A CLUSTER ANALYSIS APPROACH TO ASSESSING VOCABULARY AND TARGETING INTERVENTION

Aleka Akoyunoglou Blackwell

Middle Tennessee State University
aleka.blackwell@mtsu.edu

ABSTRACT

This study illustrates how cluster analysis can be applied to vocabulary assessment to classify students into groups with similar profiles of vocabulary knowledge so that vocabulary instruction can be designed and targeted more precisely, especially in light of the multi-dimensional nature of the vocabulary knowledge construct. We performed a cluster analysis with data from 229 native English speaking college students who completed a measure of vocabulary breadth/size (the *PPVT-III*) and a measure of vocabulary depth/quality (the *Word Associates Test*). The cluster analysis revealed 6 distinct clusters of students classified based on their relative performance on the two tests. This cluster solution was supported by significant differences among the clusters with respect to the students' scores on a third measure of both vocabulary breadth and depth (the *Verb Subordinates Test*), a variable external to the clustering process. The practical implications of such a classification of students for vocabulary instruction are discussed.

Keywords: *vocabulary assessment, cluster analysis*

Acknowledgements: The author thanks the following students for their assistance with data collection: Melissa Brock, Seokhee Cho, Dan Davis, Donna Shoults, Stefanie Price, and Deana Wildish.

1. INTRODUCTION

Researchers have long described vocabulary knowledge as a multi-dimensional construct. For example, the model proposed by Anderson and Freebody [1] distinguishes between two dimensions: *breadth* and *depth*. The breadth dimension refers to "the number of words for which the person knows at least some of the significant aspects of meaning" (93). The depth dimension represents the "sufficiently deep understanding of a word [when] it conveys to [someone] all of the distinctions that would be understood by an ordinary adult under normal circumstances" (93). By extension, when evaluating students' vocabulary knowledge, it seems essential to determine whether a student's acquired lexicon reflects both an age appropriate vocabulary size as well as a sufficiently deep knowledge of the acquired words.

With this issue in mind, we set out to investigate the variety of vocabulary knowledge profiles among native speakers of English with regard to both their vocabulary size as well as depth of their vocabulary knowledge. We then employed cluster analysis to find homogeneous groups based on the individual vocabulary knowledge profiles. The ultimate goal of such an exercise would be to design vocabulary instruction which is appropriately targeted to the specific weaknesses of each cluster of students.

2. METHODS AND MATERIALS

2.1. Materials

We selected two well established and reliable tests to assess vocabulary breadth and depth. We assessed vocabulary breadth by means of the *PPVT-III* [3], an orally-administered test of receptive English word knowledge, spanning a range of word frequency levels and relying on picture matching. We assessed depth of vocabulary knowledge by means of the *Word Associates Test* (WAT) [5], a multi-select association task for high frequency English adjectives with synonyms, noun collocates, and distractors. Scores on these two tests provided the data for performing the cluster analysis. To validate the cluster solution, we used a third vocabulary assessment measure (the *Verb Subordinates Test*) designed to simultaneously assess both breadth and depth. This test is a forced choice test of knowledge of superordinate-subordinate verb relations at four verb frequency levels [2].

2.2. Participants

Participants were 230 native English-speaking students enrolled in introductory psychology at Middle Tennessee State University (Age *Mode* = 18, *Mdn* = 19, *M* = 20.23, *Min* = 18, *Max* = 36, *SD* = 3.4). We report results on 229 participants due to a computer malfunction during the administration of the *VST* to one participant.

2.3. Procedure

We received approval from the IRB at Middle Tennessee State University prior to recruiting participants from the Psychology Department pool. Participants who selected to participate, were native speakers of English, and reported no diagnosed learning disabilities completed the three the vocabulary measures in one experimental session. The *PPVT-III* was administered by an experimenter with each participant individually in a private room. The *WAT* was administered at a conference table in an adjoining room. The *VST* was administered on a computer in an adjoining computer lab. The three vocabulary tests were administered in a counterbalanced order.

2.4. Scoring

The *PPVT-III* was scored per its instructions. The *WAT* was scored on an 8-point/item scale. The *VST* was scored by the computer software. To account for the diverse measurement scales, raw scores were normalized and *z*-scores were used in all data analyses.

3. RESULTS

3.1. Descriptive statistics

As the descriptive statistics reflect, our sample's performance reflects an adequately large range of scores in all three vocabulary measures (see Table 1). This finding ensured that this sample of participants was appropriately diverse in vocabulary profiles to yield a rich and informative typology with a cluster analysis.

Table 1: Descriptive statistics for each test ($N = 229$).

