

Local Item Dependence on the Vocabulary Levels Test Revisited

Tadamitsu Kamimoto

Kumamoto Gakuen University

doi: <http://dx.doi.org/10.7820/vli.v03.2.kamimoto>

Abstract

The purpose of this study was to address the question of local item dependence (LID) on the Vocabulary Levels Test (VLT). The test format adopts a matching format consisting of six words and three definitions. A review of the literature suggested that items presented in such a format cannot be considered independent. However, Schmitt, Schmitt, and Clapham reported that according to Rasch analysis, most of the items performed independently. The present study examined effects of LID from a non-Rasch approach. A set of three clusters was collapsed into one large cluster of 18 words and 9 definitions on the assumption that such a treatment would make items practically independent. Both the collapsed VLT and the intact VLT were given to 114 Japanese English as a Foreign Language (EFL) students at an interval of one week. Results showed that scores were 15% higher on the original VLT than on the collapsed form. Furthermore, an LID index based on the correct/wrong response types between the two tests indicated that scores on the original VLT were about 19% inflated. Implications were drawn and discussed.

1 Background

Nation's (1990) Vocabulary Levels Test (VLT) and its revised forms (Nation, 2001; Schmitt, 2000; Schmitt, Schmitt, & Clapham, 2001) are still in use (e.g., Akbarian, 2010; Li & MacGregor, 2010; Rashidi & Piran, 2011; Staehr, 2008; Webb, 2007). Validations of the VLT have been carried out by Read (1988), Beglar and Hunt (1999) and Schmitt et al. (2001). They all indicated that the test is a good measurement of vocabulary knowledge. However, there is one issue which has not been adequately addressed: local item dependence (LID). It is assumed that a response to an item should be independent of a response to another item (Hambleton, Swaminathan, & Roger, 1991; Lee, 2004; Zenisky, Hambleton, & Sireci, 2002). However, the VLT seems to violate this assumption (e.g., Nation, 2007).

Where does the possible violation of item independence come from? It comes from the test format – the matching format, as shown in Figure 1. This format, interacting with test takers' knowledge of words, is responsible for triggering LID to various degrees. The most serious effect of LID is to increase the chance of guessing correctly (Mochizuki, 1998; Shillaw, 1999; Stewart, & White, 2011).

1 business	
2 clock	_____ part of a house
3 horse	_____ animal with four legs
4 pencil	_____ something used for writing
5 shoe	
6 wall	

Figure 1. Example of the VLT Format.

In the validation studies, Beglar and Hunt (1999) and Schmitt et al. (2001) found some dependent items on the basis of Rasch analysis. However, they interpreted that most items were not dependent and caused no particular adverse effects to their following analyses. Then, they shifted items to make parallel forms [see Xing & Fulcher (2007) for criticism]. This paper shall revisit the LID issue and examine it from a non-Rasch approach.

2 Method

2.1 Material

This study used the original VLT (Nation, 1990). Among the five word levels available, four levels were used. One level consisted of six clusters, each of which had six words and three definitions, as shown in Figure 1. The original test was referred to as “6 3 form” (see Appendix 1). On the basis of this material, an experimental matching format was constructed by collapsing a set of three clusters into one large cluster consisting of 18 words and 9 definitions. This collapsed form was referred to as “18 9 form” (see Appendix 2).

2.2 Participants and Procedure

Three groups of EFL Japanese university students took the 18 9 form first and one week later the 6 3 form. Due to absence on either of the two occasions, several students were dropped from the analysis, leaving a total of 114 participants.

3 Results and Discussion

3.1 Descriptive Statistics

Table 1 shows participants' overall performance on the two forms. There are two points to be noted. First, the mean scores for the 6 3 form were larger across all the four word-frequency levels than the mean scores for the 18 9 form. Second, as the word-frequency levels became lower, the mean scores for the two forms became smaller but differences in scores gradually increased. Paired *t*-tests show that all the differences were statistically significant at the 0.01 level ($df = 113$, *t* values = 9.16 for 2K, $r = .65$; 14.90 for 3K, $r = .82$; 16.63 for 5K, $r = .84$; and 16.75 for AWL (Academic Word List), $r = .84$, two-tailed).

