Shared Identities: Our Interweaving Threads

# Measuring involvement load—"evaluation"

**Matthew Walsh**
*University of Birmingham*

**Reference data:**
Walsh, M. (2009). Measuring involvement load—"evaluation". In A. M. Stoke (Ed.), *JALT2008 Conference Proceedings*. Tokyo: JALT.

This paper will summarize quantitative research comparing task characteristics in terms of factors leading to incidental vocabulary acquisition in EFL. The investigation uses Laufer and Hulstijn's (2001a) construct of Task-Induced Involvement. The participants consisted of 223 tenth grade high school learners from two different schools. The two populations were found to have quite differing levels of L2 vocabulary knowledge from the start. Two conditions were compared: a task that creates "moderate evaluation", and a task that creates "strong evaluation". In both populations, the task creating a "strong evaluation" (original sentences) was found to be more effective at enabling retention one-week post task than the "moderate evaluation" (gap-fill) task. Furthermore, the population with the higher initial level of vocabulary knowledge seemed to benefit relatively more from the "strong evaluation" task. For high school learners in Japan, learner-initiated composition using L2 words new to the learner proved to be an effective way of acquiring them. A Task-based learning approach focusing on specific content is suggested as a way to enable this output through creating context.

語彙力の異なる二校の高校生223人を対象とした実験を実施。LauferとHulstijnの仮説に基づき、それぞれの高校の生徒をグループに分け、新しい単語を使って自で文を作るものと文章の空白を埋めるものの二種類のタスクを与える。その結果、自分で文を作るタスクを与えられたグループの方が語彙の習得が効果的であることが認められた。特に単語のレベルが高い生徒にこの効果が顕著にみられた。ゆえに、語彙を効果的に習得するには、自ら文章を創作することが重要であることが判明した。

## Review of the literature

The Construct of Task-Induced Involvement (Laufer & Hulstijn, 2001a) and the subsequently empirically tested Involvement Load Hypothesis (Laufer & Hulstijn 2001b) aims to categorize, operationalize, and quantify the extent and way second language pedagogical tasks focus on new vocabulary items so that incidental acquisition or retention of the vocabulary items occurs.

### *Theory the construct aims to address*

The Involvement Load Hypothesis (Laufer & Hulstijn, 2001b) is preceded by several proposed frameworks for the role of noticing, attention, and consciousness in SLA (Schmidt, 1994, 2000; Sharwood Smith, 1981; Gass, 1988, 1999; Swain, 1985; Robinson, 1995). These arguments together form what is referred to as a "weak interface" position as to the role of consciousness in SLA in opposition to the "no interface" (Dulay & Burt, 1973; Krashen, 1982) position, which asserts that acquisition proceeds at an unconscious level and pedagogical measures have no lasting effect. The "weak interface" position posits that consciousness-raising of language features occurring as a result of learning exercises can affect the acquisition process. It is important to note that the actual experiments evidencing the "no interface" position tested for grammar usage. Laufer and Hulstijn differ in that from the outset, they purposefully explore acquisition of vocabulary rather than grammar.

Building on a different academic discussion more about cognitive psychology than second language acquisition, The Involvement Load Hypothesis aims to examine the workings of memory and what cognitive processes might more efficiently make learned items not be forgotten. In other words, why are some things remembered and other things not. The Construct of Task-Induced Involvement sets out to operationalize the concept of depth or "Levels of Processing" (Craik & Lockhart, 1972). Craik and Lockhart summarize how preceding their "depth of processing" model, the prevalent "multistore" model of memory envisioned short and long-term containers for storage of information (Murdock, 1967; Atkinson & Shiffrin, 1968). In the framework, noticed or "registered" information is transferred to the "short-term store" (STS), where it then might again be transferred to "long term store" (LTS) via things like rehearsal, repetition or association. It was believed that holding the information in STS for a sufficient amount of time would transfer it to LTS. Critics of this model were unsure of the idea of these memory containers being independent or demarcated, or that information must necessarily "pass through" STS as a requirement to reaching LTS (Tulving & Patterson, 1968; Shalice & Warrington, 1970). Instead, Craik and Lockhart (1972) propose that information is processed at different depths or levels determining different degrees of what they term "memory trace persistence". A greater degree of semantic or cognitive analysis implies a greater depth of processing. Memory is "viewed as a continuum from the transient products of sensory analysis to the highly durable products of semantic-associative operations" (Craik and Lockhart, 1972). In terms of foreign/second language pedagogic methodology, this view was the theoretical basis for "the keyword method" of learning vocabulary (Atkison, 1975). Learners created what Atkinson called an "acoustic link" and an "imaginary link" in their minds to the target word. Empirical research carried out by Atkison (1975) and others later did prove this method improved the rate of recall. For a summary see Rodriguez and Sadoski (2000). The primary concept that "…memory performance is determined far more by the nature of the processing activities engaged in by the learner than it is by the intention to learn per se" (Eysenck, 1982) is still universally accepted by cognitive psychologists.

