SELF-ASSESSMENTS OF VOCABULARY KNOWLEDGE : PARTIAL REPLICATION

Mayumi Tsubaki

Introduction

Vocabulary is an important component in language learning, as it pervades all four skill areas of listening, speaking, reading, and writing. Vocabulary knowledge considerably helps language learners in both comprehension and production. In fact, some students of elementary proficiency, who know little grammar, can not only guess the meanings of even cognitively demanding reading materials but also produce understandable sentences as long as they have vocabulary knowledge from a dictionary or other sources. Indeed, vocabulary knowledge plays a significant role in English proficiency. Thus, research on the subject should not be neglected.

One means of measuring language proficiency that has been gaining attention recently among Teaching English to Speakers of Other Languages (TESOL) professionals has been self-assessment, also referred to as self-evaluation. The following study, which partially replicates an investigation done by Laufer and Yano (2001), examines the accuracy of self-assessment of vocabulary knowledge in context.

Literature Review

Many teachers and testing services have long considered language assessment to be synonymous with testing (McNamara and Deane, 1995). However, some of the problems inherent to testing have led a growing number of language teaching professionals to recognize the value of self-assessment as an alternative to standardized and teacher-made tests. One criticism of traditional testing has been that it does not necessarily indicate students' true abilities.

Researchers and TESOL professionals cite several reasons for their advocacy of self-evaluation. The first relates to psychological factors. Students can become independent, responsible and autonomous learners through awareness of their own learning. Learners who are given self-assessment opportunities become less dependent on their teacher's

judgment about their progress than those who are not given such opportunities. The result is positive psychological effects. Smolen, Newman, Wathen, and Lee (1995) mention that self-assessment leads students to feel ownership in and responsibility for their learning which in turn enables them to make decisions about it. For this reason, Gardner (1996) even suggests that self-assessment procedures be explicitly included as learning materials. Furthermore, Butler (1997) asserts that because of the independence, responsibility and autonomy they engender self-assessment opportunities and foster a learner-centered environment.

A second reason for using self-assessment is that it lets language learners know not only how well they are doing and but also how and what to learn. Thus, self-assessment promotes the development of meta-cognitive skills and, as a result, even students' thinking skills improve (El-Koumy, 2001). When students can assess their progress, they naturally know more precisely their linguistic strengths and weaknesses as well as the kind of learning strategies they tend to use. It is true that traditional testing somehow makes students reflect on their own learning, yet self-assessment makes learners examine their learning far more actively since they have to make their own judgments about it.

A third advantage of self-assessment is that it allows teachers to lighten their workload (Lynn, 1995). Although, traditionally, students rely on their teachers' evaluations to know if and how much they have learned, teachers do not necessarily have adequate knowledge on which to base their assessments (Schraeder, 1996). Brutten (1981) examined the differences between teachers and students in perceived vocabulary knowledge. In the experiment, both the students and teachers chose words that they thought the former did not know. The results indicated that the agreement between the two groups was only 53%, a little over the half. Since this study did not test whether these perceptions of unknown words were correct, it is not clear which group's intuitions were more accurate. In any case, teachers do not seem to necessarily show any greater awareness of which words their students do and do not know.

This problem of excessive teacher responsibility in evaluating student abilities in English as a Foreign Language (EFL) at Japanese universities is particularly acute. It is difficult enough to make grading decisions that include attendance, attitude, completion of tasks, and test and quiz results especially when it is not uncommon that there are 30 students in one class. It is all the more challenging to determine students' true proficiency gains and final ability levels, considering that each student improves in different ways, and teachers cannot pay attention to their improvement in details. One problem is that students often regard their grades and/or teachers' evaluations of their language skills as uncontestable,

as is probably often the case in EFL environments, where they do not need to use English outside of the classroom. In other words, students are dependent upon the judgment of their abilities and/or grades as given by their teachers, regardless of their fallibility. As a matter of fact, it is impossible for teachers to know each aspect of students' abilities and progress. As such, it is recommended that teachers develop classroom assessment systems that promote student self-assessment, by thus providing for teacher-student collaboration in all final evaluations (Lynn, 1995). Possible self-assessment tools include rubrics and checklists (Schraeder, 1996) and bilingual materials on TV (Gardner, 1996).

