098	39.64	**Web-GENL**	2,060	16.58
BLOG	22,092	171.77	**Web-GENL**	4,226	34.01	**Fiction**	2,024	17.11
Magazines	19,434	154.13	**Newspaper**	3,175	26.07	**Magazines**	1,805	14.32
TV/Movies	17,567	137.16	**Magazines**	3,040	24.11	**Newspaper**	1,434	11.78
Fiction	15,322	129.49	**Spoken**	2,538	20.12	**Blog**	1,263	9.82
Newspaper	17,717	145.53	**Fiction**	789	6.67	**TV/Movies**	760	5.93
Spoken	12,937	102.56	**TV/Movies**	616	4.81	**Spoken**	738	5.85
Total	169,687		**Total**	24,528		**Total**	13,241	

Table 1 illustrates that, in COCA, *teach* had the highest frequency among the three synonyms. The frequency of the word *teach* (169,687 tokens) is almost seven times higher than that of *educate* (24,528) and nearly 13 times higher than that of *instruct* (13,241). Table 1 also indicates that all three synonyms are found commonly in academic genre. This yields a very high degree of formality. This is further confirmed by the very low frequency in the spoken genre, which can be considered very informal. Noticeably, *educate* and *instruct* appear to be very low in TV/Movies, with 616 tokens and 760 tokens respectively. Interestingly, *educate* can be found frequently in the blog genre with 5,098 tokens, which highlights a difference when compared to *teach* and *instruct*. The findings are aligned with the previous studies which stated that some synonyms differ in terms of formality and styles of the context, e.g., *chance vs. opportunities* (Jarunwaraphan & Mallikamas, 2020), *error vs. fault/mistake* (Phoocharoensil, 2020) and *problem vs. trouble* (Jirananthiporn, 2018).

The next section addresses the analysis of the common noun and adverb collocates of the synonymous verbs *teach, educate* and *instruct*, responding to the second research questions. The finding that *teach, educate* and *instruct* are formally associated with academic language is consistent with Gardner and Davies' (2014) a New Academic Vocabulary List, in which all three are included.

Typical Collocations: Analysis of Noun Collocates

First, the top-forty verbs that frequently co-occur with the verb *teach*, *educate* and *instruct* are listed in Table 2. In order to approve their statistical significance, these nouns were chosen based on frequency in aggregation with a ≥ 3 MI score in COCA (Cheng, 2011).

Table 2. Noun collocates of synonyms *teach, educate* and *instruct*

Rank	*Teach*			*Educate*			*Instruct*		
	Noun Collocate	**Frequency**	**MI Value**	**Noun Collocate**	**Frequency**	**MI Value**	**Noun Collocate**	**Frequency**	**MI Value**
1	**student(s)**	4503	4.10	people	1472	3.06	**students**	463	3.50
2	children	4172	3.79	children	1295	4.89	teacher (s)	263	3.57
3	class (es)	2999	5.32	public	937	4.66	participant(s)	261	5.00
4	lesson(s)	2935	6.43	**students**	840	4.47	jury/jurors	202	5.18
5	kids	2547	3.93	kids	346	3.84	staff	106	3.18
6	skill(s)	2498	5.27	parents	230	3.44	subjects	98	4.51
7	English	1348	4.41	generation(s)	138	4.08	judge	84	3.01
8	courses	1118	5.78	youth(s)	137	4.64	employees	65	3.09
9	science	1023	3.31	teachers	137	3.44	officers	52	3.02
10	mathematics	926	4.93	consumers	134	5.12	agents	43	3.36
11	grade (s)	746	4.07	patients	127	3.60	respondents	27	3.54
12	strategies	686	4.66	citizens	125	4.29	disciples	22	5.26
13	methods	512	3.88	voters	107	3.94	clerk(s)	20	4.12
14	classroom(s)	492	4.32	girls	97	3.02	followers	19	3.63
15	literature	434	3.64	population	91	3.05	attorney	19	3.45
16	techniques	397	4.15	workers	86	3.03	experimenters	18	7.12
17	concepts	323	4.39	residents	70	3.72	pilots	17	3.85
18	subjects	317	3.51	customers	70	3.61	servant(s)	17	3.27
19	manners	301	4.36	employees	64	3.18	commanders	16	4.54
20	assistants	277	4.71	users	62	3.61	delegates	16	4.07
21	basics	271	3.60	readers	59	3.56	pupils	15	4.67
22	evolution	254	3.65	communities	59	3.39	superiors	14	5.55
23	undergraduate(s)	253	4.78	clients	58	3.81	listeners	14	4.02
24	philosophy	246	3.44	masses	56	5.55	admiral	13	3.89
25	curriculum	233	3.67	professionals	54	4.42	coders	12	7.24
26	principles	233	3.21	visitors	47	4.00	supervisor	12	3.49
27	biology	228	4.34	physicians	45	4.79	dispatcher	11	5.63
28	disabilities	225	3.88	disabilities	45	4.34	interviewers	10	6.19
29	bible	223	3.13	adults	44	3.13	tutor	10	4.89

