Intelligence 46 (2014) 122–130

Contents lists available at ScienceDirect

Intelligence

Quantity and structure of word knowledge across adulthood

Timothy A. Salthouse [1]

Department of Psychology, University of Virginia, Charlottesville, VA 22904-4400, United States

ARTICLE INFO

Article history:
Received 16 February 2014
Received in revised form 1 May 2014
Accepted 1 May 2014
Available online xxxx

Keywords:
Aging
Knowledge
Semantic memory
Construct validity

ABSTRACT

Cross-sectional and longitudinal data from moderately large samples of healthy adults confirmed prior findings of age-related declines in measures of the quantity of word knowledge beginning around age 65. Additional analyses were carried out to investigate the interrelations of different types of vocabulary knowledge at various periods in adulthood. Although the organizational structures were similar in adults of different ages, scores on tests with different formats had weaker relations to a higher-order vocabulary construct beginning when adults were in their 60s. The within-person dispersion among different vocabulary test scores was also greater after about 65 years of age. The discovery of quantitative decreases in amount of knowledge occurring at about the same age as qualitative shifts in the structure of knowledge raises the possibility that the two types of changes may be causally linked.

1. Introduction

As one would expect if knowledge accumulates over time, performance on tests of knowledge has often been reported to be greater at older ages. However, late-life declines in measures of vocabulary have been reported in cross-sectional data based on nationally representative samples (see figures in Salthouse, 1988a,b, 1991, 2003, 2010a), and also in several studies with longitudinal comparisons (e.g., Albert, Heller, & Milberg, 1988; Alder, Adam, & Arenberg, 1990; Anstey, Hofer, & Luszcz, 2003; Christensen et al., 1999; De Frias, Lovden, Lindenberger, & Nilsson, 2007; Ghisletta, McArdle, & Lindenberger, 2006; Schaie, 2005; Sliwinski & Buschke, 1999; Zelinski & Burnight, 1997). Because lack of access to previously available information may be a unique indicator of age-related decline in cognitive functioning, understanding the relations of age to word knowledge could provide valuable insights into the nature of late-life cognitive decline.

Many prior studies have examined only a single measure of vocabulary knowledge, but if multiple vocabulary measures are available relations among the measures can be examined to investigate the structure of a vocabulary construct at different ages. That is, not only can the amount of knowledge be assessed, in terms of the level of each measure, but also the cohesiveness of the vocabulary knowledge construct can be examined by the relations among the different measures.

A popular conceptualization of knowledge representation is a network in which the nodes correspond to semantic, phonological, or orthographic information (e.g., Burke, MacKay, & James, 2000; Burke, MacKay, Worthley, & Wade, 1991; Salthouse, 1988a,b). Because vocabulary tests in different formats vary in terms of the information that is provided and the information that is requested (e.g., Bowles & Salthouse, 2008; Rabaglia & Salthouse, 2011; Verhaeghen, 2003), different test formats can be postulated to involve different access routes to semantic information. That is, naming pictured objects requires that meaning is accessed and the phonological representation is activated, providing a definition of a target word requires that meaning be accessed from the phonological representation, and tests of synonyms and antonyms involve comparison of meanings with either the same or opposite connotations (cf. Rabaglia & Salthouse, 2011). If different types of vocabulary tests can be assumed to vary with respect to the

E-mail address: Salthouse@virginia.edu.
[1] This research was supported by Award Number R37AG024270 from the National Institute on Aging. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute on Aging or the National Institutes of Health. There are no conflicts of interest.

http://dx.doi.org/10.1016/j.intell.2014.05.009
0160-2896/

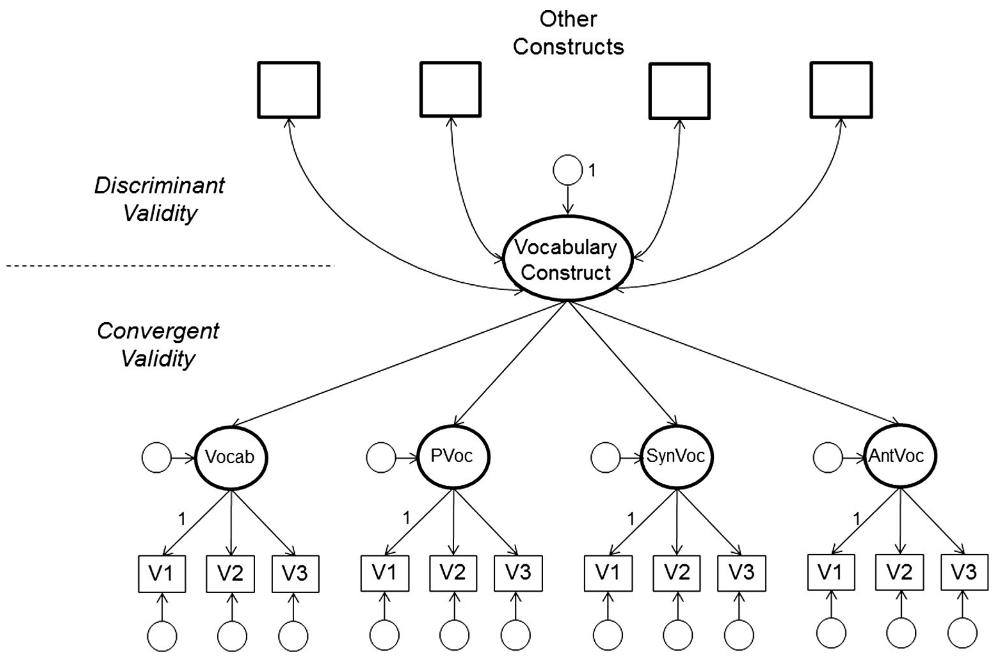

Fig. 1. Schematic representation of a vocabulary construct defined by three versions in each of four tests. Vocab refers to WAIS Vocabulary, PVoc to Picture Vocabulary, Synvoc to synonym vocabulary, and Antvoc to antonym vocabulary. V1, V2, and V3 refer to different versions of tests with the same format but involving different items. Relations with ability constructs in the top panel are relevant to discriminant validity and relations of the vocabulary construct to the vocabulary test constructs, and of the test constructs to the test versions in the bottom panel are relevant to convergent validity.

