INSTITUT FÜR TESTFORSCHUNG
UND TESTENTWICKLUNG e.V.

ITT Technical Reports on Language Testing 4

Assigning Frequency Bands to the Receptive Vocabulary Size Test According to the Total Score of the Test Taker

Erwin Tschirner[1]

[1] Leipzig University, Germany

Author Note

Erwin Tschirner https://orcid.org/0000-0002-5915-5344

Correspondence concerning this report should be addressed to Erwin Tschirner, Herder-Institut, Universität Leipzig, Beethovenstraße 15, 04107 Leipzig. email: tschirner@uni-leipzig.de

Bibliographische Informationen der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im
Internet über http://dnb.ddb.de abrufbar.

Die "ITT Technical Reports on Language Testing" sind eine Reihe des Instituts für Testforschung und Testentwicklung e. V. (ITT), in der Technische Reports zu unterschiedlichen Sprachtestevaluationen veröffentlicht werden.

Institut für Testforschung und Testentwicklung e.V. Leipzig
c/o Herder-Institut
Universität Leipzig
Beethovenstraße 15
04107 Leipzig
www.itt-leipzig.de

Herausgeberschaft:
Olaf Bärenfänger, Universität Leipzig
Jupp Möhring, Technische Universität Dresden
Erwin Tschirner, Universität Leipzig

Redaktion:
Lisa Lenort, Elisabeth Muntschick, Antonia Kurz

URN des Bandes: https://nbn-resolving.org/urn:nbn:de:bsz:15-qucosa2-936675
URN der Reihe: https://nbn-resolving.org/urn:nbn:de:bsz:15-qucosa2-902765

ISSN **2942-741X**

Table of Contents

List of Tables

List of Figures

Methods and Analysis

The Receptive Vocabulary Size Test (R-VST) measures how many of the most frequent 5,000 words of a language are known by a test taker (Institute for Test Research and Test Development, n.d.; Tschirner, 2021a; Möhring, 2022). It consists of five bands of 1,000 words each, involving 30 multiple-choice tasks. It is commonly scored according to how many of the words of each band are identified correctly, commonly 80% correct (Hacking, Rubio, & Tschirner 2019: 32). This is the percentage used in the present study.

The highest of the five bands that is at least 80 percent correct is considered the vocabulary size of the test taker. Each band has a maximum score of 30; 80 percent correct, therefore, amounts to a score of 24. These cut scores, however, increase the possibility that the result may be unduly influenced by chance. Imagine, e.g., a test taker who has a score of 23 on all five bands, whose final rating would, therefore, be 'less than 1000 words'. Moreover, reducing the results of each band to a yes/no decision, does not utilize the richness of the original scale, i.e., 150 individual items. Furthermore, an estimation of someone’s vocabulary size based on total score may be done more rapidly than having to score each band separately and combining the results. The aim of this analysis, therefore, was to use the total score rather than the 80 percent correct band approach to predict vocabulary size.

To examine the relationship between total scores and vocabulary size, all receptive VSTs administered to a total of 2,333 test takers who had been part of vocabulary studies under controlled conditions were analyzed. These included 651 test takers at a university in Germany (see, e.g., Möhring 2022), 92 students at three high schools in Hungary, Italy, and the Netherlands (see, e.g., Abitzsch et al., 2019), and 1,590 examinees at five US colleges and universities (see, e.g., Tschirner, 2019; Tschirner, 2021b; Tschirner et al., 2019). Table 1 displays the languages involved.

