Stephan J. Franciosi
JAPAN

Teaching Students to Categorize TOEFL Essay Topics

With the growing popularity of standardized instruments for testing proficiency in academic English, such as the Internet-based Test of English as a Foreign Language (TOEFL iBT), teachers of English as a second or foreign language (ESL/EFL) are increasingly called upon to meet the preparation demands for prospective test takers. For many of these learners, a particularly troublesome component of the TOEFL iBT is the 30-minute timed essay. To prepare students for the essay, teachers must provide guidance in lower-order writing skills such as grammar, vocabulary, and paragraph structure, and they must also decide on the utility of imparting an awareness of possible topics for the essay. These topics may include the qualities of a good neighbor, local public spending priorities, or local educational opportunities. Since there are many potential topics, instructors may wish to reduce them to a more manageable number by grouping individual writing prompts into larger categories of topics, for example by including the aforementioned three topics into a group labeled "community issues." This article presents a method for guiding students to generate topic categories and discusses the benefits of this approach.

It is worth asking whether prior knowledge of possible topics or categories of writing topics helps prospective test takers to write better essays. After all, the test developer, Educational Testing Service (ETS), claims that the subject matter is designed to be "familiar" to examinees in accordance with the overall purpose of the TOEFL, which is a test of language skill, not content knowledge. This claim implies that students are already sufficiently knowledgeable about topics that may appear and should therefore concentrate preparation effort on lower-order writing skills such as grammar and vocabulary.

It turns out, however, that there is a very good reason to include categories of essay topics in TOEFL writing preparation. Categorization is an inherent and omnipresent function of human cognition that is vital for organizing

and understanding the perceived world (Medin and Aguilar 1999). Knowing categories of writing prompts greatly assists test takers because higher orders of semantic groupings are more readily remembered and retrieved than lower-order labels; these groupings, once retrieved, can aid in the recovery of more-specific vocabulary (Federmeier and Kutas 1999; Kuipers, La Heij, and Costa 2006). This benefit in turn facilitates both top-down and bottom-up writing processes (van Gelderen, Oostdam, and van Schooten 2011). Accordingly, empirical studies on ESL/EFL writing outcomes have shown that TOEFL writing prompt topics do influence scores (Breland et al. 2004; Lee, Breland, and Muraki 2005), that topic familiarity is a factor (He and Shi 2012), and that attempts at achieving greater fairness by devising more general or familiar topic choices are not entirely successful (Lee and Anderson 2007).

The literature has two important takeaways. First, the more knowledgeable writers are about a particular topic, the better they can write about it—in terms of both content and structure. Second, it is unreasonable to assume, from the viewpoint of fairness, that every test taker is equally familiar with every topic. ETS appears to recognize as much, as it provides on its TOEFL website a list of 185 writing prompts and advises prospective test takers to familiarize themselves with the list from which a topic may be selected for the actual test. In addition, simple familiarity may be insufficient for addressing the task. The pitfall of "familiar" topics is the assumption that extensive experience with something equates to critical thinking about it, when more often than not we tend to take the familiar for granted. The TOEFL essay requires test takers to think critically (in many cases for the first time) about institutions, customs, and artifacts that they have grown up with, increasing the cognitive load of the task at the same time they are struggling with foreign language mechanics.

The benefits of student-generated topic categorization

A prior knowledge of the higher-order semantic groupings of prompts allows test takers to prepare more efficiently by targeting their preparation efforts at learning about unfamiliar topics and thinking critically about aspects of their lived experiences that they had heretofore taken for granted. A categorization of topics would ideally be both rigorous and meaningful from the standpoint of learners. For these two reasons, I suggest involving the learners in the categorization process as opposed to providing a teacher- or textbook-generated grouping system. Specifically, I suggest involving the students in an activity known as a *content analysis*, which is a formal method of producing a hierarchy of categories. Bryman (2008) describes this as a type of qualitative research that involves coding a data set consisting of an unstructured text in order to seek general trends. As a formal method of research, content analysis has the advantage of flexibility in that it can apply to a wide variety of textual data. This analysis generally involves the detailed scrutiny of a given text with the aim of uncovering facts about the culture that produced it (Bryman 2008); however, for instructional purposes, it is sufficient to stop short of learning about the culture of TOEFL iBT writers and instead to use the process to generate a categorization of writing topics.