Rank	*Teach*			*Educate*			*Instruct*		
	Noun Collocate	**Frequency**	**MI Value**	**Noun Collocate**	**Frequency**	**MI Value**	**Noun Collocate**	**Frequency**	**MI Value**
30	literacy	215	4.32	daughters	36	4.02	scriptures	10	4.53
31	economics	208	3.64	athletes	34	3.66	assistants	10	3.99
32	graduates	202	3.11	personnel	34	3.63	aides	10	3.45
33	physics	197	3.72	audiences	33	4.56	guides	10	3.26
34	faculty	190	3.00	peers	28	3.86	operator	10	3.00
35	ethics	180	3.63	nurses	28	3.47	apostle	9	4.42
36	tricks	178	4.57	sons	27	3.29	learners	9	3.38
37	journalism	175	3.97	immigrants	27	3.12	negotiators	8	4.88
38	discipline	175	3.54	administrators	26	4.04	intern	8	3.90
39	behaviors	163	3.16	coaches	26	3.39	diplomats	8	3.79
40	workshop	159	4.84	workforce	25	4.69	youngsters	8	3.76

Table 2 reveals that, according to the frequency and MI score (≥3), all synonymous words have a large number of nouns with which the target synonyms strongly collocate. The noun collocates *student(s)* and *teacher(s)* are shared by all of them. Some noun collocates e.g., *subjects* and *assistant* are found to be shared by *teach* and *instruct*, as opposed to *educate*. However, *teach* and *educate* share three noun collocates, namely, *children*, *kids* and *disabilities*. It cannot be firmly inferred which pairs are more closely related as synonyms of each other. Some noun collocates might be found more but they are not presented in Table 2 because of low MI value or low frequency. At this point, it can be seen that the number of noun collocates accompanying *teach*, *educate* and *instruct* helps to distinguish synonyms to a certain degree in terms of the variety of their collocations.

Table 2 also illustrates some practical information, such as that all three synonyms share the same noun collocate *student(s)* with a very high frequency of use. This indicates that all three words are synonymous when they share the same collocates (Petcharat & Phoocharoensil, 2017). Interestingly, *teach* is normally followed by object nouns that refer to different subject matter, e.g., *English, mathematics, literature, evolution, economics, biology, physics,* and *journalism*. It can also be inferred that *teach* is frequently used in educational contexts, like in school subjects or university subjects, while *educate* follows words that involve the interaction of society and social work, such as *people, youth(s), citizen, workers, communities, disabilities, immigrants* and *instruct*, and is found to be numerous in the field of law, e.g., *jury/jurors, judges, agents,*

and *attorneys*. The researchers decided to group all noun collocates into categories to find more patterns of authentic use.