Table 1

Languages Involved in the Analysis

	Number of Participants	Percent
Arabic	83	3.6
Chinese	192	8.2
French	240	10.3
German	972	41.7
Italian	144	6.2
Japanese	67	2.9
Russian	125	5.4
Spanish	510	21.9
Total	2,333	100.0

The majority of the German test takers studied in Germany or were enrolled in European high schools and had a variety of language backgrounds. All other test takers were either monolingual or bilingual English speakers. For many of these test takers, reading and listening proficiency data were also available and were used to provide evidence of convergent validity. The following test takers were removed from the analysis: (1) Test takers who speeded through the vocabulary test. Speeding was defined as spending less than six minutes on the whole test. It was assumed that these test takers simply went through the motion of taking the test without paying much, if any, attention to the items. (2) Test takers who had less than seven points on every single band. Some test takers never scrolled down to see all items of a band. Each screen presents six items. If they moved from band to band on account of the first screen only, they did not complete a single band, i.e., there is no band score that may be assigned to them. This reduced the number of examinees to 2,108.

To establish total score cut scores for bands, descriptive statistics for all test takers who were assigned a particular band according to the 80 percent correct criterion were calculated.[1] Table 2 shows the bands; the number of test takers who attained the band; the minimum, maximum, median, and mean total score; the standard error of the mean; the standard deviation; and the mean minus 1.5 times the standard deviation (SD). The mean of the test takers at each band minus 1.5 times the SD was selected to help establish cut scores. This measure includes the vast majority of the test takers who crossed the threshold of a particular band (approx. 93%).

Table 2

Descriptive Total Score Statistics of Each Band

	N	Min	Max	Median	Mean	S.E.	SD	M-1.5*SD
Below Band 1	964	0	104	41	42.83	0.72	22.34	9
Band 1	207	27	114	80	79.12	1.19	17.12	53
Band 2	171	49	125	97	94.03	1.30	16.91	69
Band 3	119	83	133	112	109.4	1.18	12.86	90
Band 4	173	50	142	126	124.2	0.87	11.46	107
Band 5	474	40	150	143	140.2	0.55	12.02	122
Total	2108							

Table 3 displays suggested correspondences between total scores and bands based on the descriptive total score statistics of each band. It shows the minimum score needed for a given band,

[1] While languages may differ in the vocabulary sizes required for particular text types and there may be differences in difficulty levels across forms or language versions of the same test, we were interested in how the R-VST performed across languages. This approach is supported by the fact that the R-VST is based on corpora designed and frequency lists analyzed the same way in the Routledge Frequency Dictionaries book series as well as on the same design and the same quality assurance processes of test construction and item validation as implemented by the Institute of Test Research and Test Development.

assuming that the test stops after the band in question; the mean minus 1.5 times the standard deviation (SD) from Table 2; and the suggested total score required to be assigned a particular band.

Table 3

Suggested Bands Based on Total Scores for the Receptive VST

Band	Minimum Score Needed	M-1.5*SD	Suggested Total Score
Below Band 1			0-53
Band 1	24	53	54-69
Band 2	48	69	70-89
Band 3	72	90	90-105
Band 4	96	107	106-119
Band 5	120	122	120-150

The lower boundaries of the suggested total scores for each band were minimally changed from the mean minus 1.5 SD to result in an even number. While bands 1 and 2 were raised by one point, band 4 was lowered by 1. They are based on the assumption that test takers know additional words at higher bands but increasingly fewer additional words at each higher band. Suggested total scores for bands 1 to 4, therefore, are 30, 22, 18, and 10 points higher, respectively, than the minimum score required, taking into account the fact that increasingly fewer words are known at higher bands. For band 5, the minimum score needed was used.

To examine the relationship between total scores, bands based on 80 percent correct and suggested new bands, bivariate correlations between these three measures were run. Table 4 shows both *Pearson* and *Spearman* correlations. All correlations were significant at the 0.01 level (2-tailed).

Table 4

Correlations Between Total Scores and Bands Assigned According to Total Score vs. According to the 80% Correct Criterion

Correlation Between Total Score and	N	*Pearson's R*	*Spearman's Rho*
Band Assigned According to Total Score	2108	0.975	0.968
Band Assigned According to 80% Correct	2108	0.904	0.903

Note: $p < 0.01$.