A student-generated categorization results in more understandable and robust groupings, and it fits well into a learner-centered, project-oriented instructional approach. Even a formalized system of categorization is fundamentally subjective because coding always involves a degree of construal by the coder (Bryman 2008). This means that categories chosen by students themselves would be more meaningful to them than those made by an instructor, materials writer, or other authority. At the same time, it also means that collaboration among students adds rigor to the categorization, making the final result more robust. The element of collaboration further makes this application ideal for educators who subscribe to constructivist approaches and who wish to engender discussion and knowledge sharing among their students. In the next sections, I illustrate a collaborative content analysis of the TOEFL writing prompts by describing my own classroom methods.

Classroom methods

An analysis requires a data source. The text source I use is the list of 185 TOEFL prompt questions available as a PDF download from

the ETS website at www.ets.org/Media/Tests/TOEFL/pdf/989563wt.pdf. An identical list is available in *The Official Guide to the TOEFL Test* (Educational Testing Service 2009), which encompasses the TOEFL iBT. An additional 35 prompts for the same task are listed on ETS's Criterion Topics website at www.ets.org/Media/Products/Criterion/topics/topics.htm; however, 25 of these are identical to prompts already included in the list of 185. Adding the 10 unique prompts to the list yields a combined corpus of 195 distinct prompts.

It should be noted that, although the primary list of 185 prompts is associated with an older Computer-based TOEFL (CBT) version of the test, it should not necessarily be considered outdated. Prompts are carefully vetted for reliability and an acceptable level of fairness, so they are not readily generated or replaced. In fact, my students have encountered many of the same prompts in actual examinations.

Depending on the facilities available, I may use either an Excel file or paper cards to support the content analysis process. Obviously the latter is much more labor-intensive in terms of preparation, but it may also be preferred by students. It is important to remember that the main purpose of the medium is to allow for the labeling and sorting of individual prompts. While this is readily accomplished in Excel, I have encountered many students who actually find writing each prompt on a separate card an easier way to sort, re-sort, and keep track of the data.

Coding and analysis

The coding and analysis of the corpus follow an abbreviated version of the procedure described by Saldaña (2009). The process is divided into two cycles.

Cycle 1: Assigning keywords

The initial cycle begins by "tagging" the prompts with keywords. I divide the class into groups of five or six students and provide each member with a list of 32–39 random prompts from the original corpus of 195. I then ask the students to tag the prompts with two or three words that are central to or representative of the prompts. Since it often takes time for each individual to read through and understand each of the prompts, this activity is best suited as homework. The result is a group of keywords taken by each student directly from the prompts themselves. Table 1 shows how three prompts might be tagged.

As an inherent semantic function of natural languages, keywords may have synonyms or nuances that are essential to understanding the topic represented by a prompt and that can be determined only by observing the words or phrases individually in context. So the next step is to compare and contrast the data by bringing together prompts that are similarly tagged in order to determine whether relationships exist. In class, I ask students to work in their groups to create groupings of prompts based on keywords. Since there should be several keywords, students are able to perceive most given prompts in various distinct groups, as illustrated in Table 2. I encourage them to look for similarities and differences among the

Prompt	Keywords
1. Do you agree or disagree with the following statement? *Television has destroyed communication among friends and family.*	television; communication; family; friends
2. If you could make one important change in a school that you attended, what change would you make?	school; change
3. Which would you choose: a high-paying job with long hours that would give you little time with family and friends or a lower-paying job with shorter hours that would give you more time with family and friends?	high-paying; low-paying; job; family; friends; time

Table 1. Tagging prompts with keywords

prompts in a grouping and to examine how the groupings are represented by the keywords.