It can be assumed that "deeper is better" but an unambiguous definition of a "level of processing" has remained elusive. As Baddely (1978) concludes, "all attempts to measure processing depth appear to have been unsuccessful". The aforementioned experiments (Rodriguez & Sadoski, 2000) compared specific techniques for memorization. However, without a construct for describing which elements within the techniques are more effective, which other pedagogic tasks would create more processing depth than others cannot be predicted. The discussion would be limited to studies comparing specific techniques.

Addressing this academic discussion Laufer and Hulstijn (2001a) attempt to quantify depth of processing as it relates to learning vocabulary by creating "The Construct of Task-Induced Involvement" which divides the nature of processing activities created by a task into three components, "need", "search", and "evaluation". A pedagogic task containing more of the components would produce better retention of the vocabulary items. This is explained in more detail below. The components themselves were abstracted from elements found in other effectiveness studies. For examples, see Table 1 below.

## The construct of task-induced involvement

Of the three components mentioned above, "need" is actually more of a driver than a "processing activity". It is the *motivational* component of the construct and simply means the need to achieve or a drive to comply with the task requirements. In the case of vocabulary items, this translates into the need to understand or perhaps produce the word in order to complete the task. "search" is one of the *cognitive* components of the construct and means the process of finding the meaning of unknown L2 word, or finding an L2 word expressing a certain concept. In second language pedagogical tasks this would typically translates into referring to a dictionary, or negotiating meaning with a pair-work partner. The last component "evaluation" is the cognitive process of comparing a given word with the surrounding words, meanings, or contexts. This could mean any sort of gap-fill exercise where the surrounding words, meanings, or contexts obligate a certain choice of words, or in more of a productive sense, where the learner must provide other words, meanings, and contexts that match the word to be used. "evaluation" is the process of making a selective decision considering the form, usage, collocations, and meaning in general of a certain word.

Part of the construct is also a moderate/strong distinction in the case of *need* and *evaluation*. A *search* is something that either is done or not done but in the case of need, there is a difference in involvement if the learner initiates the need for a word; as in if the learner had a concept that they wished to express as part of an original composition rather than if the task had simply asked them to use a given word. This is also true in the case of evaluation. Recognizing differences between words such as in a fill-in task is less "involving" than making a sentence with one and choosing additional words to combine with the target in use. The distinction between moderate and strong *need* or *evaluation* is that of externally provided versus learner initiated opportunities for use. Examining this distinction further on a theoretical level certainly would be interesting but a bottom-up examination based on the empirical data of past experimental research

also reveals the durability of the construct of task-induced involvement load including of course the moderate/strong distinction. This is illustrated in the table below.

### Table 1. Previous research in terms of Involvement Load

### (from Laufer & Hulstijn, 2001a)

| The more effective task | The less effective task | Reference |
|---|---|---|
| Meaning selected from several options +evaluation | Meaning explained by synonym | Hulstijn 1992 |
| Meaning looked up in a dictionary +search | Reading with/without guessing +/–search | Knight 1994; Luppescu and Day 1993 |
| Meaning looked up in a dictionary +search | Meaning provided in a marginal gloss | Hulstijn *et al.* 1996 |
| Meaning negotiated ++ need, +search | Meaning not negotiated | Newton 1995 |
| Negotiated input +search | Premodified input | R. Ellis *et al.* 1994 |
| Used in original sentences ++evaluation | Used in non-original sentences | Joe 1995, 1998 |
| Used in a composition (L1–L2 look up) ++evaluation | Encountered in a reading task (L2–L1 look up) –/+evaluation | Hulstijn and Trompetter 1998 |
| Interactionally modified output ++evaluation | Interactionally modified input | R. Ellis and He 1999 |
| Reading and a series of vocabulary exercises +evaluation/++evaluation | Reading only (and inferring meaning) –/+evaluation | Paribakht and Wesche 1997 |
| Reading, words looked up in a dictionary +search | Reading only, words not looked up | Cho and Krashen 1994 |

The table below (Table 2) of the scoring system illustrates that the plusses (+) indicate the degree of involvement in any of the components and the total amount of plusses indicates the total involvement load of a given task. -/+ indicates that the component was only presents for select parts within the task.