aspects of the semantic network that are involved, the cohesiveness of a vocabulary construct can be investigated by examining the strength of the relations among scores in tests involving different formats. That is, a more cohesive or tightly organized construct should have smaller variability across scores from different types of tests, and stronger relations among the scores in those tests.

There were therefore two primary goals of the present study: (1) further investigate the relations of age to vocabulary knowledge in both cross-sectional and longitudinal comparisons, and (2) investigate possible qualitative differences in word knowledge by determining whether increased age was associated with a shift in the structural organization of measures of vocabulary knowledge obtained from different test formats.

The first goal was pursued by examining the age trends on individual and composite measures of vocabulary from four different tests, each with three versions comprised of different items. The sample of participants consisted of over 4700 adults with cross-sectional data, and over 2200 adults with two-occasion longitudinal data. Because the age trends suggested the existence of two segments, spline regression analyses were conducted to determine the age corresponding to the transition between the two segments.

The second goal was investigated by examining relations among different vocabulary measures at both within-individual and between-individual levels of analysis. Within-individual comparisons were based on assessments of across-test variability. The rationale was that if the construct is becoming less cohesive with increased age, one might expect greater divergence, in the form of increased across-test variability, among the scores on different types of tests at older ages.

Between-individual comparisons were examined in the context of a hierarchical structure of word knowledge, as portrayed in the bottom of Fig. 1.[2] Note that the lowest level in the hierarchy consists of scores in the different test versions, the next level consists of constructs corresponding to the four different tests, and the highest level corresponds to a broad vocabulary construct. Relations from the first to the second levels, and from the second to the third levels, are relevant to whether the tests are all assessing the same construct, and hence serve to evaluate convergent validity of the vocabulary construct. When assessing construct validity it is also important to evaluate discriminant validity by determining whether the vocabulary construct is distinct from other constructs. Information relevant to this question is available in the magnitude of the relations of the vocabulary construct with measures of constructs representing different cognitive abilities because the correlations with cognitive abilities should be weak if the vocabulary construct represents something distinct from the other constructs.

Possible age differences in the structure of vocabulary knowledge were investigated by examining the fit of the model in different age groups, and comparing the magnitude of each parameter to determine where differences might exist in the structure. For example, differences could be evident at the lowest level, in the form of weaker relations of the test constructs to the test versions. Alternatively, if processes associated with aging have differential impact on the modes

[2] Although some (e.g., Kan et al., 2011) have advocated that a formative approach be used to model vocabulary measures, a reflective approach was used in this study because a latent variable of vocabulary knowledge was assumed to contribute to the scores on different types of tests.

of accessing information (e.g., recognition or retrieval) required by the different test formats, increased age could be associated with weaker relations of the broad vocabulary construct to the relevant test constructs. Finally, increased age might be associated with stronger relations of the vocabulary construct to other constructs, as implied by the de-differentiation concept (e.g., Cunningham, 1980).

The data in the current report were derived from the Virginia Cognitive Aging Project (VCAP), which is an on-going longitudinal study of aging and cognition involving a moderately large sample of adults ranging from 18 to over 90 years of age (e.g., Salthouse, 2009, 2010b, 2011; Salthouse, Pink, & Tucker-Drob, 2008). A unique feature of VCAP is a measurement-burst design involving the administration of parallel versions of each test on three separate sessions at each longitudinal occasion. In addition to having the potential to increase sensitivity by aggregation of scores across sessions, the availability of three versions of each test allows fine-grained analyses of the structure of vocabulary knowledge by examining relations among different versions of the same tests and relations among different types of tests, as portrayed in Fig. 1.

2. Method

2.1. Participants

Characteristics of the participants by age decade are summarized in Table 1. Approximately 79% of the participants were Caucasian, 11% African-American, and the remainder from other ethnicities or reporting multiple classifications. Because of the small number, participants over 90 years of age were not included in the current analyses. It can be seen that the ages ranged from 18 to 89, and that the average Mini Mental State Exam (Folstein, Folstein, & McHugh, 1975) scores were all above 28.

A total of 4774 adults participated at the first occasion (T1), but because the measurement burst design was not implemented for all participants at the first occasion, only 2604 of them performed parallel versions of the tests on sessions 2 and 3 of that occasion. The remaining participants performed a single version of each test on session 1, and performed different tests on sessions 2 and 3. A total of 2257 participants returned at the second occasion (T2), and all of them completed parallel versions of the vocabulary tests on the second and third sessions.

The interval between T1 and T2 averaged 3.0 years, with a range of less than 1 to over 12 years. However, there were very weak correlations of the T1–T2 interval with age ($r = -.09$) and with the T1–T2 differences in the vocabulary composite scores ($r = -.08$), and therefore interval was ignored in the analyses.