At this point, I ask students to look for similarities among the keywords and, when appropriate, to replace them with *meta-labels*, which are words or phrases that add additional or emphatic topic-related information to a prompt. Because the ultimate goal is to identify higher-order categories, I ask students to seek more general terms for describing the topics represented in the prompts. For example, of the tags in Table 1, "friends" and "family" may be replaced by "relationships," while "television" and "communication" may be replaced by "media" (see Table 3). These meta-labels are the initial categorization of the prompts.

Cycle 2: Comparing and contrasting categories

The second cycle of coding involves comparing and contrasting the provisional categories both internally and externally. That is, students examine prompts sharing the same labels for additional similarities and differences in their topics. At the initial stages of this cycle, comparison of categories usually results in subsequent reshuffling and re-categorization, which may or may not entail the creation of subcategories several levels deep. The analysis begins with an "exhaustive" set of categories that explores as many options as possible; in a qualitative investigation this process would continue until the coder is con-

Grouping 1. Keyword = *school*	Grouping 2. Keyword = *change*
If you could make one important change in a **school** that you attended, what change would you make?	If you could make one important **change** in a school that you attended, what **change** would you make?
Do you agree or disagree with the following statement? *Classmates are a more important influence than parents on a child's success in **school**.*	Nowadays, food has become easier to prepare. Has this **change** improved the way people live?
Do you agree or disagree with the following statement? *Boys and girls should attend separate **schools**.*	If you could **change** one important thing about your hometown, what would you **change**?

Table 2. Grouping of prompts by keyword

Prompt	Meta-labels
Do you agree or disagree with the following statement? *Television has destroyed communication among friends and family.*	media; relationships
If you could make one important change in a school that you attended, what change would you make?	education; change
Which would you choose: a high-paying job with long hours that would give you little time with family and friends or a lower-paying job with shorter hours that would give you more time with family and friends?	income; employment; relationships; lifestyle

Table 3. Grouping of prompts by meta-labels

fident that the categories are "mutually exclusive" (Bryman 2008, 288). However, in an educational setting the validity requirements should not be as stringent, and the analysis may conclude when students feel comfortable that they have achieved a useful result.

It is also worth noting that in practice many prompts may straddle categories. For example, a prompt regarding "the qualities of a good parent" may arguably fit into either a category entitled Ideal Personality Traits or one called Parenting (see Table 4). However, I do not believe this detracts from the utility of the exercise. Students may initially be stymied by overlapping categories but will tend to settle on a categorization through discussion as long as they are aware that ambiguity can be resolved by determining degrees of acceptability. I tell students that they may categorize such a prompt however they wish, but that they must also clearly articulate the reasons for the choice. A discussion triggered by the recognition that a prompt could be included in two separate categories serves only to deepen understanding of the prompt, the categories, and the other prompts within the relevant categories. So, in this sense, ambiguity actually has a beneficial impact on learning.

Finally, it is important to point out to students that there will inevitably be *outliers*—that is, individual prompts that do not seem to fit into any established categories or prompts whose inclusion somewhere would make it necessary to reshuffle existing categories and create

Prompt: What are some of the qualities of a good parent?	
Category: Ideal Personality Traits	**Category: Parenting**
What are some important qualities of a good supervisor (boss)?	Some people think that parents should plan their children's leisure time carefully. Other people believe children should decide for themselves how to spend their free time. Which idea do you agree with?
Neighbors are the people who live near us. In your opinion, what are the qualities of a good neighbor?	Do you agree or disagree with the following statement? *Parents or other adult relatives should make important decisions for their older (15- to 18-year-old) teenage children.*
Many students have to live with roommates while going to school or university. What are some of the important qualities of a good roommate?	Do you agree or disagree with the following statement? *Parents are the best teachers.*

Table 4. Possible ambiguity in categorization

Definition: A narrative of important events connecting past, present, and future
Learning about the past has no value for those of us living in the present. Do you agree or disagree?
What discovery in the last 100 years has been most beneficial for people in your country?
The twentieth century saw great change. In your opinion, what is one change that should be remembered about the twentieth century?
The 21st century has begun. What changes do you think this new century will bring?