The next step was to analyze the semantic preferences of the three target synonyms in order to categorize their noun collocates based on their similarities in meaning. Semantic preference refers to the limiting of lexical item co-occurrences to those that share a semantic characteristic (Sinclair, 2004). It is typical for words to be limited to certain semantic areas (Ang et al. 2017; Cheng, 2011). In other words, the semantic links that exist between words and their collocates might have an impact on semantic choice. A collocational information list assists in determining the search word's range of connections as well as the semantic linkages between its collocates (Ly & Jung, 2015). The noun collocates of the three synonyms can be divided into four pairs of main categories, which were: (1) Education and Training; (2) Social work and family; (3) Business and Office work; and (4) Laws and Politics. Those which did not belong to one of the mentioned groups were counted as (5) Others.

Table 3. Education and training

	Categorization: Education and Training		
	Teach	***Educate***	***Instruct***
Shared Collocates	student (s)		
	kids		-
	children		-
	-	teacher (s)	
	subjects	-	subjects
Different Collocates	class (es)	professionals	pupils
	lesson(s)	peers	supervisor
	skill (s)	administrators	tutor
	university(ies)	coaches	learners
	English		
	courses		
	science		
	mathematics		
	grade (s)		
	strategies		
	methods		
	classroom(s)		

	Categorization: Education and Training		
	Teach	***Educate***	***Instruct***
	literature		
	techniques		
	concepts		
	evolution		
	undergraduate(s)		
	philosophy		
	curriculum		
	principles		
	biology		
	literacy		
	economics		
	graduates		
	physics		
	faculty		
	ethics		
	journalism		
	discipline		

Table 3 shows the noun collocates that are shared among three synonyms in the education and training category. While the words *kids* and *children* are noun collocates shared by *teach* and *educate*, the word *teacher(s)* is shared only by *educate* and *instruct*. It is noticeable that even though the noun collocate *subject(s)* is found to be shared by *teach* and *instruct*, it has a different meaning when the concordance lines are scrutinized. The examples are as follows:

(1) … it is plausible that all elementary teachers who ***teach all subjects*** tended to consider science as the main content …

(2) …some trade school program ***teach subject*** like massage, baking and horseshoeing …

(3) … during each experimental visit, the ***subjects were instructed*** to lie down on a medical bed in the supine position …

(4) … In different experiments, ***subjects were instructed*** to try to maximize the number of positive images or…

It is noticeable from examples (1) to (4) that *subject* following *teach* normally refers to a thing or inanimate object like all subjects taught in school, but it refers to people (obviously in the research settings) when followed by *instruct*. Moreover, *instruct* is commonly used in the passive voice structure, while *teach* is used in both active and passive sentences.

Finally, when considering types of noun, *teach* always takes the object noun as both abstract nouns: e.g., *course, lesson, strategies, methods, technique, curriculum, concepts, curriculum, principle*, and concrete nouns: e.g., *university, classroom, undergraduates,* and *graduates*. Alternatively, *educate* and *instruct* mostly take only concrete nouns.

Table 4: Social work and family

	Categorization: Social Work and Family		
	Teach	***Educate***	***Instruct***
Shared Collocates	children		-
	kids		-
	disabilities		-
Different Collocates		people	youngsters
		daughters	servant(s)
		parents	superiors
		public	
		generations	
		youth(s)	
		citizens	
		girls	
		population	
		residents	
		communities	
		masses	
		adults	
		sons	
		immigrants	

Table 4 illustrates that in the social work and family theme, *children*, *kids*, and *disabilities* are shared by *teach* and *educate*, whereby *educate* co-occurs with

many nouns in this category: e.g., *people, public, generations, youth(s), citizen, immigrants,* etc. Interestingly, the noun collocate *girls* that follows *educate* can be interpreted regarding connotation when reading the concordance lines.