### Table 2. Scoring of Task Involvement Load

Involvement Load

| | Need | Search | Evaluation |
|---|---|---|---|
| none | ( ) | ( ) | ( ) |
| moderate | (+) | (+) | (+) |
| strong | (++) | | (++) |

## Research questions

The present research compares the weight of moderate and strong evaluation. Method is described in detail below addressing these questions:

(1) Similar research has been conducted in Israel and Amsterdam (Laufer & Hulstijn, 2001b) with higher proficiency level learners. Will the strength of "learner initiated" composition (strong evaluation) tasks hold out with lower vocabulary level learners in Japan where open productive activities are traditionally rare in formal education?

(2) What are the implications of the findings to classroom practice?

## Method

### *The tasks*

Both tasks accompany a 326-word text studied as part of the normal syllabus. Entitled "Child Labor", the text was written by the instructor/researcher with EFL pedagogical purposes in mind and was made to be within reach of the learners in terms of complexity and vocabulary (Appendix 1). Both versions of the worksheet involve the text first being read for comprehension. The comprehension questions at the end of the texts are engineered so that the target vocabulary must be understood in order to answer them. This creates a "need", the motivational component of Laufer and Hulstijn's (2001a) construct. In this case, the comprehension questions create a "moderate need" because the external agent of the task imposes the need. Both versions of the worksheet were also identical in that they didn't invoke the component "search" (Laufer & Hulstijn, 2001a). One of the tasks, "Task A", does however involve searching for word but not in the involvement load sense. Blanked out spaces in the text are matched to words with L1 glosses arranged in a jumbled matrix. According to Laufer (personal communication, February 11, 2008), this would not be considered a "search" since the learner does not need to hold the unknown word in memory (in the phonological loop) long enough to, for example, consult a dictionary. In other words, there is not enough effort in simply scanning through the matrix. Task A however does create Laufer and Hulstijn's cognitive component of involvement, "evaluation", because the learners' goal is to consider or "evaluate" whether target words and sentence context and/or language match. In this case it is a "moderate" evaluation because the contexts and surrounding language are externally provided or given. If the learner were creating original sentences with new vocabulary, the level of evaluation would be considered "strong" according to the framework. This was the case with Task B. In Task B target vocabulary items were not gapped out in the text as in Task A but underlined and on the side there was a list, in order, of the words again with L1 glosses and a blank line for the learners to compose an original sentence using the vocabulary item.

In summary, both versions of the worksheet have "need" from the comprehension questions, neither has "search", and the varying condition is between the "moderate" evaluation of the matching gap-fill Task A and the "strong" evaluation of original sentence Task B. I will next describe the environment of and participants in the experiment, choices in the process of creating it, and considerations in its execution.

### *Populations and groups*

As is illustrated below, (Table 3) the experiment described in this paper was conducted identically with two separate populations consisting of three groups each. The two populations both consisted of tenth grade high school students in Japan but from different schools with different levels of academic achievement or aptitude as determined by entrance requirements and the resulting stratification of the educational system.

In order to verify the difference between the populations specifically in terms of knowledge of vocabulary in English, Nation's (1990, pp. 265-266) 2,000-word level test was used. Combining the scores of all 3 groups from each population,

**Table 3. Populations and groups**

| Population (school) | Group (class) | Number of participants in group after missing data | Role in experiment |
|---|---|---|---|
| School A | 10th grade class 3 | 37 | Control (No Task) |
| School A | 10th grade class 4 | 38 | Task A (Gap Fill) |
| School A | 10th grade class 2 | 37 | Task B (Orig Sentence) |
| School B | 10th grade class 14 | 30 | Control (No Task) |
| School B | 10th grade class 16 | 30 | Task A (Gap Fill) |
| School B | 10th grade class 15 | 31 | Task B (Orig Sentence) |
| | | total=203 | |

School A's average score was 13.16 (73%), or presumed knowledge of 1462 words whereas School B's average was much lower: 8.07 (45%), or presumed knowledge of only 897 words.