Table 5. Definition and prompts for History category

new outliers. I usually handle this by having students include a Miscellaneous category and asking them to keep it as small as possible.

Focusing strategies

To assist the analysis, Saldaña (2009, 186) recommends two "focusing strategies," one of which is to write out definitions of the emerging categories. The students formulate these with the intent to encompass all the prompts represented in a provisional category. For example, the category of History can be defined as shown in Table 5, effectively including all listed prompts.

Category	Definition	Subcategories	Examples of Prompts
Education	The social and cultural institution of formal learning and teaching	School Policy	Schools should ask students to evaluate their teachers. Do you agree or disagree? Do you agree or disagree with the following statement? *Boys and girls should attend separate schools.*
		Prioritizing Subjects	Do you agree or disagree with the following statement? *All students should be required to study art and music in secondary school.* Do you agree or disagree with the following statement? *It is more important for students to study history and literature than it is for them to study science and mathematics.*
People and Nature	The interaction between human civilization and the natural environment	Animals	Many people have a close relationship with their pets. These people treat their birds, cats, or other animals as members of their family. In your opinion, are such relationships good? Why or why not? Do you agree or disagree with the following statement? *A zoo has no useful purpose.*
		Using Natural Resources	Some people think that human needs for farmland, housing, and industry are more important than saving land for endangered animals. Do you agree or disagree with this point of view? Many parts of the world are losing important natural resources, such as forests, animals, or clean water. Choose one resource that is disappearing and explain why it needs to be saved.

Table 6. Sample of categorization by TOEFL writing students

The act of defining categories often leads to the dissolution and reconstitution of provisional categories, and to subsequent relabeling and redefining. This is an iterative process that may or may not be carried out over several classes. I believe that it is up to the instructor to "read" the students and decide when they have reached an acceptable level of satisfaction with their categorizations. That said, some groups will reach a satisfactory conclusion relatively quickly, even within one class session. In this case, I would suggest dedicating at least some time in a subsequent meeting to re-examining the results for the same reason that revisiting any project after an interval of thought and reflection enhances knowledge construction (Kolb 1984).

Another focusing strategy suggested by Saldaña (2009) is to seek peer input, which is inherent in the project-based approach I am advocating, but the rationale still deserves mention. A sole coder is unavoidably informed as well as limited by individual experience and existing knowledge, and reducing the possible negative influences of partiality requires communication with other observers of the data. This communication may also entail groups comparing final categorizations and the creation of one hierarchical categorization for an entire class.

Results

My students typically produce between 20 and 25 major categories with subcategories up to three levels deep. The largest categories may include up to 20 prompts, and the smallest as few as three. There are also usually up to 10 outliers. Table 6 is a sample of one categorization effort and contains two major categories with two subcategories each.

Using the results

After the categorization is completed, the students have a robust and self-organized knowledge of what topics may appear on the test. Since this knowledge mostly involves higher-order semantic groupings, it can be characterized as "wide and shallow," but it provides a definite guide for students to direct their own preparation efforts.

I advise students to learn more about areas where they feel under-informed; the rationale is that they cannot provide the reasons and examples that are required by the test prompts if they have no declarative knowledge of the topic. To illustrate, I have seen several groups produce the prompt categorization Preferred Learning Styles, shown in Table 7.

The topic of learning is of course familiar to my Japanese college students, and they do have preferences, but they often find it difficult to express the reasons why. For instance, when considering the second prompt listed in Table 7, not many of my students have experience with, let alone understand the rationale for, a class where "the students do some of the talking." Yet in order for test takers to think critically about the different classroom modalities and

Category	Prompts
Preferred Learning Styles	1. Some students prefer to study alone. Others prefer to study with a group of students. Which do you prefer? 2. Some students like classes where teachers lecture (do all of the talking) in class. Other students prefer classes where the students do some of the talking. Which type of class do you prefer? 3. Some people think that they can learn better by themselves than with a teacher. Others think that it is always better to have a teacher. Which do you prefer? 4. People learn in different ways. Some people learn by doing things; other people learn by reading about things; others learn by listening to people talk about things. Which of these methods of learning is best for you?