Cross-sectional relations of individual vocabulary variables to age and to other cognitive variables were reported by Bowles and Salthouse (2008) in a sample that partially overlapped with the current one (i.e., about 16% of the current sample was included in the earlier report). However, no longitudinal data were included in the previous study, and it did not investigate structural relations among the vocabulary measures. Relations between the vocabulary construct and constructs representing cognitive abilities were examined in a subset (about 47%) of the current sample by Tucker-Drob and Salthouse (2008). However, that report did not evaluate the significance of differences in the structural relations between age groups, and all of the data were cross-sectional.

2.2. Tests

There were four primary vocabulary tests. The WAIS III vocabulary (Wechsler, 1997a) test consisted of 33 items in which the examinee was to provide the definition of a target word. Scoring of each item was on a 0-to-2 scale according to criteria in the administration manual specifying the quality of the definition. A second test was the Picture Vocabulary test from the Woodcock–Johnson Cognitive Ability Battery (Woodcock & Mather, 1990), which consisted of 30 items in which the examinee was to name the designated object in a picture. The third and fourth tests were four-alternative multiple-choice synonym and antonym vocabulary tests with 10 items each (Salthouse, 1993).

Parallel versions of the original tests were created with an identical format, but consisting of different items. Correlations among the three versions at T1 ranged from .54 to .81, and composite scores created by averaging the average *z*-scores for the four vocabulary variables at each session, had across-session correlations ranging from .83 to .87. Reliability of each test version within each age group was estimated from coefficient alpha, and these values are presented in Table 2. Although the reliability estimates of the second and third versions of the synonym and antonym vocabulary tests were relatively low, there was little evidence of systematic age differences in the reliability of any of the vocabulary tests.

Table 1
Means (and standard deviations) of demographic characteristics of the participants by age decade.

Age decade	N1/N2	T1 age	Prop. female	T1 health	Years of education	T1 MMSE	T2–T1 interval
20s	850/245	23.2 (3.2)	.59	2.0 (0.9)	14.7 (2.1)	28.8 (1.6)	3.1 (1.9)
30s	461/193	34.3 (2.8)	.70	2.1 (0.8)	15.7 (2.8)	28.5 (1.8)	3.1 (2.0)
40s	743/392	45.0 (2.9)	.71	2.1 (0.9)	15.3 (2.7)	28.5 (1.8)	3.3 (1.9)
50s	1104/596	54.4 (2.8)	.71	2.2 (0.9)	15.7 (2.6)	28.4 (1.9)	3.0 (1.6)
60s	842/429	64.1 (2.8)	.65	2.1 (0.9)	16.4 (2.8)	28.5 (1.9)	2.9 (1.4)
70s	539/301	74.2 (2.9)	.58	2.4 (0.9)	15.9 (2.9)	28.1 (1.9)	2.7 (1.2)
80s	235/101	83.1 (2.5)	.51	2.6 (0.8)	16.1 (3.1)	27.2 (2.5)	2.5 (1.1)
All	4774/2257	50.8 (18.0)	.64	2.2 (0.9)	15.6 (2.7)	28.4 (1.9)	3.0 (1.6)

Note: Health is a self-rating on a scale from 1 for excellent to 5 for poor. MMSE is the Mini Mental State Exam (Folstein et al., 1975). N1 and N2 refer to the number of participants in the cross-sectional and longitudinal samples. The T2–T1 interval is in years.

Table 2
Estimates of reliability (coefficient alpha) of vocabulary measures in the four age groups.

	18–39	40–59	60–69	70–89
WAIS vocabulary				
Version 1	.89	.88	.86	.87
Version 2	.84	.87	.86	.84
Version 3	.87	.87	.83	.82
Picture vocabulary				
Version 1	.87	.84	.78	.81
Version 2	.79	.78	.73	.74
Version 3	.80	.76	.74	.77
Synonym vocabulary				
Version 1	.79	.84	.80	.81
Version 2	.68	.74	.65	.59
Version 3	.54	.60	.63	.64
Antonym vocabulary				
Version 1	.76	.83	.78	.79
Version 2	.61	.63	.60	.57
Version 3	.54	.63	.60	.59
Median	.79	.81	.76	.78

The 35-item version of the North American Adult Reading Test (NAART, Uttl, 2002) was administered to 3165 participants at T1 and to 2027 participants at T2. The coefficient alpha at T1 was .93, which was nearly identical to the value of .92 in Uttl (2002). The correlation of the NAART score with the WAIS Vocabulary score was .76, which was the same as that reported by Uttl (2002).

In addition to the vocabulary tests, all of the participants completed a battery of cognitive tests assessing reasoning, spatial visualization, episodic memory, and perceptual speed abilities, and relations of the vocabulary construct to those constructs was examined to investigate its discriminant validity. The tests, and their sources, are briefly described in the Appendix A. Additional information about the reliabilities and validities is available in other articles (e.g., Salthouse, 2004, 2005, 2010b; Salthouse & Ferrer-Caja, 2003; Salthouse et al., 2008). Composite scores for each cognitive ability were formed by averaging relevant *z*-scores based on the mean and standard deviation of the complete sample at the first session.

3. Results[3]

3.1. Age relations

Fig. 2 portrays means (and standard errors) of *z*-scores in the four vocabulary tests on the first session of the first occasion as a function of age decade. Notice that with each measure there was a similar pattern of an increase, followed either by stability (for synonym vocabulary) or decline (for the other three vocabulary measures).