Table 1. Means and Standard Deviations of Scores for the Two Forms

Word level	Mean (%)		Differences (%)	SD (%)	
	6 3 form	18 9 form		6 3 form	18 9 form
2K	15.96 (88.67)	14.52 (80.67)	1.44 (8.00)	1.66 (9.22)	2.29 (12.72)
3K	13.63 (75.72)	10.89 (60.50)	2.75 (15.22)	2.46 (13.67)	2.92 (16.22)
5K	10.54 (58.56)	7.27 (40.39)	3.26 (18.13)	2.42 (13.72)	2.07 (11.50)
AWL	9.34 (51.89)	5.72 (31.78)	3.62 (20.11)	2.91 (16.17)	2.97 (16.50)
Sum ^a	49.47 (68.71)	38.40 (53.33)	11.07 (15.38)	7.39 (10.26)	8.49 (11.76)

The maximum score per word level is 18.

^aFigures in this row do not always add up to a sum of four figures in each column due to rounding.

3.2 Item Facility

Item facility (IF) indices were employed to identify LID items. We subtracted IF values of items for the 18 9 form from corresponding IF values of the same items for the 6 3 form. Results showed that the IF differences ranged from -0.009 to $+0.421$, with a mean of 0.154 (see Appendix 3 for details). All IF differences fell above zero except one item,¹ and three items had the same score. With regard to the remaining 68 items, participants performed better on the 6 3 form than on the 18 9 form.

Figure 2 is a graphical representation of the entire IF data. Each vertical line has two IF values at its ends; a triangle indicates an IF value for the 6 3 form and a cross indicates an IF value for the 18 9 form. Its length represents how widely the IF values of the same item are different. Although several items with no or very short length were found here and there, most of the items had various degrees of length. In short, the number of items with some length outnumbers by far the number of items with no or very short length, suggesting strongly that the 6 3 form inflated the scores of most items.

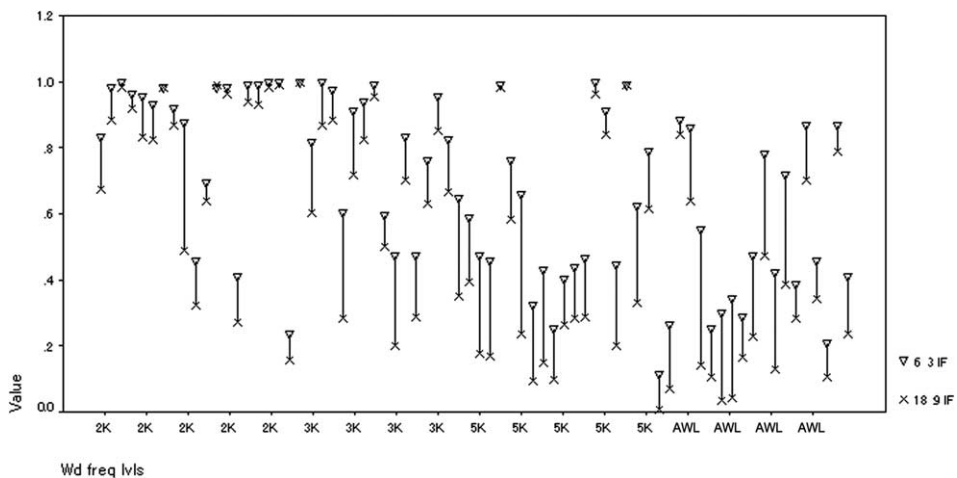


Figure 2. Drop-line Plot of IF Values between the Two Forms.