Table 4 demonstrates that the correlation between total score and bands assigned according to Table 3 is higher than assignments according to the standard formula focusing on individual bands, because it uses more of the more fine-grained information contained in total scores. It is our contention that this may estimate the vocabulary size of a test taker slightly better than yes/no decisions based on the 80 percent correct cut-off.

To look for evidence of convergent validity, correlations between two measures of receptive proficiency, i.e., listening and reading proficiency, and the two ways of defining receptive vocabulary bands were examined. Proficiency was measured using the results of official ACTFL Reading (RPT) and Listening Proficiency Tests (LPT). There were 1130 RPT and 572 LPT that had receptive vocabulary scores. Tables 5 and 6 demonstrate how many test takers were at a particular proficiency level, while figures 1 and 2 show boxplots of the R-VST results by ACTFL proficiency level.[2]

Table 5

Number of Test Takers per Reading Proficiency Level

ACTFL Level	NL	NM	NH	IL	IM	IH	AL	AM	AH	S	Total
N	36	34	70	157	187	125	237	143	73	68	1130

[2] 1 = NL (Novice Low); 2 = NM (Novice Mid); 3 = NH (Novice High); 4 = IL (Intermediate Low); 5 = IM (Intermediate Mid); 6 = IH (Intermediate High); 7 = AL (Advanced Low); 8 = AM (Advanced Mid); 9 = AH (Advanced High); 10 = S (Superior). NH and IL are roughly equivalent to A1; IM to A2; IH and AL to B1; AM to B2; AH and S to C1 (ACTFL, n.d.).

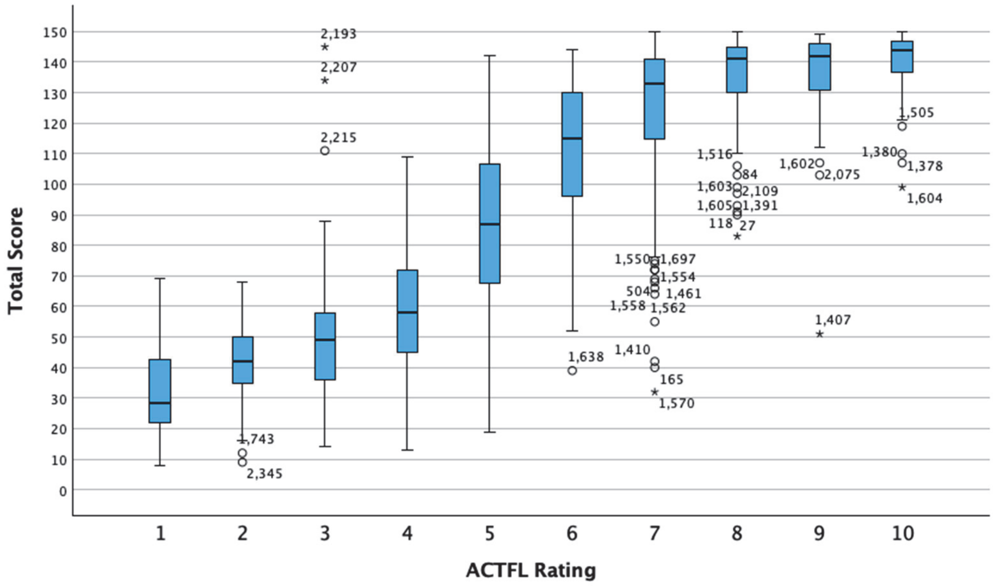
Table 6

Number of Test Takers per Listening Proficiency Level

ACTFL Level	NL	NM	NH	IL	IM	IH	AL	AM	AH	S	Total
N	7	15	28	38	82	91	208	127	20	10	626