### *Choice of lexical items to be tested*

In order to identify words from the text likely to be of difficulty to the groups, the experimental populations' peers and seniors were asked to assist. Students not taking part in the experiment but from the same schools, a total of 28 spread evenly between the 10th, 11th, and 12th, grade, were given the text and asked to read it and underline as many unfamiliar words as they could find. From the 28 students, a total of 33 words were underlined and several of them were underlined by as many of as 16 of the students spanning all three of the grades. The words chosen to be targeted were the fifteen most commonly underlined words. The words can be seen in Appendix 2.

As an interesting aside, these words were analyzed in terms of frequency using Cobb's (2008) online vocabulary profiler, which divides words from a text into frequency bands K1, K2, AWL, and Off-List. "K1" means the 1000 most common words in selected corpora of English, "K2" from 1001 to 2000 and AWL refers to "Academic Word List", a well-known list of words beyond the 2000 mark that are common in academic settings. For these populations targeting K2 and the AWL would have produced only about 56% of what they actually claimed as unknown and targeted about 31% words already known to the majority. Ascertaining the problem words for these specific populations by surveying peers produced a denser set of data to work with than would have been by using frequency lists. The table below illustrates the words selected by peers as compared to their place on the lists.

### *Pre and post-tests*

The pre and posttests of knowledge of the targeted vocabulary items was a receptive test. The translation test asks for an L1 (Japanese) equivalent of the target words. Both can be seen in Appendix 3.

During the scoring of the pre- and post-tests three raters were present, one native English speaker proficient in

## Table 4. Wordlists versus peer selected target vocabulary

| *Top peer selected as unknown* | *(out of 28 students)* | *Status of peer selected word in 'vocabprofiler' lists* | *All content words in text on 'vocabprofiler lists minus K1 (0-1000)* |
|---|---|---|---|
| gymnasium | (16) | Off-list | *From AWL* |
| agriculture | (16) | K2 (1001-2000) | aware |
| (cash) crops | (15) | K2 (1001-2000) | demonstrations |
| to be aware | (15) | AWL | globalization |
| blame | (15) | K2 (1001-2000) | labor * |
| demonstration | (14) | AWL | *From K2 (1001-2000)* |
| fiber | (13) | Off-list | agriculture |
| increase | (12) | K1 (1-1000) | blame |
| plantation | (10) | Off-list | clothing |
| warmth | (8) | K2 (1001-2000) | coffee * |
| globalization | (8) | AWL | crop |
| sweatshop | (8) | AWL | gap * |
| partly | (8) | K1 (1-1000) | international * |
| shrimp | (7) | Off-list | typical |
| clothing | (6) | K2 (1001-2000) | warmth |
| typical | (6) | K2 (1001-2000) | *From 'Off-List'* |
| | | | bananas * |
| | | | cash |
| | | | coconut * |
| | | | fiber |
| | | | gymnasium |
| | | | Japan * |
| | | | Nike * |
| | | | plantation |
| | | | shrimp |
| | | | soccer * |
| | | | * = Not peer selected as unknown |

Japanese and two native Japanese speakers proficient in English. When questions arose concerning which Japanese words were to be considered valid as equivalents to the targets, a consensus was reached and adhered to throughout the marking process, going back to mark previously marked questions when necessary. One point was given on an all-or-nothing basis for each target word correctly translated. Laufer (personal communication, February 20, 2008) suggested giving half a point for semantically close but not exact translations but there was no occasion where this was necessary.

It was announced that the pre-post tests and the vocabulary items were completely unrelated to the course and in terms of the exam, students would not be held responsible for them later. Along the same lines a similar announcement was made concerning the tasks, saying that the text "Child Labor" would be tested for general comprehension only and the words would not be on the test. Finally, pre-tests, task sheets, and post-tests were collected immediately after use in order to prevent the more diligent students from looking up the words later. As the sheets were being collected, students were asked not to talk or discuss the problems with each other.

## Results and analysis

Looking at Table 5, we can see a substantial improvement in scores resulting from both types of exercises. Multiple T-tests also verified that compared to control groups, both conditions (gap-fill moderate evaluation, and original sentences strong evaluation) for both populations (School A, and School B) resulted in improvement in scores that reach statistical significance, defined $P<0.05$ as is common in this type of research. The T-tests also verified that the slight improvements we see in the control groups between pre and post-test were not statistically significant.