Measure	Details	*Mean*	*SD*	*Range*
PPVT-III	Standard Score ($M = 100$, $SD = 15$)	106.14	10.218	80 – 138
	Raw Score (*max* = 204)	177.41	10.39	145 – 200
WAT	8-point Scoring (*max* = 320)	259.03	22.45	164 – 301
VST	Total Score (*max* = 5)	3.94	.38	2.64 – 4.88
	Frequency Band 1	0.95	0.08	0.63 – 1.00
	Frequency Band 2	0.97	0.07	0.63 – 1.00
	Frequency Band 3	0.79	0.15	0.50 – 1.00
	Frequency Band 4	0.63	0.17	0.25 – 1.00
	Frequency Band 5	0.60	0.08	0.25 – 1.00

3.2. Correlation analysis

To ensure that the different tests were, in fact, measuring different competencies related to vocabulary knowledge as designed and expected, we performed a correlation analysis of *z* scores across the three tests. As shown in Table 2, while the correlations are significant as expected, they are only moderate, a finding which suggests that the measures are, indeed, tapping into distinct competencies related to vocabulary knowledge.

Table 2: Intercorrelations among the measures.

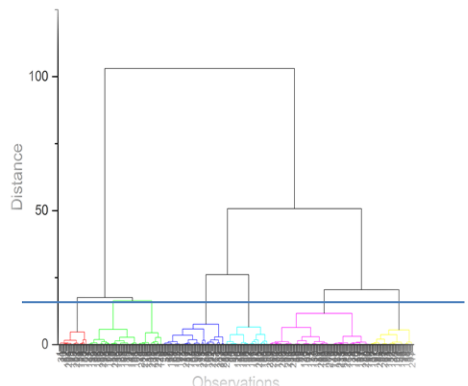
Measure	*PPVT-III*	*WAT*	*VST*
PPVT-III	1	0.62*	0.5*
WAT		1	0.45*
VST			1

3.3. Cluster analysis

The term *cluster analysis* refers to a variety of mathematical methods and procedures which allow researchers to create a classification in a data set by empirically forming discrete groups or "clusters" of similar entities based on similarities among their measured features. For the purposes of the current study, the value of cluster analytic methods is that they can, therefore, generate a typology of individuals based on their vocabulary knowledge profiles reflected by their scores in a combination of vocabulary tests.

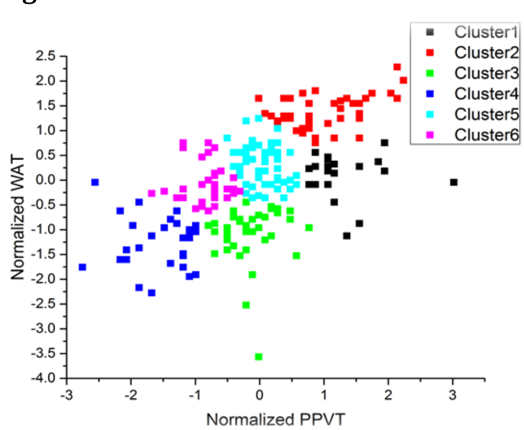
Among cluster analytic methods, hierarchical cluster analysis is particularly suited to situations in which the researcher cannot state *a priori* how many clusters are present in a data set. The method is hierarchical because higher order clusters are formed from lower order clusters. In agglomerative cluster analysis, pairs of entities that are most similar (i.e., entities with the smallest distance between them in terms of their scores in the measured performance or behavior) are merged to form a single cluster, and these clusters are then merged in a stepwise fashion following the same principle until a single cluster groups all the entities in the data set. Hierarchical clustering is best reflected in a dendogram (see Figure 1). Since cluster analysis is an exploratory approach, the interpretation of the resulting dendogram is context dependent. In other words, a researcher must decide what number of clusters is most informative and meaningful for the specific data set on which cluster analysis is performed. We relied on this method as the preliminary analysis of the data in this study. Specifically, we employed Ward's method of hierarchical clustering with Euclidean distance [7]. Analyses were performed in OriginPro (OriginLab, Northampton, MA). Based on the resulting dendogram (see Figure 1), we selected the 6-cluster solution reflected on the dendogram by the horizontal blue line. Note that the difference in length among the branches in the dendogram reflects the size of the distance in performance between the clusters.

Figure 1: Hierarchical cluster analysis dendogram.



Next, we performed a k-means clustering procedure to ensure that participants were grouped into the best fitting cluster given the selected 6-cluster solution. K-means clustering aims to partition *n* entities into *k* clusters in which each entity belongs to the cluster with the nearest mean, serving as a prototype of the cluster. In this case, using the 6 cluster centers drawn from the hierarchical cluster analysis, we performed a k-means clustering with 10 iterations based on maximum distance with the OriginPro statistical software. The cluster membership of this sample appears in Figure 2 with the vocabulary size scores plotted along the x-axis and the vocabulary depth scores plotted along the y-axis.

Figure 2: K-means cluster solution.



3.4. Cluster descriptive statistics

The clusters range in size from 21 to 62 participants (see Table 3). The mean scores on the *PPVT-III* and the *WAT* for each cluster of participants appear in Table 3.

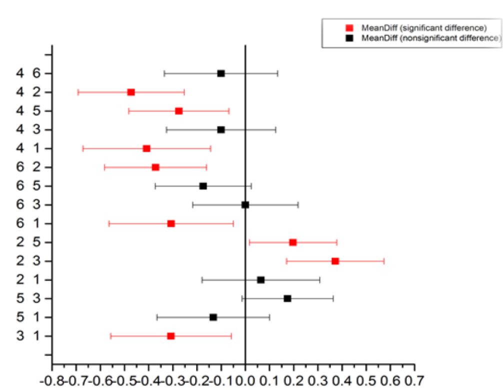
Table 3: Cluster size and mean normalized score on *PPVT-III* and *WAT*.