Despite the many advantages of self-assessment, some disadvantages have also been pointed out, including questions of validity (i.e., Does the test measure what it should measure?), reliability (i.e., How stable is the measurement?), and objectivity (i.e., Is the test objective?) (Huerta-Macias, 1995). Additionally, self-assessment is not only difficult but also susceptible to student cheating (El-Koumy, 2001). On the subject of its validity, however, some reassuring studies exist. For example, the study of the Wellington Indo-Fijians, immigrants to New Zealand showed the self-reported data and assessment of listening and vocabulary had high correlation, and, therefore, it proved the validity of self-assessment (Shameem, 1998). Ross (1998) also reviewed 60 correlation studies of self-assessment to examine the validity of self-assessment. His research indicated that there were mixed results depending on language skills and learners' language learning experiences. More studies regarding the validity of self-assessment are essential.

The importance of vocabulary knowledge cannot be overemphasized. Not knowing the degree of one's lexical knowledge can hinder student learning. To date, only Laufer and Yano (2001) have extensively investigated the validity of vocabulary knowledge. In their study, 106 university students from China, Israel, and Japan served as participants. While referring to a text, the participants rated their lexical knowledge of a number of target words (stage 1), either translated them into their native language or explained them in English (stage 2), and then repeated the self-assessment task (stage 3). After comparing the lexical knowledge scores from the translation test and the two self-assessments, the researchers found that all the learners had overestimated their vocabulary knowledge. While there were no significant gender differences, discrepancies were found between the self-assessment and lexical knowledge, depending on the country and the level of lexical knowledge of the learner. As the subjects in Laufer and Yano's study (2001) were at an advanced level of language proficiency, much higher than the majority of students in Japan, a replication study was deemed necessary to determine the generalizability of their findings to less proficient students. Thus, the purpose of the present study was to investigate the

accuracy of the vocabulary knowledge self-evaluations of elementary and lower intermediate students in Japan.

Research Questions

The present study set out to answer the following questions:

1. How accurately can language learners assess their own vocabulary knowledge in context?

2. Is there any relationship between lexical knowledge and the accuracy of self-evaluation of vocabulary in context?

3. Are there gender differences in the accuracy of self-assessment?

Method

Participants

The participants in this study were 67 non-English majors taking compulsory English courses at two universities in Japan. They had each already studied English for a minimum of 6 years, focusing on their written skills since junior high school. Being at an elementary to lower intermediate, as opposed to an advanced level, as in the original study, they were assumed to have different learner characteristics as well.

Materials

As the reading passage from the original study was not appropriate for the less proficient participants of this investigation, the researcher, who also taught the participants regularly, chose a text from *News for Now 2* (Blanchard and Root, 1998) (see Appendix A). The topic was El Niño, which the researcher deemed male and female students alike to have some, though not a great deal of, background knowledge about, and to find sufficiently interesting.

Twenty words from the passage were selected and listed on a self-assessment sheet, in the same format as that of the original study (see Appendix B). Taking motivation into consideration, the target words varied in difficulty level (e.g., "weather," and "deluge"). For each word, line numbers were given to indicate its location in the reading passage. A rating scale of 0 to 2 was also provided. A rating of 0 would indicate that the subject had no knowledge of the word. Ratings of 1 or 2 would mean that a subject had partial or full knowledge of the word, respectively. As a test of their knowledge, the subjects were then asked to translate the 20 target words into Japanese (see Appendix C). To avoid confusion, these subjects were not given the option of explaining the words in English as had been the subjects in the original study.

This experiment was conducted in a normal instructional period. To begin, the students were given a copy of the reading passage to refer to throughout each of the following three stages: self-evaluation of knowledge of each of the targeted vocabulary words (stage 1), translation of the words (stage 2), and a follow-up self-assessment identical to the first.

Data Analysis

The researcher calculated scores for each word in the self-assessment sheets and labeled them. Objective lexical knowledge scores were then obtained from the translation in the second stage. Two points were given for a correct translation while none were given for an incorrect translation or for no attempt. If partial knowledge was evident, one point was given. The researcher compared the scores of the translation test with those of the two self-assessment sheets to check the accuracy of the latter. The values obtained by subtracting the self-assessment scores from the objective (i.e., lexical knowledge) scores were used to measure the discrepancy between them. Positive (+) discrepancy values indicated over-estimation, while negative (−) ones showed under-estimation. For each participant, the following values were calculated and analyzed. The range for each value is given in parentheses.