Segmented (spline) regression analyses were used to estimate the age corresponding to the best-fitting boundary between the two segments of the functions (cf. Bowles & Salthouse, 2008). The analyses specified two segments and

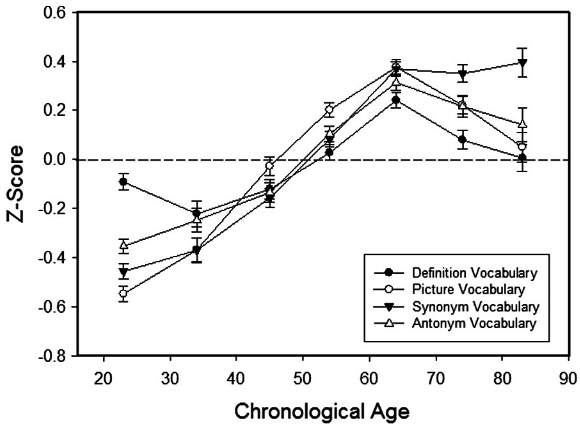

Fig. 2. Means (and standard errors) of scores on four vocabulary tests on the first session of the first occasion by age decade. Definition vocabulary was assessed with the WAIS Vocabulary test.

then used an iterative process to estimate the boundary age between the segments, and the slopes in each segment. The estimated ages from these analyses, and the age slopes of the functions before and after the boundary age, are reported in Table 3. Inspection of the table reveals that the transition ages were between 64 and 66 years of age, with positive slopes before the transition age and, with the exception of synonym vocabulary, negative slopes after the transition age.

Also reported in Table 3 are the results of segmented regression analyses for the composite vocabulary scores on each session. Note that the estimated transition ages between the two segments were 65.0, 65.4, and 69.0 years in sessions 1, 2, and 3, respectively, with positive slopes in the first segment and negative slopes in the second segment.

Table 3
Results of two-segment regression analyses.

Variable	N	Transition age	Slope 1	Slope 2
T1 session 1				
Definition vocabulary	4769	66.0	.008*	−.012*
Picture vocabulary	4765	64.4	.025*	−.021*
Synonym vocabulary	4646	65.0	.020*	.001
Antonym vocabulary	4626	65.0	.017*	−.010
Session 1				
T1 mean	4609	65.0	.017*	−.009*
T1 within-person SD	4770	65.3	−.002*	.009*
T2–T1 difference	2158	76.0	−.004*	−.013
T2–T1 within-person SD	2252	68.7	−.001*	.017*
Session 2				
T1 mean	2543	65.4	.008*	−.012*
T1 within-person SD	2608	66.0	−.001	.004*
T2–T1 difference	1002	67.0	−.001	−.010
T2–T1 within-person SD	1046	42.0	−.003	.003*
Session 3				
T1 mean	2531	69.0	.009*	−.010
T1 within-person SD	2580	76.0	.000	.011
T2–T1 difference	994	68.7	−.005*	−.004
T2–T1 within-person SD	1031	76.0	.000	.014

Note: The slope is the z-score difference per year of age before (Slope 1) and after (Slope 2) the transition age.
* $p < .01$.

[3] Because of the moderately large sample size, an alpha of .01 was used in all significance tests.

The vocabulary composite scores in the first session at T1 and T2 for participants with data on at least two occasions are portrayed as a function of age decade in Fig. 3. It can be seen that the scores were higher with increased age in both the cross-sectional (separate sets of lines) and longitudinal (solid lines) comparisons until the decade of the 60s, after which there was decline in both between-person (cross-sectional) and within-person (longitudinal) contrasts.

Segmented regression analyses on the T2–T1 difference scores revealed transition points between the two segments at age 76.0 for the session 1 composite score difference, age 67.0 for the session 2 difference, and age 68.7 for the session 3 difference. Although not always significantly different from zero, the slopes relating T2–T1 difference to age were negative in both the first and second segments in each session.

3.2. Across-test variability

Cohesiveness of the vocabulary construct at the level of individual participants was assessed by computing standard deviations across the four vocabulary scores for each individual. The means (and standard errors) of the within-person standard deviations are portrayed by age decade in Fig. 4, where it can be seen that the standard deviations were relatively constant across successive ages until they increased after the decade of the 60s. Because the synonym vocabulary measure had a somewhat different age trajectory than the other measures, standard deviations were also computed without the synonym vocabulary measure. The pattern without the synonym vocabulary measure was very similar to that portrayed in Fig. 4 as the average standard deviations for adults in their 60s, 70s, and 80s were .45, .59, and .60, respectively.

Results of segmented regression analyses for the within-person standard deviations at the first occasion are reported in Table 3. Note that the transitions were at ages 65.3, 66.0, and 76.0 for the within-person standard deviations on sessions 1, 2, and 3. As expected from Fig. 4, the slopes of the within-person standard deviations were more positive in the second segment than in the first segment.

Within-person standard deviations were also computed for the T2–T1 differences in the individuals with longitudinal data. The transition age for the two segments with the T2–T1 differences in within-person standard deviations were 68.7, 42.0, and 76.0, respectively, for sessions 1, 2, and 3. The means (and standard errors) of the standard deviations for the T2–T1 differences are plotted in Fig. 4, where it can be seen that the pattern with the longitudinal differences was very similar to that with the T1 means as in both cases the within-person variability was relatively flat until the decades of the 70s and 80s.