Table 2. A Frequency Table of Items by Word Levels and IF Differences

	2K	3K	5K	AWL	Total (%)
<i>IF diff ranges</i>					
$0.50 \leq \sim < 0.60$					
$0.40 \leq \sim < 0.50$			1	1	2 (2.78)
$0.30 \leq \sim < 0.40$	1	1		2	4 (5.56)
$0.20 \leq \sim < 0.30$		3	6	5	14 (19.44)
$0.10 \leq \sim < 0.20$	5	9	7	8	29 (38.89)
$0.00 \leq \sim < 0.10$	11	5	4	2	22 (26.39)
$-0.10 \leq \sim < 0.00$	1				1 (4.17)
Total	18	18	18	18	72 (100)

In the hope of sketching out the general features of the data on item difficulty, the frequency counts were tabulated according to both the IF difference range and the word levels in Table 2. For identification of LID items, two cut-off points can be adopted, depending on the level of severity. If an IF difference of 0.10 is adopted on an assumption that measurement error can happen within an IF difference range of less than 0.10, LID items account for 49 (68.06%) of 72 items. They include 6 2K items, 13 3K items, 14 5K items and 16 AWL items. If a lenient criterion of 0.20 is adopted, LID items account for 20 (27.78%) of all the items: one 2K item, four 3K items, seven 5K items and eight AWL items. It should also be noted that the number of the LID items increased as word difficulty increased in the order of word frequency. In other words, it appears that LID tends to become more marked as words become less frequent.

3.3 LID Items

IF indices are a handy but rough measure for the identification of LID items because IF values on the 6 3 form represent the end product of a series of test-taking behaviours within a cluster. For more accurate identification, it is preferable to see whether the participants' responses are consistent or not between the 6 3 and 18 9 forms. Figure 3 illustrates four possible response types on items between the two forms. When participants matched an item correctly to a definition

		6 3 form	
		correct	wrong
18 9 form	correct	1	3
	wrong	2	4

Figure 3. A Taxonomy of Response Types between the Two Forms.

on both the 6 3 and 18 9 forms, their response can be considered as LID-free. Participants would answer right irrespective of other words within a cluster. This response type corresponds to the first cell. On the other hand, if a response was correct on the 6 3 form but wrong on the 18 9 form, it suggests that the participants did not actually know an item and they unduly gained a score on the 6 3 format. This response type is where guessing effects by LID are found. The second cell is the case in point. The opposite case where a response is correct on the 18 9 form but wrong on the 6 3 form is expected to be rare. This corresponds to the third cell. Finally, a wrong response on both forms suggests that the participants did not know the item. This is a response type of the fourth cell. Our chief concern is with response Type 2.

Table 3 shows frequencies of the following three types of data: (1) the number correct on the two forms, (2) an overlap between them corresponding to the first cell in Figure 3 and (3) deviations on the 6 3 and 18 9 forms from an overlap corresponding, respectively, to the second and third cells. Items are displayed in the order of appearance on the 6 3 form. A glance at the table indicates that the frequency for response Type 2 ranged from 0 to 53 (46.49%), indicating that guessing effects of LID varied considerably from one item to another.

This frequency table is also very useful in comprehending how items in a cluster affect one another. As an illustrative example, let us take up the case of “spoil” from the 2K word level, which had the second highest LID index of 46 (40.35%) among all the items. As the table shows, 56 participants were correct on the 18 9 form, and one week later the number correct for the item on the 6 3 form increased to 100. How did such a large increase take place? The other two target items with which “spoil” shared the same cluster were “invite” and “hide”. They had an overlap of 110 and 99, respectively, and the LID indices for them were 2 and 6. In short, these two items were extremely easy when compared with “spoil”. After matching them correctly to definitions, the participants apparently tackled with “spoil”. In the end, it appears that 46 participants ended up gaining a score, while the remaining number of the participants chose one of the three distractors of “blame,” “hit” and “pour” which were relatively easy.

Table 4, summing up the data above, displays descriptive statistics of the guessing rates according to word frequency level. Means ranged from 9.89% for the 2K word level to 24.61% for the AWL level with a mean of 18.84%. The guessing rates increased as word frequency decreased, suggesting that the less frequent the word levels may be, the more dependent items tend to be. It is interesting to note that the guessing rates are higher than the corresponding IF differences shown in Table 1. This evidence, adequately supporting Stewart & White (2011), showed that successful guessing may be more prevalent than their theoretical estimation of guessing even when items are less frequent.