1. first self-assessment scores (from 0 to 40)

2. translation test scores (also referred to as objective scores or lexical knowledge scores) (from 0 to 40)

3. second self-assessment scores (from 0 to 40)

4. discrepancy in first self-assessment (from −40 to +40)

5. discrepancy in second self-assessment (from −40 to +40)

6. overestimated scores (from 0 to 40) —— scores with plus values in discrepancy scores

7. underestimated scores (from 0 to 40) —— scores with minus values in discrepancy scores

In terms of statistical analysis of differences, t tests were used. On the account of multiple t tests, alpha was adjusted in order to judge the statistical differences. Spearman's rank order correlation (Hatch and Lazaraton, 1991) was employed to examine the relationship between the lexical knowledge and self-assessment scores.

Results

Accuracy of Self-assessments

The statistical analysis of paired t tests indicated that the research subjects were not able to accurately evaluate their own vocabulary knowledge (see Table 1). They had a mean score of 18.49 points on the first self-assessment, although their average objective

score was only 13.42. Furthermore, the mean on the second self-evaluation was only 1.25 lower than that on the first.

Three interrelated phenomena may account for the subjects' inaccuracy. First, in the initial self-assessment, they assumed that they were familiar with some of the target words, whereas, in fact, they were not. Second, they did not realize that they had overestimated their knowledge on the first self-assessment, even after having completed the translation test. In other words, they were unaware that many of their answers on the translation test were incorrect. Third, they were not able to guess or remember the meanings of the words from context.

Table 1   Objective Test Score versus Self-assessment Score

	1st Self-assessment				2nd Self-assessment			
	mean	s.d.	min.	max.	mean	s.d.	min.	max.
Objective scores	13.42	6.66	0	32	13.42	6.66	0	32
Self-assessment	18.49	6.46	3	34	17.24	6.95	3	36
Discrepancy	5.07**	5.90	−9	17	3.82**	4.95	−13	13

** $p < .01$
objective score── scores on the translation test
self-assessment── scores on the self-assessment task
discrepancy── difference between self-assessment and objective score
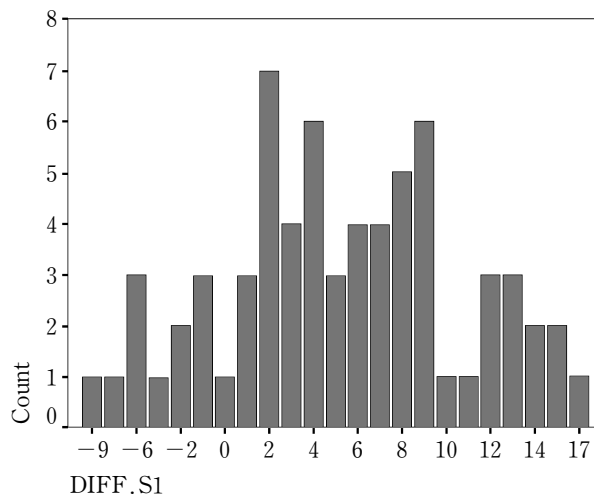Note.   Scores will be negative in the case of underestimation.

Table 2   The Discrepancy between Self-assessment and Objective Score

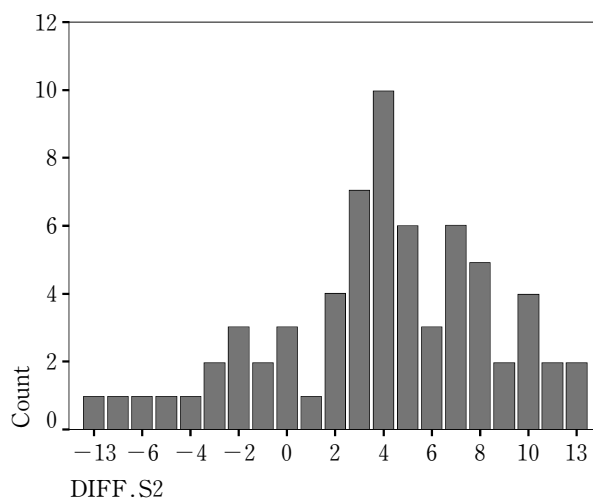|	1st Self-assessment					2nd Self-assessment				
	N	mean	s.d.	min.	max.	N	mean	s.d.	min.	max.
Over-estimation	55	7.04	4.30	1	17	52	5.88	2.98	1	13
Under-estimation	11	4.27	2.97	1	9	12	4.17	3.49	1	13
No Difference	1	0	N.A.	N.A.	N.A.	3	0	N.A.	N.A.	N.A.
Total	67	5.07	5.90	−9	17	67	3.82	4.95	−13	13

Note.   The minimum and maximum values for underestimation here were actually negative.