3.3. Cohesiveness of a vocabulary construct

Convergent and discriminant validity of the vocabulary construct was investigated with the model portrayed in Fig. 1. Because the results reported above suggested that there were different patterns before and after about age 65, separate analyses were conducted in four age groups, two groups (ages 18-to-39 and ages 40-to-59) prior to the transition age, one group encompassing the transition age (age 60-to-69), and one group after that age (age 70-to-89). Identical analyses were conducted for each age group, and the results in Table 4 indicate that the hierarchical model had a good fit to the data in each age group. Moreover, the estimates were reasonably precise as the 99% confidence intervals for the unstandardized coefficients were all relatively small. The high coefficients for the relations of the test constructs to the test versions, and for the relations of the highest-order vocabulary construct to the test constructs, provide evidence for convergent validity of a vocabulary construct. In addition, with the exception of the NAART variable, which can be postulated to be another measure of vocabulary, the correlations of the vocabulary construct with composite scores representing other cognitive abilities were generally modest, which is consistent with discriminant validity of the vocabulary construct.

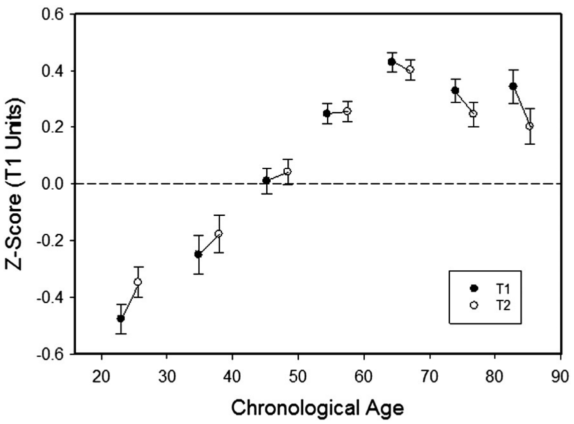

Fig. 3. Means (and standard errors) of composite vocabulary scores for participants with data on the first session at both T1 and T2 by age decade. The T2–T1 differences were significantly ($p < .01$) positive in the decades of the 20s and 30s, and significantly ($p < .01$) negative in the decades of the 70s and 80s.

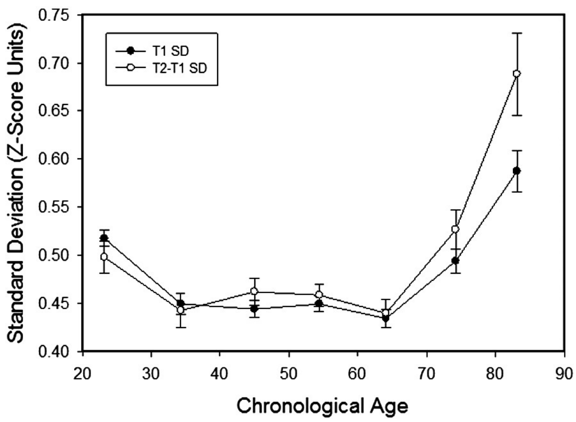

Fig. 4. Means (and standard errors) of within-person standard deviations for vocabulary test scores at T1, and for the T2 – T1 differences in vocabulary test scores by age decade.

Table 4
Unstandardized coefficients (and 95% confidence intervals) for the model in Fig. 1 in adults in four age groups.

	Age group			
	18-to-39	40-to-59	60-to-69	70-to-89
Vocabulary construct				
➔ WAIS Vocabulary	.92 (.87–.96)	.93 (.89–.97)	.70 (.65–.75)	.69 (.64–.74)
➔ Picture vocabulary	.81 (.76–.85)	.84 (.80–.88)	.61 (.56–.66)	.56 (.50–.62)
➔ Synonym vocabulary	.86 (.81–.90)	.90 (.86–.94)	.73 (.68–.77)	.69 (.64–.74)
➔ Antonym Vocabulary	.76 (.72–.80)	.88 (.84–.92)	.78 (.73–.83)	.74 (.67–.80)
WAIS vocabulary				
➔ Version 1[a]	1.0	1.0	1.0	1.0
➔ Version 2	.62 (.59–.66)	.66 (.63–.69)	.74 (.69–.80)	.76 (.67–.84)
➔ Version 3	.69 (.65–.73)	.70 (.66–.74)	.63 (.58–.69)	.57 (.50–.63)
Picture vocabulary				
➔ Version 1[a]	1.0	1.0	1.0	1.0
➔ Version 2	.72 (.68–.76)	.67 (.64–.70)	.69 (.64–.74)	.80 (.74–.87)
➔ Version 3	.67 (.61–.72)	.71 (.67–.74)	.70 (.64–.76)	.87 (.79–.95)
Synonym vocabulary				
➔ Version 1[a]	1.0	1.0	1.0	1.0
➔ Version 2	.71 (.66–.76)	71 (.67–.75)	.64 (.58–.69)	.64 (.57–.71)
➔ Version 3	.56 (.51–.61)	.60 (.56–.63)	.73 (.68–.79)	.76 (.68–.84)
Antonym vocabulary				
➔ Version 1[a]	1.0	1.0	1.0	1.0
➔ Version 2	.55 (.50–.59)	.48 (.45–.51)	.45 (.40–.50)	.45 (.39–.51)
➔ Version 3	.61 (.56–.66)	.59 (.55–.62)	.61 (.56–.67)	.63 (.55–.70)
Relations with vocabulary construct				
NAART	.86 (.82–.91)	.99 (.94–1.03)	.79 (.74–.84)	.70 (.65–.76)
Memory composite	.46 (.42–.50)	.51 (.47–.54)	.38 (.33–.43)	.37 (.31–.43)
Speed composite	.28 (.24–.32)	.28 (.25–.31)	.25 (.20–.29)	.29 (.24–.34)
Reasoning composite	.54 (.50–.58)	.60 (.57–.64)	.53 (.49–.58)	.52 (.46–.57)
Space composite	.56 (.51–.60)	.46 (.42–.49)	.34 (.29–.38)	.29 (.24–.33)
Chi-square/df	4.53	6.02	3.08	3.60
CFI	.97	.97	.97	.95
RMSEA	.05	.05	.05	.06

Note: CFI refers to comparative fit index, and RMSEA refers to root mean squared error of approximation. CFI values greater than .90 and RMSEA values less than .08 are often considered to reflect good fit (Kline, 2005).
[a] This relation was fixed at 1.0 to provide a scale for the construct and therefore confidence intervals could not be computed.

