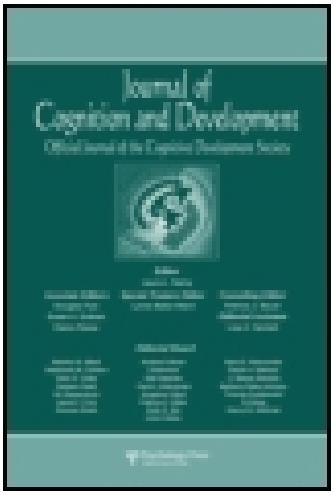

Journal of Cognition and Development
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hjcd20

A Developmental Vocabulary Assessment for Parents (DVAP): Validating Parental Report of Vocabulary Size in 2- to 7-Year-Old Children

Melissa E. Libertus[a], Darko Odic[a], Lisa Feigenson[a] & Justin Halberda[a]
[a] Johns Hopkins University

Accepted author version posted online: 12 Nov 2013.Published online: 12 Nov 2013.

To cite this article: Melissa E. Libertus, Darko Odic, Lisa Feigenson & Justin Halberda (2013): A Developmental Vocabulary Assessment for Parents (DVAP): Validating Parental Report of Vocabulary Size in 2- to 7-Year-Old Children, Journal of Cognition and Development, DOI: 10.1080/15248372.2013.835312

To link to this article: http://dx.doi.org/10.1080/15248372.2013.835312

JOURNAL OF COGNITION AND DEVELOPMENT, 0(0):1–13

ISSN: 1524-8372 print/1532-7647 online
DOI: 10.1080/15248372.2013.835312

Routledge
Taylor & Francis Group

A Developmental Vocabulary Assessment for Parents (DVAP): Validating Parental Report of Vocabulary Size in 2- to 7-Year-Old Children

Melissa E. Libertus, Darko Odic, Lisa Feigenson, and Justin Halberda

Johns Hopkins University

Measuring individual differences in children's emerging language abilities is important to researchers and clinicians alike. The 2 most widely used methods for assessing children's vocabulary both have limitations: Experimenter-administered tests are time-consuming and expensive, and parent questionnaires have only been designed for children up to 37 months of age. Here, we test the validity of a new assessment to fill this gap: the Developmental Vocabulary Assessment for Parents (DVAP). In 4 experiments, we assess the reliability of this measure and its concurrent and predictive validity in samples of 2- to 7-year-old children. We found that the DVAP provides a rapid, cost-effective alternative to experimenter-administered vocabulary tests for children.

Language learning is a major component of children's cognitive development, and individual differences in emerging language skills are important predictors for later abilities. For example, vocabulary size at 24 months is a significant predictor of later language and literacy abilities through 11 years of age (and likely beyond; Lee, 2011). Moreover, vocabulary size at 25 months predicts not only linguistic ability, but also more general cognitive skills such as IQ and working memory at 8 years of age, the oldest age tested so far (Marchman & Fernald, 2008). Hence, researchers who study language acquisition and those who study other domains of development have a shared interest in quantifying children's linguistic abilities. In the present article, we focus on measuring individual differences in 2- to 7-year-old children's expressive vocabulary as one marker of language development.

Two major types of standardized measures for vocabulary size in early childhood are experimenter-administered tests and caregiver questionnaires. Each of these has advantages and disadvantages. Experimenter-administered tests are usually individually given by a trained clinician or experimenter, and they seek to provide objective assessments of children's observable abilities (Gray, Plante, Vance, & Henrichsen, 1999). However, these tests have to be brief enough so they can be completed within the constraints of children's attention span, thus limiting the number and kinds of tasks that can be administered. Moreover, they require children to cooperate with an unknown person, which may be especially difficult for young children or those who are shy or apprehensive. Furthermore, administration of these tests takes significant blocks of time

Melissa E. Libertus is now at the University of Pittsburgh.

Correspondence should be sent to Melissa E. Libertus, Department of Psychology, Learning Research and Development Center, University of Pittsburgh, 607 LRDC, 3939 O'Hara Street, Pittsburgh, PA 15260, USA. E-mail: libertus@pitt.edu

and requires the availability of a trained experimenter. The most widely used experimenter-administered tests for measuring children's vocabulary size in English are the Peabody Picture Vocabulary Test (PPVT-4; Dunn & Dunn, 2007), the Receptive and Expressive One-Word Picture Vocabulary Tests (EOWPVT-4; Martin & Brownell, 2011a, 2011b), the Expressive Vocabulary Test (EVT-2; Williams, 2007), and the Comprehensive Receptive and Expressive Vocabulary Test (Wallace & Hammill, 2002; see Figure 1 for the age ranges that are covered by these tests). In tests that assess receptive vocabulary, an experimenter shows children an array of images and asks them to point to the image that best corresponds to a spoken word. In tests that assess expressive vocabulary, participants either are shown a picture and asked to name the depicted object or action, or are asked to define a word. These tests require about 30 min of testing time, depending on participants' age and ability, and they require visual and auditory processing along with an explicit choice or a verbal response. A few other standardized tests assess vocabulary size using similar methods, in addition to measuring other language and cognitive abilities (e.g., The New Reynell Developmental Language Scales [Edwards, Letts, & Sinka, 2011]; Test of Language Development-Fourth Edition [Newcomer & Hammill, 2008]; Bayley Scales of Infant and Toddler Development [Bayley, 2005]; see Kamphaus, Petoskey, & Rowe, 2000, for a list of further standardized language assessments).