Now let us focus on the differences in the number of participants who overestimated or underestimated their knowledge. In the first task, 55 out of the 67 subjects (82%) overestimated their vocabulary knowledge, while 11 of them (16 %) underestimated it. Likewise, in the second self-assessment, 52 subjects (77%) overestimated and 12 (18%) underestimated. However, as illustrated in Figures 1 and 2, the constituent data from first and second self-assessments are different: the score distribution in the second self-evaluation was closer to bell-shaped curve, with a mode of four. The paired t test demonstrates that the discrepancies in the first and second assessments were statistically different. That is, the participants assessed their knowledge in the second assessment more accurately than they did in the first assessment. The mean difference between the objective score and the first self-assessment was 5.07 with a standard deviation of 5.90. On the second self-assessment, however, the mean was 3.82 with a standard deviation of 4.95, and both the mean and standard deviation are approximately one point lower. We should not overlook the fact that the subjects did not accurately assess their knowledge even in the second assessment, as mentioned earlier. The improvement on the second assessment was insufficient; thus, there were still significant differences between the objective scores and the second assessment scores.

Figure 1 Number of students for each difference between first self-assessment and objective score

Note. These numbers were calculated by subtracting the objective scores from the self-assessment scores.
negative value — underestimate
positive value — overestimate

Figure 2 Number of students for each difference between second self-assessment and objective score

Note. These numbers were calculated by subtracting the objective scores from the self-assessment scores.
　　negative value ── underestimate
　　positive value ── overestimate

Lexical Knowledge

With respect to the relationship between lexical knowledge and the discrepancy of the two self-assessment scores, Spearman's rank-order correlation (rho) revealed significant relationships on the first assessment (see Table 3). The correlation was $-.45$ at $p=.01$, which is considered moderate (Hatch and Lazaraton, 1991). It is reasonable to say the higher the lexical knowledge, the less the difference between the self-assessment scores and the objective score. In other words, the more vocabulary knowledge the subjects had, the more accurately they evaluated themselves on the first self-assessment. By contrast, there were no statistically significant correlations in the second assessment, which means that vocabulary knowledge was not relevant to accuracy of self-evaluation.

Table 3 The correlation between the Lexical Knowledge and the Discrepancy between the Self-assessment Scores and the Objective Scores

	Discrepancy First Assessment	Discrepancy Second Assessment
Correlation	$-.450^{**}$	$-.277$
Significant Differences	Yes	No

Note.　Accuracy of self-assessment was obtained by subtracting objective score scores from self-assessment score.
** $p<.01$.

Gender Differences

The results obtained in this study were not the same as in the original. Unlike in the latter, statistically significant gender differences appeared in the objective scores between males and females, with males consistently performing better on all the tests. Although no statistical difference was seen, females also tended to overestimate their knowledge more than males did. Their lexical knowledge could possibly have influenced their self-assessment scores and they could not assess their knowledge as well as males did.

Table 4 The Differences between Male and Female Learners

	First Assessment			Second Assessment		
	Self-assessment	Objective Score	Differences	Self-assessment	Objective Score	Differences
Female n=37	17.41 (5.65)	10.76 (5.25)	6.65 (6.41)	15.24 (5.23)	10.76 (5.25)	4.49 (5.08)
Male n=30	19.83 (7.20)	16.70 (6.83)	3.13 (4.61)	19.70 (8.04)	16.70 (6.83)	3.00 (4.75)
Significant Difference	No	Yes**	No	No	Yes**	No

upper number — mean
lower number — s.d.
** $p<.01$.

Conclusion

The subjects in this study were not able to accurately assess their vocabulary knowledge in either of the two self-assessments. In both trials, they thought that they knew words that, in fact, they did not. Interestingly, though the two assessment results were different. After the translation test, the self-evaluations were slightly more accurate, even though the subjects were still not able to rate their vocabulary knowledge well. We may say that self-assessment performance improved as the participants became aware that their word knowledge was lower than they had initially thought. Nevertheless, the improvement was minor.