(5) … This money is going to ***educate women and girls*** in the developing world, in commemoration of…

(6) … help the child" and whose goal is to ***educate girls*** and prevent them from being sold into brothels…

(7) … her to project to ***educate lower-caste girls*** in India…

(8) … its mission is to inspire and ***educate girls*** while equipping them with the computing skills they need…

(9) … other religions, enshrine basic human rights, ***educate girls***, preserve peace treaties…

It can be inferred from examples (5) to (9) that *educate girls* suggests gender bias, and that female tends to be portrayed as an inferior position and undereducated.

Table 5: Business and office work

	Categorization: Business and Office Work		
	Teach	***Educate***	***Instruct***
Shared Collocates	-	Employees	
Different Collocates		consumers	staff
		customers	officers
		employees	agents
		clients	clerk(s)
		visitors	guides
		personnel	operator
		audiences	intern
		workers	negotiators
			dispatcher

In the category of business and office work, *employees* is a shared noun collocate of *educate* and *instruct.* However, the noun collocates of *educate* do not appear here since they mostly take place in the education and training theme.

Table 7. Laws and politics

	Categorization: Laws and Politics		
	Teach	***Educate***	***Instruct***
Different Collocates	-	voters	jury/jurors
			judge
			attorney
			delegates
			diplomats

Table 7 presents the laws and politics category which shows that *instruct* is found to have the most noun collocates. As mentioned earlier, *instruct* can have another meaning related to legal discipline; the term could be defined as "to employ a lawyer to deal with a legal case" (Longman Dictionary of Contemporary English Online, 2021). Examples (10) to (14) illustrate the use of the word in the data:

(10) … noting that the court below ***had instructed the jury*** that a fiduciary duty may arise informally from a relationship…

(11) … to commit a felony, but judge Morrison ***instructed the jury*** it could convict him of lesser…

(12) … he had repeatedly ***instructed the jury*** that they may only use statements Holmes makes in the video in…

(13) … for example, I am ***instructing my assistant district attorneys*** that we are no longer going to prosecute petty marijuana…

(14) …on Thursday voted ***to instruct attorneys*** representing the university of enter a voluntary meditation session…

Table 8: Others

	Categorization: Others		
	Teach	***Educate***	***Instruct***
Different Collocates	basics	users	participants
	tricks	readers	respondents
	bible	athletes	interviewers
		physicians	experimenters
		nurses	pilots

	Categorization: Others		
	Teach	***Educate***	***Instruct***
		patients	listeners
			coders
			followers
			scriptures
			commanders
			admiral
			aides
			disciples
			apostle

Our analysis of the noun collocates in the Others category found that there can be more sub-themes of specific terms in particular disciplines; for example, *educate* co-occurs with nouns in medical-related fields, namely *physicians, nurses* and *patients*. Some terms in research methodology, including *participants, respondents, interviewers,* and *experimenters*, can be found to co-occur with *instruct*. Military-related terms that collocate with *instruct* are *commanders, admiral*, and *aides*, and religion-related words like *disciples* and *apostle* can also be found collocated with such words.

The findings of this study show that the three synonymous verbs *teach, educate,* and *instruct* are regarded "near-synonyms", which is in line with Harley (2006) and Jackson and Amvela (2007), who found that near-synonyms may be employed differently in terms of collocation and semantic preferences while sharing similar meanings.

The study's findings also indicate that corpus-based synonym analysis may yield more startling and enlightening results than checking dictionaries or depending only on native speaker intuition (Aroonmanakun, 2015; Biber et al., 1998; Conrad, 2010). Furthermore, the current findings are broadly consistent with previous research in that the near-synonyms *teach, educate* and *instruct* cannot be used interchangeably in all contexts, and while they share some collocations, they have distinct semantic properties (Aroonmanakun, 2015; Gu, 2017; Jarunwaraphan & Mallikamas, 2020; Phoocharoensil, 2020; Selmistraitis, 2020; Sinclair, 2004; Thornbury, 2002).

Furthermore, this study implies that corpora might be a useful language teaching resource for language teachers, and it is vital that teachers teach their students about the complex usage of synonymous words and how to correctly use them. Because of the reliability and authenticity, linguists and ELT practitioners should promote corpus approaches in synonym analysis. Additionally, the use of corpus data in EFL or ESL classes, particularly in vocabulary teaching and learning, should be encouraged in order to extend students' viewpoints.

