

NUMB3RS Activity: Message of the Bottle

Episode: "Toxin"

Topic: Code Breaking

Grade Level: 7 - 9

Objective: Breaking single letter substitution codes using knowledge of the English language.

Time: 20 minutes

Introduction

In the episode "Toxin," the FBI tracks a serial prisoner who uses contaminated drugs. The FBI intercepts an encrypted message without the encryption key and Charlie explains to Don that Information Theory and knowledge of the English language makes deciphering easier.

For example, some letters in the English language, like a, s, t, r are used more than others, like x, q, or z. Some words like "who" or "what" may be used more in questions than others. Some letters appear together more frequently than others, for example "q" almost always is followed by a "u," and there is a vowel in almost all words. If you can decode some of the letters in the word, you can often tell what the word is.

For example if you know that "_ _ e r e" is the first word in a question, you can probably guess that the missing letters are "w" and "h."

The messages contained in this activity use **single-letter substitution**. Each letter is only substituted for one other letter within a message.

Discuss with Students

You probably know that most codes can be broken with the use of a computer. Many simple codes can also be broken by hand if you apply your knowledge of the English language. As a very simple example, if you decoded the first two letters of a three-letter word as 'TH-', you'd have a very good idea of what the last letter is.

You know other patterns in the English language that can help to break codes, though you rarely think about them. For example:

1. Give several words you might expect to find in any question. [*who, what, why, where, how*]
2. What letters often appear as double letters in common English words? [*b, c, d, e, f, g, l, m, n, o, p, r, s, t, z*]
3. What are some 3- or 4-letter words that start and end with the same letter? [*sample responses: mom, dad, did, tot, eve, that, sees, ease, seas, rear, prop*]
4. What letters do you think are the most used in the English language? [*The ten most commonly used letters, in order, are: e, t, a, o, i, n, s, r, h, l*]
5. What are some common 1- and 2- letter words that appear often in the English language? [*l, a, am, is, it, to, of, do, be, we, us, by, my, he*]

Student page answers: 1. a. *what, when* b. Using *WHEN* as the first word makes the fourth word *NHEN*, which is not a word. Using *WHAT* as the first word, the message would appear as **WHAT AMF THFKF UQGGK THAT ZQGG?** c. The verb likely will be "are." d. **WHAT ARE THESE PILLS THAT KILL?** e.

X	J	E	W	E	M	F	W	J	F	K	F	U	Q	G	G	K	W	J	E	W	Z	Q	G	G
W	H	A	T	A	R	E	T	H	E	S	E	P	I	L	L	S	T	H	A	T	K	I	L	L

2. a. *WHO IS THE INTENDED VICTIM?* b. *Strategies will vary.* 3. a. *You can only use letters that have been contained in previous messages.* b. *messages will vary* A sample message: *When and where to meet?*

Name: _____ Date: _____

NUMB3RS Activity: Message of the Bottle

The FBI finds a writing pad used by the suspect in a series of poisonings. On it, they find the imprint of an encoded message, but no key to decipher it. Don thinks the coded message is from a killer who is communicating with an insider in the pharmacy. Charlie explains that deciphering the message requires knowledge of the structure of English words and sentences. The message they find is:

XJEW EMF WJFKF UQGGK WJEW ZQGG?

1. **a.** The message is a question. Identify common four-letter words that might start a question: _____
- b.** Use one of the words from part **a** and replace all the other identical letters in the message used in the word you chose. _____
- c.** The second word in a question is often a verb. Based on your answers in parts **a** and **b**, what is the most likely verb in this question? _____
- d.** If you know that "G" is really an "L," complete the rest of the message using the letters you've found and by making an educated guess:

- e.** Make a key for the code based on the letters you've decoded from the message.

Agents intercept a reply to the message above. Charlie suggests using the key from the first message to decode this one. The intercepted message is below:

XJC QK WJF QAWFASFS HQIWQP?

2. **a.** Decode this new message. _____
- b.** Describe the strategy you used to decode this message.

Agent Epps suggests trying to catch the accomplice by sending a message in the same code. In this fake message, they would answer the question and ask for a meeting with the "insider."

3. **a.** What limit is there on the letters you can use in the message you write?

- b.** Write a possible return message using the key you have developed.

The goal of this activity is to give your students a short and simple snapshot into a very extensive math topic. TI and NCTM encourage you and your students to learn more about this topic using the extensions provided below and through your own independent research.

Extensions

For the student:

- Below is a chart showing the proportional use of each letter in the English language. Look at the most and least used letters. Do these make sense to you?

Letter	Probability	Letter	Probability	Letter	Probability
a	0.0642	j	0.0008	s	0.0514
b	0.0127	k	0.0049	t	0.0796
c	0.0218	l	0.0321	u	0.0228
d	0.0317	m	0.0198	v	0.0083
e	0.1031	n	0.0574	w	0.0175
f	0.0208	o	0.0632	x	0.0013
g	0.0152	p	0.0152	y	0.0164
h	0.0467	q	0.0008	z	0.0005
i	0.0575	r	0.0484	space	0.1859

- The probabilities above add up to 1. This means that punctuation is not being considered in the analysis. How do you think different punctuation, like parentheses, commas, semi-colons, dashes, and exclamation marks might be distributed?
- Consider that a lot of code-breaking is not conducted in English. If you're familiar with a foreign language, think about how the chart above would change for other languages. What other parts of speech or letter-usage will be more common in that language than in English?
- The board game Scrabble has a number of tiles with letters on them used to form words. Consider the distribution of the letters on the tiles. Based on the table above, are the letter tiles in Scrabble distributed according to their common use in English?
- Cryptoquotes are word puzzles that resemble the encoded messages found in this activity. Many newspapers offer them to their readers to solve on a daily basis. Below are two web pages that give you an opportunity to try your code-breaking skills. One of them uses the digits 0-9 as well as the letters of the alphabet to encode the quote.

<http://www.cryptoquotes.com/>

http://www.eastoftheweb.com/cgi-bin/go_daily_game.pl?game_id=Cryptoquote&id=0