Inspection of the values in Table 4 reveals that although the test-version coefficients were fairly similar across age groups, the relations from the vocabulary construct to the constructs representing different test formats were weaker beginning in the decade of the 60s. This pattern, together with the lack of age differences in the reliability estimates in Table 2, suggests that age-related differences are evident in the coherence of the broad vocabulary construct, but not in the degree to which individual items assess different test versions or in the degree to which different test versions assess constructs corresponding to different types of vocabulary tests.

4. Discussion

The results reported above indicate that the relations between age and vocabulary knowledge in this study are distinctly non-linear, with an increase followed by either stability or a decrease. The cross-sectional and longitudinal data were consistent in indicating that negative between-person differences and within-person changes appear to begin when people are about 65 years of age. Moreover, because the patterns were similar in three different sets of items (based on different test versions administered in separate sessions), and are consistent with the findings of earlier studies cited in the introduction, the results can be considered robust.

In addition to evaluating level of vocabulary performance, cohesiveness of a vocabulary construct was assessed both within participants and across participants. The within-person assessment was evaluated in terms of the magnitude of within-person variability, or dispersion, among the vocabulary test scores. The results in Fig. 4 and in Table 3 indicate that there was a marked increase in the within-person variability of the test scores beginning in the decade of the 60s. These findings suggest that different measures of vocabulary were exhibiting greater dispersion at the level of individual participants at nearly the same age as the cross-sectional and longitudinal decreases in level of performance.

The across-participant analyses were based on the hierarchical model in Fig. 1 in which observed scores in the three versions of each test were at the lowest level, the test constructs were at the second level, and a vocabulary construct was at the highest level. The results in Table 4 indicated that the model had a good fit in four age groups, and specific contrasts revealed that the groups had generally similar relations of the test constructs to the test versions, and of the vocabulary construct with other cognitive abilities. However, the relations of the higher-order vocabulary construct to the test constructs were weaker among adults 60-to-69 years of age compared to adults 40-to-59 years of age. These results suggest that although the vocabulary construct was distinct from other cognitive abilities in each age group, it was less cohesive, in the sense of weaker relations among scores from different test formats, beginning sometime in the decade of the 60s.

What might account for the loss of previously available word knowledge found in this and earlier studies? Two possible interpretations can be distinguished based on deficits in either the retrieval or representation of semantic information. The retrieval view postulates that decreases in vocabulary test performance are attributable to impaired access to intact representations of the semantic information, whereas the representation view attributes the decrease in vocabulary test performance to degradation of either the strength or the organization of the relevant semantic information.

Among the evidence that could be cited as evidence for the retrieval view are differential deficits for vocabulary measures presumed to vary in their retrieval requirements (i.e., Goral, Spiro, Albert, Obler, & Connor, 2007), higher levels of performance after providing a cue that presumably facilitates access (i.e., Au et al., 1995), and inconsistent patterns of performance across longitudinal occasions (i.e., Barresi, Nicholas, Connor, Obler, & Albert, 2000). Although not specifically formulated to account for age trends in vocabulary tests, the Node Structure Theory (e.g., Burke et al., 1991, 2000) can also be considered a retrieval interpretation because not only have the theorists claimed that there is little effect of age on the representation of semantic knowledge, but also a central postulate of the theory is that retrieval or access problems increase with age.

The Node Structure Theory has primarily been applied to the tip-of-the-tongue (TOT) phenomenon, in which the individual is presumed to possess the relevant knowledge but cannot access it when probed. The TOT phenomenon almost certainly reflects problems with retrieval, but there are reasons to question whether these problems are a major factor contributing to late-life decreases in vocabulary knowledge. Results of a recent study by Salthouse and Mandell (2013) are relevant to this point. Participants across a wide range of ages were presented pictures or definitions, and asked to name the target item. If the target could not be immediately named, the participant was asked to indicate whether he or she was in the TOT state, which was defined as knowing the name but not being able to report it. Finally, four response alternatives were presented with the participant asked to select which one was the correct target. In order to ensure that the participant actually knew the relevant information, responses were classified as TOTs only if the participants reported being in a TOT state and the correct alternative was selected in the subsequent multiple choice test. One of the major findings in the Salthouse and Mandell (2013) study was that TOT rates were higher across the same age range when vocabulary test performance was increasing. These results suggest that difficulties in retrieval increase gradually across adulthood, whereas the results reported above indicate that decreases in vocabulary test performance are only apparent in the decade of the 60s.

The primary evidence cited in favor of intact semantic representations in late life has been the lack of age differences in the pattern of category exemplars (i.e., Howard, 1980) and word associations (i.e., Burke & Peters, 1986). However, these studies only involved cross-sectional data, and age differences within the samples of older adults were not analyzed. Furthermore, the power to detect possible differences in the decade of the 60s may have been low because there were only 50 adults across an age range of 60 to 79 years of age in the Howard (1980) study, and 80 adults across a range of 62 to 87 years of age in the Burke and Peters (1986) study.