EFL and ESL instructors can apply these research findings to teaching their students regarding the three near-synonyms. They can also provide students with a clear explanation of the collocations frequently combined with the three verbs, with an emphasis on the fact that each verb has different contexts and collocational information that is not provided in dictionaries. A lack of awareness of the characteristics of each synonym (Szudarski, 2018), such as, formality or collocational patterns, often leads students to produce unnatural language. Teachers may point out that even though the three verbs have the same core of meaning, information from the dictionary only is not sufficient to select appropriate words. In fact, corpus data provides more information than dictionaries about how native speakers commonly communicate in both writing and speaking. There are, however, a huge number of English synonyms, and it would be beneficial for teachers to aim at effective synonym instruction in order to find a suitable way to present them to students. Teachers should search for alternative solutions to obtain additional information regarding synonyms from corpus databases. They should employ corpus databases as an innovative tool in addition to dictionaries to provide further information on vocabulary during English language teaching.

It is also highlighted that a corpus-based method has various favorable implications, including improved command of lexicogrammar and collocation (Liu & Jiang, 2009). This data-driven learning activity may be incorporated and utilized to increase learner awareness of semantic preferences, grammar, and collocation, particularly among intermediate and advanced students (Lee & Lee, 2010). It is possible that creating awareness is more essential than teaching particular pairings of synonyms.

We propose that, when designing in-house materials or textbooks, vocabulary selection is an essential process. Corpus data can provide guidelines for material developers in selecting words and their order of importance at various levels. Highly frequent words should be shown prior to those with lower

frequencies as well as other major differences in the usage of synonyms, especially with respect to formality, collocations, and grammatical patterns.

CONCLUSIONS

The present study aimed at investigating, by means of a corpus-based approach, the similarities and differences of three near-synonymous verbs, *teach*, *educate,* and *instruct,* in terms of their frequencies, distribution patterns across genres and noun collocation. The data were collected from COCA and analyzed both quantitatively and qualitatively. The analysis yielded a number of insightful advantages applicable to English language teaching. First of all, through distribution patterns across text types, *teach* is revealed to be a much more extensively and generally used than *educate* and *instruct*, with a frequency outnumbering the others across eight genres. The frequency information also indicated that all three synonyms happen to be favored in formal genres, represented by academic texts, rather than in spoken discourse. This observation is further supported by the low number of occurrences in informal genres, that is, spoken, fictions and TV/Movie.

The results also show that when categorizing the noun collocates, more information was yielded about the co-occurring authentic use of each target synonym. This can be more beneficial to EFL students than just obtaining the definition from a dictionary or merely relying on native speakers' intuition, which is not always reliable and trustworthy.

It should be noted that the study was limited to only one corpus-based data source, COCA. Therefore, the findings of this study may not be generalized to other varieties of English, e.g., British English or Australian English. A comparative study using another corpus, such as, the British National Corpus (BNC), may result in more valid generalizability on collocation and distribution across genres.

The following recommendations are proposed for future research. Further studies should employ other criteria to distinguish synonyms, which might include grammatical patterns, connotation, colligation (Jackson & Amvela, 2007) or adverbial and prepositional collocations and their semantic prosody (Phoochareonsil, 2020). Furthermore, the statistical use in collocation extraction used in this study may have been a limitation. To determine a common collocation, the combination of frequency and MI scores were applied. Nonetheless, Schmitt (2010) stated that the collocations extracted from a corpus differ based on the type of statistical measure applied. Future studies should employ other statistical tests, including t-score, z-score, or log-likelihood tests,

with a possibility of yielding different results (Gablasova et al., 2017). It is also highly recommended that other English synonyms be investigated using a corpus-based analysis to generate a clearer understanding for researchers in the fields of English language teaching and applied linguistics.