The correlation (r＝ −.45) between lexical knowledge and the discrepancy in the first assessment was statistically significant. In the first assessment, the more vocabulary knowledge the subjects had, the better they were able to assess their vocabulary knowledge. As their skills improve, they know where they are and what they have to learn. On the other hand, the second assessment showed no such correlation with the lexical knowledge. Perhaps, subjects with less vocabulary knowledge became more aware of unfamiliar words because of the translation task.

Gender differences were seen not in the self-assessment or discrepancy scores but only in the objective scores, which seems contradictory to the correlation between the lexical knowledge and the discrepancy in the first assessment. Thus, there is room to explore this result. Unexpectedly, males did not necessarily overestimate their vocabulary more than females.

The results from the present study were inconsistent with the findings of the original research of Laufer and Yano (2001).

1. The subjects performed better in the second assessment than in the first though not enough to demonstrate accurate self-assessment.

2. There was no significant correlation between lexical knowledge and discrepancy of self-assessment in the second assessment task.

3. As for gender differences, the objective test showed statistically significant differences (although the discrepancy scores and self-assessment were the same as in the original study).

These differences may stem from differences in learner characteristics. The subjects in the original study were English majors of advanced language proficiency, and, thus, there may have been differences in motivation, confidence, and other psychological factors.

Concerning vocabulary learning in the classroom, teachers have to make sure that students know the meanings of even important or target words that they say they already

know. It is a good idea to ask them to explain those words in English or give a translation in their L1 even when they are confident about their knowledge. Additionally, TESOL professionals should keep in mind the need to make English learners aware of unfamiliar words from the first exposure. On the issue of gender differences, the findings are still inconclusive. Thus, at this point, we do not need to worry about gender differences in dealing with self-assessment of vocabulary knowledge.

Further research will be necessary in four areas. First, gender differences should be investigated since this study did not yield a clear result. Second, researchers should explore lexical factors that lead to over- and under-assessment. Third, the accuracy of self-assessment should be looked at in the light of reading and listening comprehension. One of the reasons why words are important is that they influence comprehension, so the relationships between comprehension levels and vocabulary self-assessment skills deserve attention. Fourth, the same students need to assess their own ability with different materials and the same materials at different opportunities. Various differences (e.g. background knowledge and differences in text) influences students' performance, and the same students react differently after a little while. Finally, the relationship between self-assessment ability and other language learning factors, such as motivation, should be explored.

References

Blanchard, K. & Root C. (1998). *News for Now 2*. Albert Complex, Singapore, Thomson Asia.

Brutten, S. R. (1981). An analysis of student and teacher indications of vocabulary difficulty. *RELC Journal 12*, 66-71.

Butler, P. (1997). *Toward a definition of alternative assessment in English as-Second-Language (ESL) instruction*. Issues in alternative assessment: The Japanese perspective. Washington, DC, ERIC Clearinghouse on Languages and Linguistics. (ERIC Document Reproduction Service No. ED 408 819).

El-Koumy, A. S. A. (2001). *Effects of student self-assessment on knowledge achievement and academic thinking*. Washington, DC, ERIC Clearinghouse on Languages and Linguistics. (ERIC Document Reproduction Service No. ED 452 731).

Gardner, D. (1996). Self-assessment for self-access learners. *TESOL Journal, 5* (3), 18-23.

Hatch, E. & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York, Newbury House.

Huerta-Macias, A. (1995). Alternative assessment: Responses to commonly Asked Questions. *TESOL Journal, 5* (1), 8-11.

Laufer, B. & Yano, Y. (2001). Understanding unfamiliar words in a text: Do L2 learners understand how much they don't understand? *Reading in a Foreign Language, 13* (2), 549-569.

Lynn, M. J. (1995). Caveat: Using innovative classroom assessment. *TESOL Journal, 5* (1), 36-43.

McNamara, M. J. & Deane D. (1995). Self-assessment activities: Toward autonomy in language learning. *TESOL Journal, 5* (1), 17-21.

Ross, S. (1998). Self-assessment in second language testing: a meta-analysis and analysis of experiential factors. *Language Testing, 15* (1), 1-20.

Schraeder, L. L. (1996). *Empowering ESL students in the mainstream through self-assessment and contracted learning.* Washington, DC, ERIC Clearinghouse on language and Linguistics. (ERIC Document Reproduction Service No. ED 407 850.