4 Conclusion

These findings have shown that the original matching format of the VLT has a mean guessing rate of 18.84%. This rate may vary according to test takers' lexical proficiency and the number of word levels involved. It seems reasonable to

Table 3. A Frequency Table of Response Types

Word ID	Item	Number correct		Response		
		6 3	18 9	Type 1 Overlap	Type 2 6 3 (%)	Type 3 18 9 (%)
2K01	original	95	77	76	19 (16.67)	1 (0.88)
2K02	private	112	101	100	12 (10.53)	1 (0.88)
2K03	total	114	112	112	2 (1.75)	0 (0.00)
2K04	manufacture	110	105	104	6 (5.26)	1 (0.88)
2K05	elect	109	95	91	18 (15.79)	4 (3.51)
2K06	melt	106	94	93	13 (11.40)	1 (0.88)
2K07	invite	112	112	110	2 (1.75)	2 (1.75)
2K08	hide	105	99	99	6 (5.26)	0 (0.00)
2K09	spoil	100	56	54	46 (40.35)	2 (1.75)
2K10	roar	52	37	27	25 (21.93)	10 (8.77)
2K11	debt	79	73	63	16 (14.04)	10 (8.77)
2K12	pride	112	113	111	1 (0.88)	2 (1.75)
2K13	temperature	112	110	109	3 (2.63)	1 (0.88)
2K14	flesh	47	31	30	17 (14.91)	1 (0.88)
2K15	salary	113	107	107	6 (5.26)	0 (0.00)
2K16	sport	113	106	105	8 (7.02)	1 (0.88)
2K17	victory	114	112	112	2 (1.75)	0 (0.00)
2K18	birth	114	113	113	1 (0.88)	0 (0.00)
3K01	herd	27	18	13	14 (12.28)	5 (4.39)
3K02	angel	114	114	114	0 (0.00)	0 (0.00)
3K03	administration	93	69	59	34 (29.83)	10 (8.77)
3K04	bench	114	99	99	15 (13.16)	0 (0.00)
3K05	charity	111	101	101	10 (8.77)	0 (0.00)
3K06	province	69	32	29	40 (35.09)	3 (2.63)
3K07	echo	104	82	78	26 (22.81)	4 (3.51)
3K08	darling	107	94	93	14 (12.28)	1 (0.88)
3K09	slice	113	109	109	4 (3.51)	0 (0.00)
3K10	palm	68	57	46	22 (19.30)	11 (9.65)
3K11	scheme	54	23	22	32 (28.07)	1 (0.88)
3K12	thrill	95	80	78	17 (14.91)	2 (1.75)
3K13	illustrate	54	33	23	31 (27.19)	10 (8.77)
3K14	encounter	87	72	66	21 (18.42)	6 (5.26)
3K15	toss	109	97	96	13 (11.40)	1 (0.88)
3K16	annual	94	76	75	19 (16.67)	1 (0.88)
3K17	savage	74	40	33	41 (35.97)	7 (6.14)
3K18	definite	67	45	37	30 (26.32)	8 (7.02)
5K01	phase	54	20	14	40 (35.09)	6 (5.26)
5K02	mess	52	19	16	36 (31.58)	3 (2.63)
5K03	apron	113	112	111	2 (1.75)	1 (0.88)
5K04	trumpet	87	66	61	26 (22.81)	5 (4.39)
5K05	stool	75	27	22	53 (46.49)	5 (4.39)
5K06	sermon	37	11	9	28 (24.56)	2 (1.75)
5K07	compliment	49	17	13	36 (31.58)	4 (3.51)

Table 3 (Continued)