REFERENCES

Ang, L. H., Tan, K. H., & He, M. (2017). A corpus-based collocational analysis of noun premodification types in academic writing. *3L: Language, Linguistics, and Literature*, *23*(1), 115-131. http://doi.org/10.17576/3L-2017-2301-09

Aroonmanakun, V. (2015). Quick or fast: A corpus based study of English synonyms. *LEARN Journal: Language Education and Acquisition Research Network*, *8*(1), 53-62. https://so04.tci-thaijo.org/index.php/LEARN/article/view/102685

Biber, D., Douglas, B., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Cheng, W. (2011). *Exploring corpus linguistics: Language in action*. Routledge.

Conrad, S. (2010). What can a corpus tell us about grammar? In A. O'Keeffe & M. J. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (2nd ed) (pp. 227-240). Routledge.

Crawford, W., & Csomay, E. (2016). *Doing corpus linguistics*. Routledge.

Daskalovska, N. (2015). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, *28*(2), 130-144. https://doi.org/10.1080/09588221.2013.803982

Davies, M. (2020, June 25). *The new corpus of contemporary American English (COCA 2020)*. Language Institute Thammasat University (LITU) Webinar, Bangkok, Thailand.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning: A Journal of Research in Language Studies*, *67*(S1), 155-179. https://doi.org/10.1111/lang.12225

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, *35*(3), 305-327. https://doi.org/10.1093/applin/amt015

Gu, B. J. (2017). Corpus-based study of two synonyms: Obtain and gain. *Sino-US English Teaching*, *8*, 511-522. https://doi.org/10.17265/1539-8072/2017.08.006

Harley, H. (2017). *English words: A linguistic introduction*. John Wiley & Sons.

Hashemi, M., Azizinezhad, M., & Dravishi, S. (2012). Collocation a neglected aspect in teaching and learning EFL. *Procedia-Social and Behavioral Sciences*, *31*, 522-525. https://doi.org/10.1016/j.sbspro.2011.12.097

Hu, H. C. M. (2015). A semantic prosody analysis of three synonymous adjective pairs in COCA. *Journal of Language and Linguistic Studies*, *11*(2), 117-131. https://www.jlls.org/index.php/jlls/article/view/418

Jackson, H., & Amvela, E. Z. (2007). *Words, meaning and vocabulary: An introduction to modern English lexicology*. Bloomsbury Publishing.

Jafarpour, A. A., Hashemian, M., & Alipour, S. (2013). A corpus-based approach toward teaching collocation of synonyms. Theory & Practice in Language Studies, *3*(1), 51-60. https://doi.org/10.4304/tpls.3.1.51-60

Jarunwaraphan, B., & Mallikamas, P. (2020). A corpus-based study of English synonyms: Chance and opportunity. *rEFLections*, *27*(2), 218-245. https://files.eric.ed.gov/fulltext/EJ1283247.pdf

Jirananthiporn, S. (2018). Is this problem giving you trouble? A corpus-based examination of the differences between the nouns problem and trouble. *Thoughts*, *2018*(2), 1-25. https://so06.tci-thaijo.org/index.php/thoughts/article/view/163389

Lapangdee,M & Phoocharoensil, S. (2017). A corpus-based study of collocations of English synonyms: Student and learner [Master's independent study paper, Thammasat University, Bangkok, Thailand]. http://ethesisarchive.library.tu.ac.th/thesis/2017/TU_2017_5921042189_8556_6970.pdf

Lee, H. J., & Lee, E.-J. (2010). The effects of corpus-based vocabulary tasks on high school students' English vocabulary learning and attitude. *English Language & Literature Teaching*, *16*(4), 239-265. https://www.koreascience.or.kr/article/JAKO201003939212597.pdf

Leech, G. (1981). *Semantics: The study of meaning* (2nd ed). Penguin Books.

Leech, G., & Svartvik, J. (2003). *A communicative grammar of English* (3rd ed). Routledge.