Figure 1

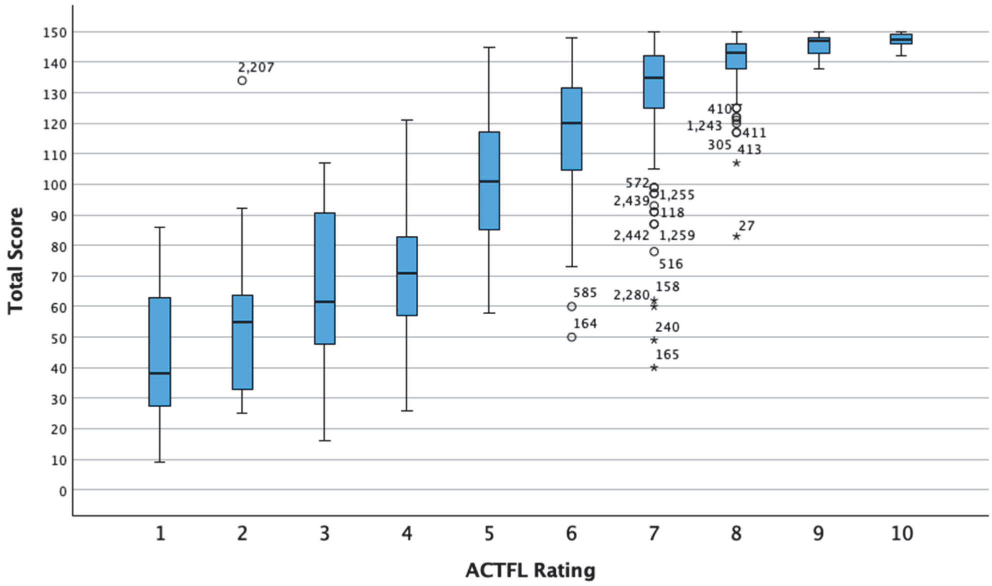
Boxplots of R-VST Scores by ACTFL Reading Proficiency Rating



Note: 1 = NL; 2 = NM; 3 = NH; 4 = IL; 5 = IM; 6 = IH; 7 = AL; 8 = AM; 9 = AH; 10 = S.

Figure 2

Boxplots of R-VST Scores by ACTFL Listening Proficiency Rating



Note: 1 = NL; 2 = NM; 3 = NH; 4 = IL; 5 = IM; 6 = IH; 7 = AL; 8 = AM; 9 = AH; 10 = S.

Figures 1 and 2 exhibit a relatively large number of outliers for some proficiency ratings. They were removed from the following calculations. This reduced the number of RPT to 1,099 and LPT to 607.

To examine the differences between the two scoring methods more closely, correlations were calculated between listening and reading proficiency levels and the two methods of calculating vocabulary size. Table 7 shows both *Pearson* and *Spearman* correlations between the two receptive skills and the two ways of calculating vocabulary bands. All correlations were significant at the 0.01 level (2-tailed).

Table 7

Correlations Between Bands Assigned According to Total Score and 80 Percent Correct per Band and Listening and Reading Proficiency Levels

		Total Score Bands		80% Bands	
	N	*Pearson's R*	*Spearman's Rho*	*Pearson's R*	*Spearman's Rho*
Listening	607	0.798	0.761	0.783	0.759
Reading	1099	0.824	0.844	0.798	0.818

Note: p < .0.01.

Table 7 demonstrates that correlations between listening and reading proficiency and bands are slightly better for total score bands for reading but essentially the same for bands derived from total scores and bands defined by 80 percent correct for listening. This provides convergent validity evidence to suggest that the classification of bands according to both models would be justifiable. Because the model based on total scores uses more of the information provided by the test and may, therefore, be less influenced by chance than yes/no decisions based on individual bands, we suggest that it may be a better estimate of the vocabulary size of test takers. In addition, the strong correlations between total score bands and listening and reading proficiency provide evidence that the total score cut scores are valid.

To further examine the relationship between vocabulary and reading, tables 8 and 9 show crosstabulations of vocabulary bands and ACTFL reading proficiency levels for the two methods of assigning bands. Cells that include at least 10 percent of the total number of the respective band are highlighted.