Word ID	Item	Number correct		Response		
				Type 1	Type 2	Type 3
		6 3	18 9	Overlap	6 3 (%)	18 9 (%)
5K08	apparatus	29	11	6	23 (20.18)	5 (4.39)
5K09	revenue	46	30	15	31 (27.19)	15 (13.16)
5K10	precede	50	32	21	29 (25.44)	11 (9.65)
5K11	collapse	53	33	26	27 (23.68)	7 (6.14)
5K12	skip	114	110	110	4 (3.51)	0 (0.00)
5K13	blend	104	96	91	13 (11.40)	5 (4.39)
5K14	devise	51	23	16	35 (30.70)	7 (6.14)
5K15	hug	113	113	113	0 (0.00)	0 (0.00)
5K16	gloomy	71	38	34	37 (32.46)	4 (3.51)
5K17	fragrant	90	70	68	22 (19.30)	2 (1.75)
5K18	wholesome	13	1	0	13 (11.40)	1 (0.88)
AWL01	affluence	30	8	4	26 (22.81)	4 (3.51)
AWL02	episode	101	96	89	12 (10.53)	7 (6.14)
AWL03	innovation	96	73	68	28 (24.56)	5 (4.39)
AWL04	deficiency	63	16	15	48 (42.11)	1 (0.88)
AWL05	prestige	29	12	5	24 (21.05)	7 (6.14)
AWL06	oscillation	34	4	2	32 (28.07)	2 (1.75)
AWL07	configuration	39	5	2	37 (32.46)	3 (2.63)
AWL08	discourse	33	19	9	24 (21.05)	10 (8.77)
AWL09	hypothesis	54	26	20	34 (29.83)	6 (5.26)
AWL10	supplement	89	54	52	37 (32.46)	2 (1.75)
AWL11	expel	48	15	11	37 (32.46)	4 (3.51)
AWL12	deprive	82	44	40	42 (36.84)	4 (3.51)
AWL13	restrict	44	32	22	22 (19.30)	10 (8.77)
AWL14	transform	99	80	76	23 (20.18)	4 (3.51)
AWL15	assume	52	39	36	16 (14.04)	3 (2.63)
AWL16	indigenous	24	12	2	22 (19.30)	10 (8.77)
AWL17	minimum	99	90	86	13 (11.40)	4 (3.51)
AWL18	anonymous	47	27	19	28 (24.56)	8 (7.02)

conclude, however, that this study indicates that items in clusters tend to be more dependent than the previous Rasch-based studies suggested. This may lead us to wonder if they have met the item response theory (IRT) requirement of item independence between responses to any pair of items.

Table 4. Descriptive Statistics of Guessing Rates by Word-frequency Level

	Mean (%)	SD (%)
2K	11.28 (9.89)	11.35 (9.96)
3K	21.28 (18.67)	11.70 (10.26)
5K	25.28 (22.17)	14.26 (12.51)
AWL	28.06 (24.61)	9.85 (8.64)
Four levels	21.47 (18.84)	13.29 (11.65)

Note

1. The item is “pride” from the 2K word frequency level. This was the only item with a minus IF difference from among a total of 72 items. Its value (−0.009) was extremely small and equivalent to one participant. This irregularity may be caused by a participant’s carelessness. It can be dismissed as negligible.

References

- Akbarian, I. (2010). The relationship between vocabulary size and depth for ESP/EAP learners. *System*, 38, 391–401. doi:10.1016/j.system.2010.06.013
- Beglar, D., & Hunt, A. (1999). Revising and validating the 2000 word level and university word level. *Language Testing*, 16, 131–162. doi:10.1177/026553229901600202
- Hambleton, R. K., Swaminathan, H., & Roger, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). London, UK: Sage.
- Lee, Y.-W. (2004). Examining passage-related local item dependence (LID) and measurement construct using Q_3 statistics in an EFL reading comprehension test. *Language Testing*, 21, 74–100. doi:10.1191/0265532204lt260oa
- Li, L., & MacGregor, L. J. (2010). Investigating the receptive vocabulary size of university-level Chinese learners of English: How suitable is the Vocabulary Levels Test? *Language and Education*, 24, 239–249. doi:10.1080/09500781003642478
- Mochizuki, M. (1998). Nihonjin Eigo Gakushusha no tameno goi saizu tesuto [A vocabulary size test for Japanese learners of English]. *The IRLT (Institute for Research in Language Teaching) Bulletin*, 12, 27–53.
- Nation, I.S.P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle & Heinle.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.
- Nation, I.S.P. (2007). Fundamental issues in modelling and assessing vocabulary knowledge. In H. Daller, J. Milton & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 35–43). Cambridge, UK: Cambridge University Press.
- Rashidi, N., & Piran, M. (2011). The effect of extensive and intensive reading on Iranian EFL learners’ vocabulary size and depth. *Journal of Language Teaching and Research*, 2, 471–482. doi:10.4304/jltr.2.2.471-482
- Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal*, 19(2), 12–25. doi:10.1177/003368828801900202
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge, UK: Cambridge University Press.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18, 55–88. doi:10.1177/026553220101800103