An interesting emerging finding can be seen when comparing the gains from the gap-fill exercise to those from the original sentences exercise across populations. Although

## Table 5. Results of the study

| | Control (No-Task) | Moderate Evaluation (Gap-Fill) | Strong Evaluation (Original Sentences) |
|---|---|---|---|
| School B (Vocabulary level test score 897) | Pre 2.52(17%)<br>Post 2.7 (18%)<br>*Improved 0.18 (1%)* | Pre 1.9 (13%)<br>Post 6.24 (42%)<br>*Improved 4.34 (29%)* | Pre 2.14 (14%)<br>Post 6.79 (45%)<br>*Improved 4.65 (31%)* |
| School A (Vocabulary level test score 1462) | Pre 5.93 (40%)<br>Post 6.43 (43%)<br>*Improved 0.5 (3%)* | Pre 6.18 (41%)<br>Post 10.11 (67%)<br>*Improved 3.93 (26%)* | Pre 6.89 (46%)<br>Post 11.81 (79%)<br>*Improved 4.92 (33%)* |
| * Pre and post-tests scores are out of a total of 15 possible correct answers. | | | |
| * Vocabulary level indicates mean score on Nation's 2000 word vocabulary test. | | | |

| Comparing pre and post-tests, how much better was original sentences than gap fill? (Moderate vs. Strong Evaluation) | School B (Vocabulary level 897) = +0.31 (2%) P=0.39 *(not significant)* | School A (Vocabulary level 1462) = +0.99 (7%) P=0.009 *(significant)* |
|---|---|---|

original sentences (strong evaluation) results in more improvement than gap-fill (moderate evaluation) in either population, there is a difference as to the extent that is does. School B, the population with the lower initial vocabulary score according to Nation's 2000 word vocabulary test (Nation, 1990) gained relatively less from doing the original sentences exercise over the gap-fill exercise than did the population with the higher initial vocabulary knowledge (School A). In fact, the difference does not reach statistical significance in the lower vocabulary population. A T-test gives us a p-value of 0.39 and the null hypothesis cannot be rejected. In this study for the population with a lower level of knowledge of vocabulary, making original sentences seems not to have had that much advantage over gap-fill exercises, but for the higher vocabulary knowledge population the difference was drastic.

The results of this experiment suggest there is a difference in the "processing depth", or weight of strong evaluation compared to moderate evaluation according to the learner's level of vocabulary knowledge, but does not offer clues as to why. A number of possible explanations for this difference could be explored. Perhaps for population with more limited vocabulary knowledge the original sentences weren't much better at creating a context for the unknown word than the gap-fill exercise. The definition of "strong evaluation" includes that the learner creates the context whereas "moderate evaluation" connotes fitting the new word into a given context as part of the task. For a learner with more limited vocabulary knowledge, the context provided in the reading text may be less understood than the L1 gloss provided in the worksheet. Such a learner would be tempted to pay less attention to the text and create an original sentence depending greatly on the L1 gloss. Along the same lines, a lower level learner would have less of a L2 lexicon to draw from while creating a context for the original sentence. The original sentence could easily end up a lexically sparse transliteration of a context one would more likely find the L1 equivalent in. Although both types of tasks provided an L1 gloss, the gap-fill may have had low-level learners attention more focused on the context within the text. Looking at the text-provided contexts could not be avoided. This process of considering the new word's context, usage and collocation may have created a deeper level of processing than a learner who merely scanned the text for gist and directly referred to

the L1 gloss in order to make an original sentence, as would be more the case in the situation described above.

The question remains as to whether for the lower proficiency level population the gap-fill was more effective or the original sentences were less effective.

Even if not reaching statistical significance, the original sentence task did outperform the gap-fill task for the population with the lower level of vocabulary knowledge. This perhaps demonstrates the strength of learner-created contexts for vocabulary acquisition even for the population with the lower level of vocabulary knowledge, those who might have been at a disadvantage to create original sentences or understanding the contexts in the L2 text.

Turning to the population with higher vocabulary knowledge level, to explain the drastic difference in retention between the groups, the opposite may be true. With a higher level of vocabulary knowledge this population is able to tease contexts, usage, and collocations from input (the text). They would also have greater freedom to create contexts that fit with the new word utilizing their much larger L2 lexicon. In the literature review, the keyword method was explained. This method aimed to create deeper processing by creating mental associations and purposefully connecting them to the new word. Although artificial, it creates a context containing associations. For the higher vocabulary group it may be easier to create associations with other words in a task where strong evaluation is a component.

The answer to the first research question is yes, strong evaluation will prove more effective than moderate evaluation with high school learners in Japan whose level of proficiency is much lower than the learners in Laufer and Hulstijn's (2001b) research. However, this seems to be more true for learners with a higher level of knowledge of vocabulary to start out with. These findings have implications for language pedagogy which will be explained below.