Table 8

Crosstabulation of ACTFL Reading Proficiency Levels and Vocabulary Bands Assigned According to Total Score

Band	NL	NM	NH	IL	IM	IH	AL	AM	AH	S	Total
0	32	25	42	70	26	1	0	0	0	0	196
1	4	7	20	39	24	5	0	0	0	0	99
2	0	0	5	31	48	22	11	0	0	0	117
3	0	0	0	16	40	19	19	0	0	0	94
4	0	0	0	1	29	24	28	9	8	0	99
5	0	0	0	0	20	53	169	126	62	64	494
Total	36	32	67	157	187	124	227	135	70	64	1099

Note: Cells that include at least 10% of the total number of the band are highlighted.

Table 9

Crosstabulation of ACTFL Reading Proficiency Levels and Vocabulary Bands Assigned According to 80% Correct of Individual Bands

Band	NL	NM	NH	IL	IM	IH	AL	AM	AH	S	Total
0	36	31	60	111	74	8	5	0	0	0	325
1	0	1	4	33	32	19	17	0	0	0	106
2	0	0	3	11	41	22	25	2	0	0	104
3	0	0	0	2	13	18	19	12	5	2	71
4	0	0	0	0	15	33	37	17	15	12	129
5	0	0	0	0	12	24	124	104	50	50	364
Total	36	32	67	157	187	124	227	135	70	64	1099

Note: Cells that include at least 10% of the total number of the band are highlighted.

Tables 8 and 9 demonstrate that a scoring system that focuses on yes/no decisions per band (Table 9) appears to make it more difficult to receive a higher vocabulary score. 325 test takers, e.g., received a score of below band 1 in Table 9, whereas only 196 did in Table 8. Table 9 also shows a surprisingly large number of IM readers with a vocabulary score of less than 1,000 words, while exhibiting a large number of advanced readers with relatively low vocabulary scores of 1,000 and 2,000 words. In general, however, both tables show a clear association of higher vocabulary scores with higher proficiency levels.

Tables 10 and 11 show crosstabulations of vocabulary bands and ACTFL listening proficiency levels for the two methods of assigning bands. Because of the lower overall numbers, only the cells with the highest numbers are highlighted.

Table 10

Crosstabulation of ACTFL Listening Proficiency Levels and Vocabulary Bands Assigned According to Total Score

Band	NL	NM	NH	IL	IM	IH	AL	AM	AH	S	Total
0	4	7	11	8	0	0	0	0	0	0	30
1	2	5	5	9	8	0	0	0	0	0	29
2	1	1	4	15	18	7	1	0	0	0	47
3	0	1	6	3	21	15	6	0	0	0	52
4	0	0	2	2	19	21	25	0	0	0	69
5	0	0	0	1	16	46	166	121	20	10	380
Total	7	14	28	38	82	89	198	121	20	10	607

Note: Cells with the highest numbers are highlighted.

Table 11

Crosstabulation of ACTFL Listening Proficiency Levels and Vocabulary Bands Assigned According to 80% Correct of Individual Bands

Band	NL	NM	NH	IL	IM	IH	AL	AM	AH	S	Total
0	6	10	19	24	13	4	0	0	0	0	76
1	0	3	6	6	11	7	3	0	0	0	36
2	1	1	1	5	23	19	11	0	0	0	61
3	0	0	2	0	16	8	15	2	0	0	43
4	0	0	0	3	10	26	39	11	1	0	90
5	0	0	0	0	9	25	130	108	19	10	301
Total	7	14	28	38	82	89	198	121	20	10	607

Note: Cells with the highest numbers are highlighted.

Tables 10 and 11 again demonstrate that a scoring system that focuses on yes/no decisions per band (Table 11) makes it more difficult to receive a higher vocabulary score. 76 test takers, e.g., received a score of *below band 1* in Table 11, whereas only 30 did in Table 10. Table 11 also shows a relatively large number of IM listeners with a vocabulary score of less than 1,000 words. In general, however, both tables show a clear association of higher vocabulary scores with higher proficiency levels.