- Shillaw, J. (1999). *The application of the Rasch model to yes/no Vocabulary Tests* (Unpublished Ph.D. thesis). University of Wales Swansea, Swansea, Wales.
- Staehr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36, 139–152. doi:10.1080/09571730802389975
- Stewart, J., & White, D. A. (2011). Estimating guessing effects on the Vocabulary Levels Test for differing degrees of word knowledge. *TESOL Quarterly*, 45, 370–380.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28, 46–65. doi:10.1093/applin/aml048
- Xing, P., & Fulcher, G. (2007). Reliability assessment for two versions of Vocabulary Levels Tests. *System*, 35, 182–191. doi:10.1016/j.system.2006.12.009
- Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2002). Identification and evaluation of local item dependencies in the medical college admissions test. *Journal of Educational Measurement*, 39, 291–309. doi:10.1111/j.1745-3984.2002.tb01144.x

Appendix 1

6 3 Form (Sample)

3K word frequency level

1 administration

2 angel _____ 群れ

3 frost _____ 天使

4 herd _____ 運営

5 fort

6 pond

1 bench

2 charity _____ 長椅子

3 mate _____ 慈善

4 jar _____ 地方

5 mirror

6 province

1 coach

2 darling _____ こだま

3 echo _____ 最愛の人

4 interior _____ 薄切り 1 枚

5 opera

6 slice

1 marble

2 palm _____ 手のひら

3 ridge _____ 計画

4 scheme _____ わくわくする感じ

5 statue

6 thrill

1 discharge

2 encounter _____ 例証する

3 illustrate _____ 出合う

4 prevail _____ を投げる

5 restore

6 toss

1 annual

2 previous _____ 年 1 回の

3 mental _____ 野生の

4 concealed _____ はっきりと確定した

5 definite

6 savage

Appendix 2

18 9 Form (Sample)

3K word frequency level

(A)

- 1 administration
- 2 angel _____ 薄切り 1 枚
- 3 bench _____ 運営
- 4 charity _____ こだま
- 5 coach _____ 最愛の人
- 6 darling _____ 慈善
- 7 echo _____ 地方
- 8 fort _____ 天使
- 9 frost _____ 長椅子
- 10 herd _____ 群れ
- 11 interior
- 12 jar
- 13 mate
- 14 mirror
- 15 opera
- 16 pond
- 17 province
- 18 slice

(B)

- 1 annual
- 2 concealed _____ 計画
- 3 definite _____ 出合う
- 4 discharge _____ 手のひら
- 5 encounter _____ 年 1 回の
- 6 illustrate _____ はっきりと確定した
- 7 marble _____ 野生の
- 8 mental _____ 例証する
- 9 palm _____ わくわくする感じ
- 10 prevail _____ を投げる
- 11 previous
- 12 restore
- 13 ridge
- 14 savage
- 15 scheme
- 16 statue
- 17 thrill
- 18 toss