## Implications

To answer the second research question, the findings suggest that for the typical high school student in Japan learner-initiated composition tasks with new vocabulary items is efficient for acquisition of them. It appears that context is important in both input and output. The task for the educator is then how to construct tasks that take advantage of this. A framework is needed that facilitates output and at the same time pulls attention to context in input so that new words will be noticed. Utilizing The Construct of Task-Induced Involvement would entail seeking to create as many of the conditions as possible, noting that according to the findings of this study, output, in specific learner-initiated composition involving target vocabulary is definitely an advantageous, if not the most advantageous element of the construct to have in place.

Two things would address these aims: First, the tasks should be set in a framework that facilitates learner-initiated output such as Task-based Learning (TBL), and second, that the task share a theme, topic or a series of interlinked topics as in approaches like CBI (Content-Based Instruction; Brinton, 2003) or ESP (English for Specific Purposes). The use of a theme increases the possibility that the same vocabulary items will appear repetitively in a series of tasks and contexts. This would hopefully be of particular use to

learners with a lower vocabulary proficiency. The theme would also lend itself to use of infrequent words like those on the upper bands of frequency lists by creating a meaning space particular to that domain.

## The TBL framework

Laufer and Hulstijn (2001b) mention TBL, in particular the Skehan (1996) model as a sound framework to put the theory and empirical research they aim to stimulate to pedagogical practice. To illustrate how the findings of this research can be utilized, it is suggested to adapt the Willis (1996) framework for TBL to Content-based instruction with an added element of learner-initiated composition.

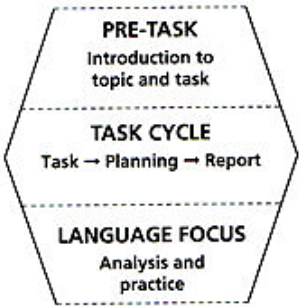


**Figure 1. Willis's model of Task-Based Learning (TBL) (Willis (1996, p. 60)**

As can be seen above, in the Willis framework, after some sort of activity introducing the topic and task, the students do the task. After that, time is given for the learners to prepare a "report". The effect of planning time on the quality of language production in testing situations has been thoroughly researched. For a summary see Ellis (2003, p. 293). Willis (1996) however suggests the planning time more as a time for the learners to get organized as to what meanings they wish to convey. In a typical group-work or pair-work situation, a report would mean one representative presenting the pair or group's findings to the classroom as a whole. For the current purposes, this report phase can be done twice but in different modes. The first report would be oral and done shortly after task completion. The next report is a written one and shared within groups, and then with the class as a whole the next time the group meets. TBL differs from methodological frameworks for language learning such as "presentation, practice, production" (PPP) in that, as is the case with most interpretations of communicative language teaching (CLT), for production, the linguistic resources to be used by the learner to complete the task are not to be dictated by the learning materials or the teacher, as Willis (1996, p.24) states "…learners are free to choose whatever language forms they choose to convey what they mean, in order to fulfill, as well as they can, the task goals". Samuda and Bygate (2008) suggest running the task cycle twice in order to introduce the target structure after a meaning space has been created by the first cycle. In the proposed framework, no target structures are introduced. However, as a necessity learners will have to use unfamiliar vocabulary items which they will have encountered earlier in the introduction or task-cycle. These writing assignments can be recycled for a final consciousness-raising activity by lifting common mistakes and focusing on them in accordance with the Willis (1996) precept of focus on form *after* the task cycle.

This research has suggested merely using new vocabulary in learner-initiated composition very efficiently promotes acquisition of them. By following a TBL framework with an added written follow-up report phase, aside from the use of language during the task itself, there are two obligatory occasions when learner-initiated composition with new vocabulary items occurs; once during the in-class group report and again for the written report.

## An example of a theme

Any number of topics or themes would introduce new vocabulary items particular to the domain, but the following is an example appropriate for "International Education" an area often dealt with by the EFL Department in high school in Japan. UNESCO's "Decade of Education for Sustainable Development" (DESD) started in 2005 and represents the consensus of what issues the UN member countries think should be addressed in education. In a high school in Japan, courses aimed at such topics would be typically considered "International Education".