On the basis of the preceding discussion, Table 12 presents the suggested total score cut scores for frequency bands of the most frequent 5,000 words of a language as measured by the ITT R-VST. Note that these cut scores are the same as in Table 3.

Table 12

Total Score Cut Scores for Frequency Bands

Total Score	0-53	54-69	70-89	90-105	106-119	120-150
Bands	Below 1000	1000	2000	3000	4000	5000

Conclusion

The present analysis based on 2,108 examinees taking the Receptive Vocabulary Size Test (R-VST) provides evidence for the argument that frequency band definitions based on the total score of examinees appear to be as valid and reliable as the established model of rating each band separately and may be more useful because they can be assigned rapidly and they use more of the information provided by the test. In addition, the present study provides solid validity evidence of the cut scores established by the analysis.

There is one caveat, however. Assigning bands according to total score using the formula above assumes test takers having completed all five bands. In case of doubt, it is suggested to calculate bands using both methods of assigning bands and to allocate the higher band.

References

Abitzsch, D., van der Knaap, E., Abbate, R., Dawidowicz, M., Feld-Knapp, I., Hoffmann, S., Perge, G., & Schramm, K. *Freies Lesen im LEELU-Projekt – Vom Forschungsstand zur Unterrichtshandreichung*. https://leelu.eu/ueber-leelu/publikationen/

ACTFL. (n.d.). *Assigning CEFR Ratings to ACTFL Assessments*. https://www.actfl.org/assessments/assigning-cefr-ratings-to-actfl-assessments

Hacking, J. F., Rubio, F., & Tschirner, E. (2019). Vocabulary size, reading proficiency and curricular design: The case of college Chinese, Russian and Spanish. *Foreign Language Proficiency in Higher Education*, *37*, 25–44. https://doi.org/10.1007/978-3-030-01006-5_3

Institute for Test Research and Test Development. (n.d.). *Vocabulary Tests*. https://itt-leipzig.de/about-the-vocabulary-tests-2/?lang=en

Möhring, J. (2022). Sprachliche Studierfähigkeit im Spiegel produktiver und rezeptiver Wortschatzkompetenz. *Informationen Deutsch Als Fremdsprache*, *49*(4), 384–410. https://doi.org/10.1515/infodaf-2022-0058

Tschirner, E. (2019). Der rezeptive Wortschatzbedarf im Deutschen als Fremdsprache. In E. Peyer, T. Studer, & I. Thonhauser (Eds.), *IDT 2017: Band 1: Hauptvorträge* (pp. 98–111). Erich Schmidt Verlag. https://www.esv-elibrary.de/book/10.37307/b.978-3-503-18161-2/00016/download

Tschirner, E. (2021a). *Examining the validity and reliability of the ITT Vocabulary Size Tests*. *Research Papers in Assessment: Vol. 3*. Universitätsbibliothek Leipzig.

Tschirner, E. (2021b). Extensives Lesen mit Leichten Lektüren: Eine Pilotstudie zum Erwerb von Lesekompetenz bei DaF-Anfänger* innen. In O. C. Díaz Pérez, E. Tschirner, & K. Wisniewski (Eds.), *Deutsch als Fremd- und Zweitsprache: Band 16. Mexikanisch-deutsche Perspektiven auf Deutsch als Fremdsprache: Beiträge zum zehnjährigen*

Bestehen des binationalen Masterstudiengangs Leipzig - Guadalajara (pp. 238–254). Stauffenburg Verlag.

Tschirner, E., Hacking, J. F., & Rubio, F. (2019). Reading proficiency and vocabulary size: An empirical investigation. In: Ecke, P., & Rott, S. (Eds.), Understanding Vocabulary Learning and Teaching: Implications for Language Program Development (pp. 58-77). Cengage Learning.