Appendix 3

All the 72 Items in the Increasing Order of IF Differences

No	Item ID*	Items	6 3 IF	18 9 IF	Differences
1	2K12	pride	0.982	0.991	-0.009
2	2K07	invite	0.982	0.982	0.000
3	3K02	angel	1.000	1.000	0.000
4	5K15	hug	0.991	0.991	0.000
5	2K18	birth	1.000	0.991	0.009
6	5K03	apron	0.991	0.982	0.009
7	2K13	temperature	0.982	0.965	0.018
8	2K03	total	1.000	0.982	0.018
9	2K17	victory	1.000	0.982	0.018
10	3K09	slice	0.991	0.956	0.035
11	5K12	skip	1.000	0.965	0.035
12	2K04	manufacture	0.965	0.921	0.044
13	AWL02	episode	0.886	0.842	0.044
14	2K08	hide	0.921	0.868	0.053
15	2K15	salary	0.991	0.939	0.053
16	2K11	debt	0.693	0.640	0.053
17	2K16	sport	0.991	0.930	0.061
18	5K13	blend	0.912	0.842	0.070
19	3K02	herd	0.237	0.158	0.079
20	AWL17	minimum	0.868	0.789	0.079
21	3K05	charity	0.974	0.886	0.088
22	2K02	private	0.982	0.886	0.096
23	3K10	palm	0.596	0.500	0.096
24	AWL13	restrict	0.386	0.283	0.103
25	2K06	melt	0.930	0.825	0.105
26	3K15	toss	0.956	0.851	0.105
27	5K18	wholesome	0.114	0.009	0.105
28	AWL16	indigenous	0.211	0.105	0.105
29	AWL15	assume	0.456	0.342	0.114
30	3K08	darling	0.939	0.825	0.114
31	2K05	elect	0.956	0.833	0.123
32	AWL08	discourse	0.289	0.167	0.123
33	2K10	roar	0.456	0.325	0.132
34	3K04	bench	1.000	0.868	0.132
35	3K12	thrill	0.833	0.702	0.132
36	3K14	encounter	0.763	0.632	0.132
37	5K09	revenue	0.404	0.265	0.138
38	2K14	flesh	0.412	0.272	0.140
39	AWL05	prestige	0.254	0.105	0.149
40	5K10	precede	0.439	0.283	0.155
41	5K08	apparatus	0.254	0.097	0.157
42	2K01	original	0.833	0.675	0.158
43	3K16	annual	0.825	0.667	0.158

Appendix 3 (Continued)

No	Item ID*	Items	6 3 IF	18 9 IF	Differences
44	AWL14	transform	0.868	0.702	0.167
45	5K11	collapse	0.465	0.289	0.175
46	5K17	fragrant	0.789	0.614	0.175
47	AWL18	anonymous	0.412	0.237	0.175
48	5K04	trumpet	0.763	0.584	0.179
49	3K13	illustrate	0.474	0.289	0.184
50	3K07	echo	0.912	0.719	0.193
51	AWL01	affluence	0.263	0.070	0.193
52	3K18	definite	0.588	0.395	0.193
53	3K03	administration	0.816	0.605	0.211
54	AWL03	innovation	0.860	0.640	0.219
55	5K06	sermon	0.325	0.096	0.228
56	AWL09	hypothesis	0.474	0.228	0.246
57	5K14	devise	0.447	0.202	0.246
58	AWL06	oscillation	0.298	0.035	0.263
59	3K11	scheme	0.474	0.202	0.272
60	5K07	compliment	0.430	0.149	0.281
61	5K02	mess	0.456	0.168	0.288
62	AWL11	expel	0.421	0.132	0.289
63	5K16	gloomy	0.623	0.333	0.289
64	5K01	phase	0.474	0.177	0.297
65	3K17	savage	0.649	0.351	0.298
66	AWL07	configuration	0.342	0.044	0.298
67	AWL10	supplement	0.781	0.474	0.307
68	3K06	province	0.605	0.283	0.322
69	AWL12	deprive	0.719	0.386	0.333
70	2K09	spoil	0.877	0.491	0.386
71	AWL04	deficiency	0.553	0.140	0.412
72	5K05	stool	0.658	0.237	0.421

Note: Item ID* consists of a combination of a word frequency level and a two-digit number (from 01 to 18) which corresponds to the appearing order of a definition within a word-frequency level in the original VLT.