Shameem, N. (1998). Validating self-reported language proficiency by testing performance in an immigrant community: the Wellington Indo-Fijians. *Language Testing, 15* (1), 86-108.

Smolen, L., C. Newman, C., Wathen, T. & Dennis, L. (1995). Developing student self-assessment strategies. *TESOL Journal, 5* (1), 22-27.

Niño is coming and people in every part of the world are getting worried. El Niño is a mysterious *weather* pattern. It begins about every four years with a sudden warming of eastern Pacific Ocean along the *equator*.

El Niño is caused by changes in air movement. These changes *disrupt* ocean currents. The warm water in the Pacific Ocean moves south along the west coast of South America. It makes the cold water off the coast of Peru and Ecuador warmer. The warming of the ocean water kills many fish, sea animals and sea birds. But El Niño's effects are not limited to the local areas. The warming of the ocean *affects* weather patterns throughout the world.

El Niño is a Spanish term. It *means* "the male child and usually *refers* to the baby Jesus. Fishermen in Peru began calling the strange weather event El Niño because it usually arrived around Christmas time. Today the term El Niño is only used when the event is big enough and warm enough to bring worldwide *effects*.

The first El Niño was recorded in 1726. Since then, it has returned *approximately* every four years. This year, El Niño is warmer and started earlier than any *previous* El Niño. It is already bigger than the entire continental United States. It covers more than 6,000 miles, which is a *quarter* of the Earth's circumference, and it is still growing.

A big El Niño *occurred* during the winter of 1982 -1983. It is often referred to as the worst of the *century*. That powerful El Niño killed nearly 2,000 people and caused $ 13 *billion* in damage worldwide. It brought severe *drought* to Australia and Indonesia, storms to California, and floods to Ecuador and Peru. Advances in computers and weather satellites help weather forecasters *predict* when an El Niño will begin.

The overall effect of an El Niño is *complicated*. But forecasters are *certain* that this year's El Niño will have a *tremendous* impact of global weather patterns. The question is how serious the impact will be and where it will occur.

Already, northeastern Australia is experiencing a severe drought. South Africa, India, Ethiopia, and Indonesia are preparing for droughts also. Fires have burned out of control in Indonesia and Malaysia because El Niño *prevented* the monsoons from bringing heavy rains.

While some areas of the world may experience drought, parts of South America and the western United States may face *deluges* of rain and floods.

There is no doubt that El Niño will have a devastating effect on the world's food supply. Climatologists think changes in the world's weather patterns will peak this winter and return to normal next summer. If their predictions are right, summer cannot come too soon.

Taken from Blanchard, K. & Root C. (1998) *News for Now*

Appendix B: Self-assessment Sheet

大学名：   学部：

名前：

丸をつけなさい　男性　女性

下線の単語をみて，この読み物に使われている意味がわかるかどうか，次の中から選びなさい。

単語	行	理解の程度		
1．weather	3	0	1	2
2．equator	5	0	1	2
3．disrupt	7	0	1	2
4．affect	15	0	1	2
5．mean	16	0	1	2
6．refer	17	0	1	2
7．effect	23	0	1	2
8．approximately	25	0	1	2
9．previous	27	0	1	2
10．quarter	30	0	1	2
11．occur	32	0	1	2
12．century	34	0	1	2
13．billion	35	0	1	2
14．drought	36	0	1	2
15．predict	40	0	1	2
16．complicated	42	0	1	2
17．certain	43	0	1	2
18．tremendous	44	0	1	2
19．prevent	52	0	1	2
20．deluge	56	0	1	2

0．全くわからない

1．少しわかる

2．よくわかる

Note.　The second column gives the line number where each word can be found in the text. The same format was used for the translation test.

Appendix C: The translation test

大学名：　　　　　　　　　　　　　　　　　　　　学部：

名前：

丸をつけなさい　男性　女性

下線の単語をみて，この読み物に使われている意味を日本語で書きなさい。

単語	行	意味
1．weather	3	
2．equator	5	
3．disrupt	7	
4．affect	15	
5．mean	16	
6．refer	17	
7．effect	23	
8．approximately	25	
9．previous	27	
10．quarter	30	
11．occur	32	
12．century	34	
13．billion	35	
14．drought	36	
15．predict	40	
16．complicated	42	
17．certain	43	
18．tremendous	44	
19．prevent	52	
20．deluge	56	

Note.　The number next to each word is line number.