Leel, H. C. (2011). In defense of concordancing: An application of data-driven learning in Taiwan. *Procedia-Social and Behavioral Sciences*, *12*, 399-408. https://doi.org/10.1016/j.sbspro.2011.02.049

Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh University Press.

Liu, D. (2010). Is it a chief, main, major, primary, or principal concern?: A corpus-based behavioral profile study of the near-synonyms. *International*

Journal of Corpus Linguistics, *15*(1), 56-87. https://doi.org/10.1075/ijcl.15.1.03liu

Liu, D., & Jiang, P. (2009). Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal*, *93*(1), 61-78. https://doi.org/10.1111/j.1540-4781.2009.00828.x

Longman Dictionary of Contemporary English. (n.d.). Citation. *In ldoceonline.com.* Retrieved February 13, 2021, from https://www.ldoceonline.com/

Ly, T. H., & Jung, C. K. (2015). A corpus investigation: The similarities and differences of cute, pretty and beautiful. *3L: Language, Linguistics, and Literature, 21*(3), 125-140. http://journalarticle.ukm.my/9073/

McKeown, M. G., Beck, I. L., & Sandora, C. (2012). Direct and rich vocabulary instruction needs to start early. *Vocabulary instruction: Research to practice*, *2*, 17-33.

Merriam-Webster. (n.d.). Citation. In *Merriam-Webster.com dictionary*. Retrieved February 13, 2021, from https://www.merriam-webster.com/dictionary/citation

Meunier, F., & Reppen, R. (2015). Corpus versus non-corpus-informed pedagogical materials: Grammar as the focus. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 498-514). Cambridge University Press. https://doi.org/10.1017/CBO9781139764377.028

Petcharat, N., & Phoocharoensil, S. (2017). A corpus-based study of English synonyms: Appropriate, proper, and suitable. *LEARN Journal: Language Education and Acquisition Research Network*, *10*(2), 10-24. https://so04.tci-thaijo.org/index.php/LEARN/article/view/111700

Phoocharoensil, S. (2010). A corpus-based study of English synonyms. *International Journal of Arts and Sciences*, *3*(10), 227-245.

Phoocharoensil, S. (2020). A genre and collocational analysis of consequence, result, and outcome. *3L: Language, Linguistics, Literature*, *26*(3). http://ejournals.ukm.my/3l/issue/view/1326

Phoocharoensil, S. (2021). Semantic prosody and collocation: A corpus study of the near-synonyms persist and persevere. *Eurasian Journal of Applied Linguistics*, *7*(1), 240-258.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Springer.

Selmistraitis, L. (2020). Semantic preference, prosody and distribution of synonymous adjectives in COCA. *GEMA Online Journal of Language Studies*, *20*(3), 1-18. http://doi.org/10.17576/gema-2020-2003-01

Sinclair, J., & Carter, R. (2004). *Trust the text: Language, corpus and discourse*. Routledge.

Szudarski, P. (2017). *Corpus linguistics for vocabulary: A guide for research*. Routledge.

Thornbury, S. (2002). *How to teach vocabulary*. Harlow.

Timmis, I. (2015). *Corpus linguistics for ELT: Research and practice*. Routledge.

Toye, M. (1971). To educate…. or to instruct? *Industrial and Commercial Training*, *3*(12), 588-589. https://doi.org/10.1108/eb003184

Tsui, A. B. (2005). ESL teachers' questions and corpus evidence. *International Journal of corpus linguistics*, *10*(3), 335-356. https://doi.org/10.1075/ijcl.10.3.03tsu

Yeh, Y., Liou, H. C., & Li, Y. H. (2007). Online synonym materials and concordancing for EFL college writing. *Computer Assisted Language Learning*, *20*(2), 131-152. https://doi.org/10.1080/09588220701331451

Yoo, I. W., & Shin, Y. K. (2020). Determiner use in English quantificational expressions: A corpus-based study. *TESOL Quarterly, 54*(1), 90-117. https://doi.org/10.1002/tesq.539