As can be seen by the graphical representation, ESD can address economic, environmental, and social issues or those that lie in between. Willis gives examples of 6 types of tasks that can be used in her model of TBL: listing, ordering and sorting, comparing, problem solving, sharing personal experiences, and creative tasks (writing, media projects). It is easy to imagine how to combine TBL with ESD. An example could be as simple as "Rank the UN's Millennium Development Goals in terms of importance". A topic like this will immediately provide a clear meaning space and a plethora of unfamiliar L2 vocabulary items to be

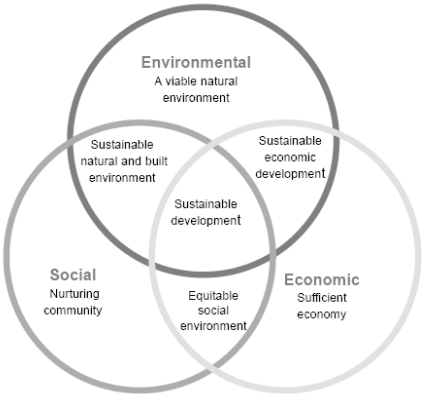


**Figure 2. Issues within the realm of Education for Sustainable Development (ESD)**

used in learner-initiated output in both the oral and written reports. Continuing such a content course with a series of related topics as could be imagined with ESD, the same words would most probably be encountered again giving the learners the chance integrate new words into input as the theme and words associated with it becomes more familiar.

## Conclusion

As in Laufer and Hulstijn's (2001b) research on their Involvement Load Hypothesis in Amsterdam and Israel with higher level learners, with all other variables equal, for high school learners in Japan, with lower levels of vocabulary acquisition, strong evaluation, or learner-initiated

composition using new vocabulary words appears to be more effective at facilitating acquisition than moderate evaluation, or comparing new words with contexts and matching them. The data suggest however that the advantages of learner composition may be lower if the learner does not have enough vocabulary knowledge to start out with. These results imply that for the language teacher in Japan wishing to aid learners in increasing their L2 lexicon, using a teaching framework that involves learner-initiated composition is advantageous. One way to do this would be to adapt TBL to a specific content area in order that learners would explore new vocabulary associated particularly with that domain. It would be reasonable to suppose that doing so would aid learners in creating a context to understand, use, and acquire new vocabulary.

**Matthew Walsh** lives in Osaka, Japan and teaches EFL at high school.

## References

Atkinson, R. C. (1975). Mnemotechnics in second-language learning. *American Psychologist, 30,* 821-828.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory, 2,* New York: Academic Press.

Baddely, A. (1978). The trouble with levels: A reexamination of Craik and Lockharts framework for memory research. In *Psychological Review*, *85(3)*, 139-152.

Brinton, D. (2003). Content-based instruction. In D. Nunan (Ed.), *Practical English language teaching* (pp. 199-224). New York: McGraw Hill.

Cobb, T. Web *Vocabprofile* [accessed October 2008 from http://www.lextutor.ca/vp/ ], an adaptation of Heatley & Nation's (1994) Range.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour 11*, 671-84.

Dulay, H. C., & Burt, M. K. (1973). Should we teach children syntax? *Language Learning, 23*(2)*,* 245-258.

Ellis, R. (2003). *Task-based language learning and teaching*, Oxford: Oxford University Press.

Eysenck, M. W. (1982). Incidental learning and orienting tasks. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 197-228). New York: Academic Press.

Gass, S. M. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics, 9,* 198-217.

Gass, S. M. (1999). Discussion: Incidental vocabulary acquisition. *Studies in Second Language Acquisition, 21,* 319-333.

Krashen, S. (1982). *Principles and practice in second language acquisition.* Oxford: Pergamon.

Laufer, B. (2008a). Conversation with Matthew Walsh, 11th February.

Laufer, B. (2008b). Conversation with Matthew Walsh, 20th February.

Laufer, B., & Hulstijn, J. (2001a). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22(1),* 1-26.

Laufer, B., & Hulstijn, J. (2001b). Some empirical evidence for the involvement load hypothesis. *Language Learning, 51*(3)*,* 539-558.

Murdock, B. B. Jr. (1967). Recent developments in short-term memory. *British Journal of Psychology, 58,* 421-433.

Nation, I. S. P. (1990). *Teaching and learning vocabulary.* Boston: Heinle & Heinle.

Robinson, P. (1995). Attention, memory and the noticing hypothesis. *Language Learning, 45*(2)*,* 283-331.

Rodriguez, M., & Sadoski, M. (2000). Effects of rote, context, keyword, and context/keyword methods on retention of vocabulary in EFL classrooms. *Language Learning, 50,* 385-412.