In contrast to these early studies, the number of adults in their 60s in the current study was 842 in the cross-sectional sample, and 429 in the longitudinal sample. In addition, both within-person (across-test variability) and between-person (relations of different test formats to a latent vocabulary variable) analyses suggested that the representation of semantic information pertaining to words was less cohesive beginning in the decade of the 60s.

A tentative interpretation of the existing results is that both retrieval problems and representational deficits contribute to late-life declines in vocabulary test performance. However, retrieval problems do not appear to be the primary source of the declines because they are evident at all ages in adulthood, even in the range when vocabulary performance is increasing. In contrast, because the disorganization of semantic representations is pronounced only after about age 60, it may be the major factor contributing to the age-related decreases in vocabulary performance. This framework is obviously speculative, but it leads to several implications that should be testable with moderately large samples of adults. For example, age-related declines beginning in the decade of the 60s would be expected both with sensitive behavioral methods of evaluating the organization of semantic information, and with measures of functional connectivity among brain regions involved in the representation of semantic information.

As with all studies, the current study has a number of limitations that could restrict the generalizability of the results. For example, the average longitudinal interval of 3 years was relatively short, and results are only reported for two measurement occasions. The participants were also relatively high functioning, with an average of nearly 16 years of education, and different patterns might be evident among individuals with fewer years of education. In addition, all of the knowledge tests assessed word knowledge, and it is not known whether similar results would be found with tests of broader knowledge.

In summary, the results of this study confirm earlier findings that the relation of age to vocabulary knowledge is not linear, but instead is characterized by an increase followed by a decrease beginning approximately when people are in the decade of the 60s. Because the same pattern was evident in both cross-sectional and longitudinal comparisons, and across different items with different types of tests, the phenomenon of an increase followed by a decrease in vocabulary knowledge can be considered robust. An intriguing shift in the strength of the interrelations among measures from different types of vocabulary types was also evident at about the same age. Important goals for future research are to determine whether a single explanation will be able to account for the discontinuity in the relations of age to word knowledge and the weaker cohesiveness of the vocabulary construct at older ages, and whether one or both of these shifts is related to the onset of pathological decline.

Appendix A. Description of cognitive variables and their sources

Variable	Description	Source
Reasoning		
Matrix reasoning	Determine which pattern best completes the missing cell in a matrix	Raven (1962)
Shipley abstraction	Determine the words or numbers that are the best continuation of a sequence	Zachary (1986)
Letter sets	Identify which of five groups of letters is different from the others	Ekstrom, French, Harman, and Dermen (1976)
Spatial visualization (space)		
Spatial relations	Determine the correspondence between a 3-D figure and alternative 2-D figures	Bennett, Seashore, and Wesman (1997)
Paper folding	Determine the pattern of holes that would result from a sequence of folds and a punch through folded paper	Ekstrom et al. (1976)
Form boards	Determine which combinations of shapes are needed to fill a larger shape	Ekstrom et al. (1976)
Memory		
Logical memory	Number of idea units recalled across three stories	Wechsler (1997b)
Word recall	Number of words recalled across trials 1 to 4 of a word list	Wechsler (1997b)
Paired associates	Number of response terms recalled when presented with a stimulus item	Salthouse, Fristoe, and Rhee (1996)
Speed		
Digit symbol	Use a code table to write the correct symbol below each digit	Wechsler (1997a)
Letter comparison	Same/different comparison of pairs of letter strings	Salthouse and Babcock (1991)
Pattern comparison	Same/different comparison of pairs of line patterns	Salthouse and Babcock (1991)

References

Albert, M. S., Heller, H. S., & Milberg, W. (1988). Changes in naming ability with age. *Psychology and Aging*, *3*, 173–178.

Alder, A. G., Adam, J., & Arenberg, D. (1990). Individual differences assessment of the relationship between change in and initial level of adult cognitive functioning. *Psychology and Aging*, *1990*(5), 560–568.

Anstey, K. J., Hofer, S. M., & Luszcz, M. A. (2003). A latent growth curve analysis of late-life sensory and cognitive function over 8 years: Evidence for specific and common factors underlying change. *Psychology and Aging*, *18*, 714–726.

Au, R., Joung, P., Nicholas, M., Obler, L. K., Kass, R., & Albert, M. L. (1995). Naming ability across the adult life span. *Aging and Cognition*, *2*, 300–311.

Barresi, B., Nicholas, M., Connor, L. T., Obler, L. K., & Albert, M. L. (2000). Semantic degradation and lexical access in age-related naming failures. *Aging, Neuropsychology, and Cognition*, *7*, 169–178.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1997). *Differential aptitude test.* San Antonio, TX: The Psychological Corporation.

Bowles, R. P., & Salthouse, T. A. (2008). Vocabulary test format and differential relations to age. *Psychology and Aging*, *23*, 366–376.

Burke, D. M., MacKay, D. G., & James, L. E. (2000). Theoretical approaches to language and aging. In T. J. Perfect, & E. A. Maylor (Eds.), *Models of cognitive aging* (pp. 204–237). Oxford, England: Oxford University Press.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language*, *30*, 542–579.

Burke, D. M., & Peters, L. (1986). Word associations in old age: Evidence for consistency in semantic encoding during adulthood. *Psychology and Aging*, *1*, 283–292.