Table 7. Preferred Learning Styles prompts

provide a reasoned response to the prompt, knowledge of the pros and cons of various learning approaches is essential. To contextualize the entire category, I introduce the concept of learning theories, provide resources for students to gain declarative knowledge of the concepts, and follow up with group discussions.

Of course, the goal is not to convert students into educational psychologists, so I concentrate on finding short, introductory articles and videos. Happily, my students have good Internet connectivity and are net savvy, so I can easily direct them to a multitude of free resources. For example, Wikipedia has a page introducing the four major theories of learning—behaviorism, cognitivism, humanism, and constructivism—that provides a brief description of the main principles and some rationale; the page also includes further information on topics such as learning styles and instructional theory. Other free resources for this particular topic include Learning-Theories.com, or for media other than text, simply performing a keyword search in YouTube for "learning theories" will yield several dozen multi-media overviews at various levels of detail. I follow up the exploration of these resources with discussions about my students' own learning experiences, asking them to identify theories that would underpin or explain those experiences, and to articulate the aspects of knowledge acquisition that the various theories were devised to address.

Finally, it is worth emphasizing that prospective test takers do not typically have the time to delve into every one of the 20 or so categories during a preparation course. Yet this fact highlights the advantage of having a categorization in that it allows learners to target their areas of weakness and plan their preparation efforts accordingly. For instance, education majors who are already familiar with learning theories would probably not need to spend time reading about them, whereas engineering majors would more likely benefit from the knowledge. In this way, the categorization affords the ability to tailor preparation to meet the needs of individual learners.

Conclusion

In this article I have described a method of creating a student-generated categorization of TOEFL iBT writing topics using materials freely or readily available online. Although my particular course focuses on the TOEFL iBT, the flexibility of this type of qualitative analysis allows it to be applied to various other sources of unstructured data, including essay topics for the International English Testing Service, which are also freely available online. Involving students in the categorization process deepens their understanding of the categories and provides them with a tool for directing their own learning efforts.

References

Breland, H., Y. Lee, M. Najarian, and E. Muraki. 2004. *An analysis of TOEFL CBT writing prompt difficulty and comparability for different gender groups*. Princeton, NJ: Educational Testing Service.

Bryman, A. 2008. *Social research methods*. 3rd ed. Oxford: Oxford University Press.

Educational Testing Service. 2009. *The official guide to the TOEFL test*. 3rd ed. New York: McGraw-Hill.

Federmeier, K. D., and M. Kutas. 1999. A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language* 41 (4): 469–495.

He, L., and L. Shi. 2012. Topical knowledge and ESL writing. *Language Testing* 29 (3): 443–464.

Kolb, D. A. 1984. *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.

Kuipers, J., W. La Heij, and A. Costa. 2006. A further look at semantic context effects in language production: The role of response congruency. *Language and Cognitive Processes* 21 (7–8): 892–919.

Lee, H., and C. Anderson. 2007. Validity and topic generality of a writing performance test. *Language Testing* 24 (3): 307–330.

Lee, Y., H. Breland, and E. Muraki. 2005. Comparability of TOEFL CBT writing prompts for different native language groups. *International Journal of Testing* 5 (2): 131–158.

Medin, D. L., and C. Aguilar. 1999. Categorization. In *The MIT encyclopedia of the cognitive sciences*, ed. R. A. Wilson and F. C. Keil, 104–106. Cambridge, MA: MIT Press.

Saldaña, J. 2009. *The coding manual for qualitative researchers*. London: Sage.

van Gelderen, A., R. Oostdam, and E. van Schooten. 2011. Does foreign language writing benefit from increased lexical fluency? Evidence from a classroom experiment. *Language Learning* 61 (1): 281–321.

Stephan J. Franciosi is a Doctor of Education in Learning Technology, also possessing a Master of Arts in Linguistics. He has been instructing TOEFL preparation and Academic Writing for 12 years.