Cluster	*N*	*Mean PPVT-III*	*Mean WAT*
1	21	1.3	0.09
2	45	1.1	1.39
3	39	-0.16	-1.09
4	29	-1.54	-1.25
5	62	0.06	0.27
6	33	-0.84	-0.06

3.5. Cluster solution validation

To validate our cluster solution, we analyzed the performance of the clusters with respect to the participants' performance on the third measure (the *Verb Subordinates Test*) which was designed to measure simultaneously both vocabulary breadth and depth and which served as the variable external to the clustering process. An analysis of variance showed that the effect of cluster membership on *VST* mean score was significant, $F(5, 223) = 12.79$, $p < .001$. Post hoc comparisons using the Fisher's Least Significant Differences test revealed significant differences in the clusters' mean scores on the *VST* between the following cluster pairs: 4-2, 4-5, 4-1, 6-2, 6-2, 5-2, 3-2, 3-1 (see Figure 4; cluster pairs plotted along the y-axis and *VST* mean scores plotted along the x-axis).

Figure 4: Between cluster comparisons on *VST* scores



4. DISCUSSION

This study illustrates how cluster analysis can be fruitfully used to classify students into groups with similar vocabulary knowledge profiles in terms of both how *many* words a student knows as well as how *well* the student knows individual words. In the current example, we performed a cluster analysis with data from 229 native English speaking college students who completed a measure of vocabulary breadth/size (the *PPVT-III*) and a measure of vocabulary depth/quality (the *Word Associates Test*). The cluster analysis revealed 6 distinct clusters of students classified based on their relative performance on these two tests. This cluster solution was subsequently supported by an ANOVA which confirmed statistically significant differences between a number of pairs of clusters with respect to the students' scores on a third vocabulary test designed specifically as a measure of both vocabulary *breadth* and vocabulary *depth* (the *VST*).

Based on our analyses, the following vocabulary profiles emerged among the college students in our sample. The largest cluster consists of 27.1% of participants. This group's performance was average on both measures. Next in size with 19.65% of participants is the cluster whose vocabulary profile is characterized by an overall above average performance in both vocabulary breadth and depth. 17% of participants compose the cluster characterized by an average sized vocabulary with a below average depth of knowledge of individual words. 14.4% of participants compose the cluster characterized by an average performance on depth of word knowledge with a below average vocabulary size. 9.16% of the sample consists of participants with an above average vocabulary size with an average performance in terms of depth of knowledge of individual words. Lastly, 12.6% of participants compose the cluster characterized by an overall below average vocabulary knowledge in terms of both breadth and depth.

Several practical implications arise from these findings. First, this type of classification of students reveals how to best focus vocabulary instruction to each individual cluster of students given the cluster's strengths and weaknesses with regard to vocabulary knowledge. Second, this type of classification reveals which cluster of students is most in need of vocabulary instruction. Third, a 6-cluster solution offers a more detailed understanding of the vocabulary profiles of students who perform largely in the average range on both measures, as is shown here by the assignment of participants to each of 3 different clusters when performing mostly in the average range on both tests (clusters pink, green, aqua in Figure 2). Finally, assuming that a sample representative of the population of college students enrolled in similar large public universities in the United States, the findings suggest that approximately 13% of admitted students are likely to struggle academically given their vocabulary knowledge profile [6] and could benefit from targeted vocabulary instruction.

5. CONCLUSION

This study confirmed that applying cluster analysis to a model of vocabulary assessment which takes into account the multi-dimensionality of vocabulary knowledge is a promising and valuable endeavor. Based on such an analysis, targeted instruction and intervention can be designed to match the needs of students composing the different clusters that emerge in a population of students.

REFERENCES

1. Anderson, R. C., & Freebody, P. (1981). Vocabulary Knowledge. In John T. Guthrie (Ed.), *Comprehension and Teaching: Research Reviews*, (pp. 77-117). Newark, DE: International Reading Association.
2. Blackwell, A. A. (2012). *The Verb Subordinates Vocabulary Test*. Unpublished manuscript.
3. Dunn, Lloyd M. (1997). *PPVT-III : Peabody picture vocabulary test.* Circle Pines, MN: American Guidance Service.
4. Pearson, P. D., Hiebert, E., & Kamil, M. (2007). "Vocabulary assessment: What we know and what we need to learn," *Reading Research Quarterly*, *42*(2), 282-296.
5. Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in language assessment*, (pp. 41-60). Mahwah, NJ: Erlbaum. (https://www.lextutor.ca/tests/associates/)
6. Treffers-Daller, J., & Milton, J. (2013) Vocabulary size revisited: The link between vocabulary size and academic achievement. *Applied Linguistics Review*, *4*(1), 151-172. http://centaur.reading.ac.uk/29879/
7. Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association 58*(301), 236-244.