Christensen, H., Mackinnon, A. J., Korten, A. E., Jorm, A. F., Henderson, A. S., Jacomb, P., et al. (1999). An analysis of diversity in the cognitive performance of elderly community dwellers: Individual differences in change scores as a function of age. *Psychology and Aging*, *14*, 365–379.

Cunningham, W. R. (1980). Age comparative factor analysis of ability variables in adulthood and old age. *Intelligence*, *4*, 133–149.

De Frias, C. M., Lovden, M., Lindenberger, U., & Nilsson, L. -G. (2007). Revisiting the dedifferentiation hypothesis with longitudinal multi-cohort data. *Intelligence*, *35*, 381–392.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Ghisletta, P., McArdle, J. J., & Lindenberger, U. (2006). Longitudinal cognition–survival relations in old and very old age. *European Psychologist*, *11*, 204–223.

Goral, M., Spiro, A., Albert, M. L., Obler, L. K., & Connor, L. T. (2007). Change in lexical retrieval skills in adulthood. *The Mental Lexicon*, *2*, 215–240.

Howard, D. V. (1980). Category norms: A comparison of the Battig and Montague (1969) norms with the responses of adults between the ages of 20 and 80. *Journal of Gerontology*, *35*, 225–231.

Kan, K-J., Kievit, R. A., Dolan, C., & van der Maas, H. (2011). On the interpretation of the CHC factor Gc. *Intelligence*, *39*, 292–302.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.

Rabaglia, C., & Salthouse, T. A. (2011). Natural and constrained language production as a function of age and cognitive abilities. *Language and Cognitive Processes*, *26*, 1505–1531.

Raven, J. (1962). *Advanced progressive matrices, set II.* London: H.K. Lewis.

Salthouse, T. A. (1988a). Effects of aging on verbal abilities: An examination of the psychometric literature. In L. L. Light, & D. M. Burke (Eds.), *Language and memory in old age*. New York: Cambridge University Press.

Salthouse, T. A. (1988b). Initializing the formalization of theories of cognitive aging. *Psychology and Aging*, *3*, 3–16.

Salthouse, T. A. (1991). *Theoretical perspectives on cognitive aging.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Salthouse, T. A. (1993). Speed and knowledge as determinants of adult age differences in verbal tasks. *Journal of Gerontology: Psychological Sciences*, *48*, P29–P36.

Salthouse, T. A. (2003). Interrelations of aging, knowledge, and cognitive performance. In U. Staudinger, & U. Lindenberger (Eds.), *Understanding human development: Lifespan psychology in exchange with other disciplines*. Berlin: Kluwer Academic Publishers.

Salthouse, T. A. (2004). Localizing age-related individual differences in a hierarchical structure. *Intelligence*, *32*, 541–561.

Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology*, *19*, 532–545.

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, *30*, 507–514.

Salthouse, T. A. (2010a). *Major issues in cognitive aging.* New York: Oxford University Press.

Salthouse, T. A. (2010b). Influence of age on practice effects in longitudinal neurocognitive change. *Neuropsychology*, *24*, 563–572.

Salthouse, T. A. (2011). Effects of age on time-dependent cognitive change. *Psychological Science*, *22*, 682–688.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*, 763–776.

Salthouse, T. A., & Ferrer-Caja, E. (2003). What needs to be explained to account for age-related effects on multiple cognitive variables? *Psychology and Aging*, *18*, 91–110.

Salthouse, T. A., Fristoe, N., & Rhee, S. H. (1996). How localized are age-related effects on neuropsychological measures? *Neuropsychology*, *10*, 272–285.

Salthouse, T. A., & Mandell, A. R. (2013). Do age-related increases in tip-of-the-tongue experiences signify episodic memory impairments? *Psychological Science*, *24*, 2489–2497.
Salthouse, T. A., Pink, J. E., & Tucker-Drob, E. M. (2008). Contextual analysis of fluid intelligence. *Intelligence*, *36*, 464–486.
Schaie, K. W. (2005). *Developmental influences on adult intelligence: The Seattle longitudinal study*. New York, NY: Oxford University Press.
Sliwinski, M., & Buschke, H. (1999). Cross-sectional and longitudinal relationships among age, cognition, and processing speed. *Psychology and Aging*, *14*, 18–33.
Tucker-Drob, E. M., & Salthouse, T. A. (2008). Adult age trends in the relations among cognitive abilities. *Psychology and Aging*, *23*, 453–460.
Uttl, B. (2002). North American Adult Reading Test: Age norms, reliability, and validity. *Journal of Clinical and Experimental Neuropsychology*, *24*, 1123–1137.
Verhaeghen, P. (2003). Aging and vocabulary scores: A meta-analysis. *Psychology and Aging*, *18*, 332–339.
Wechsler, D. (1997a). *Wechsler adult intelligence scale* (3rd ed.). San Antonio, TX: The Psychological Corporation.
Wechsler, D. (1997b). *Wechsler memory scale* (3rd ed.). San Antonio, TX: The Psychological Corporation.
Woodcock, R. W., & Mather, N. (1990). *Woodcock–Johnson tests of cognitive ability. Standard and supplemental batteries: Examiner's manual.* Allen, TX: DLM.
Zachary, R. A. (1986). *Shipley Institute of Living Scale — Revised.* Los Angeles, CA: Western Psychological Services.
Zelinski, E. M., & Burnight, K. P. (1997). Sixteen-year longitudinal and time lag changes in memory and cognition in older adults. *Psychology and Aging*, *12*, 503–513.