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning.* New York: Palgrave Macmillan.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. In J. H. Hulstijn & R. Schmidt (Eds.), *Consciousness in second language learning.* AILA Review 11 (pp. 11-26).

Schmidt, R. (2000). Attention. In P. Robinson (Ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press.

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuro-psychological study. *Quarterly Journal of Experimental Psychology 22,* 261-273.

Sharwood Smith, M. (1981). Consciousness raising and the second language learner *Applied Linguistics, 11*(2)*,* 159-168.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics, 17*(1)*,* 38-62.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.

Tulving, E., & Patterson, R. D. (1968). Functional units and retrieval processes in free recall. *Journal of Experimental Psychology, 77,* 239-248.

Willis, J. (1996). *A framework for task-based learning.* Edinburgh Gate: Longman.

## Appendix 1

### *The "child labor" text*

Is globalization a good thing? Many people think not. Sometimes globalization causes problems. One of them is the increase of "child labor".

In some countries, children work and don't go to school. Most work in agriculture (farming) for cash crops. A cash crop is a crop made to sell for money, not to eat or trade with. These crops are usually sold to a foreign country. Sometimes a foreign company owns the plantation the children work at. Some examples of cash crops are coffee, bananas, or shrimp. Coconut fiber, or "natadecoco" is a famous cash crop sold to Japan.

Some children work in factories that make things for foreign countries like Japan and the U.S.. Some people call these factories "sweatshops" (like a hot gymnasium) because they are so hot and the work is so hard. A typical "sweatshop" is a big clothing factory that makes things for big international companies like Nike or The Gap. Many of the soccer balls we use are made by children. Whether it be cash crops or factories, the children are working to sell things to developed countries. We are partly to blame for this problem.

Some people in Japan and other countries know about this problem and buy things from fair trade companies like the people tree that don't use child labor and pay the workers better than the "sweatshops". Some people have demonstrations against free trade because "free trade" means that more foreign companies can do things like use child labor in developing countries. Because of these demonstrations, some big companies like Nike changed and became more fair, but many haven't yet.

Public opinion can change the world, but first people must be aware of the problems. We will march on Sunday 6/8 and we will ask people: Do you know about this problem? At the least, people will see our warmth and kindness and they will start to think. It is a start. Let's go!

## Appendix 2

### *Peer selected unknown words from the text*

gymnasium 16
agriculture 16
cash crops / crops 15
(to be) aware 15
to blame 15
demonstration 14
fiber 13
increase 12
plantation 10
warmth 8
globalization 8
sweatshop 8
partly 8
shrimp 7
clothing 6
typical 6
farming 5

trade 5
to march 4
coconut 3
fair 3
developed 2
against 2
companies 2
public 2
whether 1
developing 1
causes 1
Nike 1
whether 1
(child) labor 1
opinion 1
(at the) least 1

## Appendix 3

### *Pre and post-tests*

### *Pre-test version*

How many of these words do you know? Write the translation on the answer sheet.

1. This is a big gymnasium.
2. Many people work in agriculture.
3. This is where they sells their crops.
4. I am aware of the problem.
5. He is to blame for what happened.
6. There was a demonstration against the new law.
7. This plant has a lot of fiber.
8. Crime has increased in 2007.
9. The coffee is made at a plantation.
10. Some people think globalization is a bad thing.
11. The factory was like a sweatshop.
12. I partly agree with you.
13. His farm makes shrimp.
14. I buy my clothing on the Internet.
15. He is a typical university student.

### *Post-test version*

How many of these words do you know? Write the translation on the answer sheet.

1. Many people work in agriculture.
2. He is to blame for what happened.
3. His farm makes shrimp.
4. Some people think globalization is a bad thing.
5. I am aware of the problem.
6. I buy my clothing on the Internet.
7. Crime has increased in 2007.
8. I partly agree with you.
9. This is a big gymnasium.
10. This is where they sells their crops.

11. He is a typical university student.
12. The factory was like a sweatshop.
13. This plant has a lot of fiber.
14. There was a demonstration against the new law.
15. The coffee is made at a plantation.

## *Answer Sheet*

Vocabulary check: Write the Japanese word for the words squared in the sentences.

| | |
|---|---|
| 1 | |
| 2 | |
| 3 | |
| 4 | |
| 5 | |
| 6 | |
| 7 | |
| 8 | |
| 9 | |
| 10 | |
| 11 | |
| 12 | |
| 13 | |
| 14 | |
| 15 | |