The other method for assessing children's vocabulary, a caregiver questionnaire, takes advantage of caregivers' extensive experience with their children and can be completed at the caregivers' convenience (Dale, Bates, Reznick, & Morisset, 1989). Questionnaires do not require the presence of a trained clinician or experimenter, and thus, a large number of children can be assessed in a cost-effective fashion (Bates & Carnevale, 1993; Fenson et al., 1994; Kamphaus et al., 2000). Although caregivers may be subjective in their evaluation of their child and may differ in their ability to accurately report their children's knowledge (Stiles, 1994), Fenson and colleagues (2000) have suggested that the accuracy of caregivers' reports can be increased by

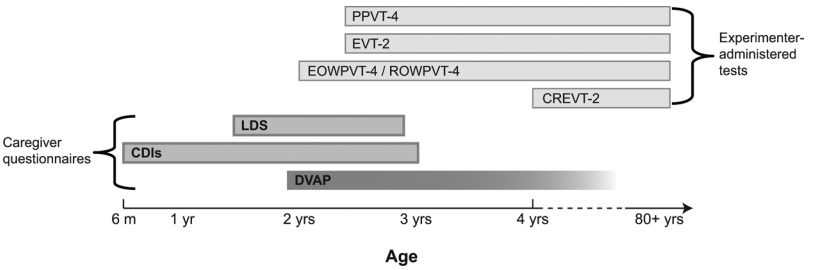

FIGURE 1 Overview of existing measures of expressive and receptive vocabulary, the current measure being investigated (DVAP), and the age ranges covered by each. Experimenter-administered tests are shown in light gray, caregiver questionnaires are shown in darker gray. The DVAP extends the age range of caregiver questionnaires to include preschool and early elementary school years (tested in the current samples) and may be relevant beyond these years as well (not tested here). PPVT-4 = Peabody Picture Vocabulary Test-Fourth Edition (Dunn & Dunn, 2007); EVT-2 = Expressive Vocabulary Test-Second Edition (Williams, 2007); EOWPVT-4 = Expressive One-Word Picture Vocabulary Test-Fourth Edition (Martin & Brownell, 2011a); ROWPVT-4 = Receptive One-Word Picture Vocabulary Test-Fourth Edition (Martin & Brownell, 2011b); CREVT-2 = Comprehensive Receptive and Expressive Vocabulary Test-Second Edition (Wallace & Hammill, 2002); LDS = Language Development Survey (Rescorla, 1989); CDIs = MacArthur-Bates Communicative Development Inventories (Fenson et al., 2007).

focusing on children's current and emerging behaviors and by using a recognition format, in which caregivers are provided with a list of words and are asked to report the child's use or understanding of each. The most widely used questionnaires to assess children's early vocabulary are the MacArthur-Bates Communicative Development Inventories (CDIs; Fenson et al., 2007) and the Language Development Survey (LDS; Rescorla, 1989). As can be seen in Figure 1, these questionnaires only provide estimates of vocabulary size for children up to 37 months of age. Some other questionnaires assess vocabulary and early language skills in older children in conjunction with other cognitive abilities (e.g., Vineland Adaptive Behavior Scales [Sparrow, Cicchetti, & Balla, 2005]; Ages & Stages Questionnaires [Squires, Bricker, Twombly, & Potter, 2009]; Child Development Inventory [Ireton, 1992]). However, due to their broad scope, these questionnaires often are not as sensitive to a single cognitive domain as a focused questionnaire would be. Thus, a caregiver questionnaire to assess vocabulary size in children 3 years and older would be a valuable measure for researchers and clinicians.

In the four experiments described in this article, we created and validated such a measure. One concern about using questionnaires to index vocabulary in older children is that caregivers may not be able to accurately report children's growing vocabulary (Dale, 1991); therefore, it is critical to ensure that the words selected are appropriate and informative about overall vocabulary size and complexity. To this end, we constructed a Developmental Vocabulary Assessment for Parents (DVAP) from a list of 212 words from the PPVT-4 (Dunn & Dunn, 2007).[1] We chose to assess children's word production rather than comprehension because previous work suggests that it is easier for caregivers to track production than comprehension in children 2 years and older (Fenson et al., 2007). In Experiment 1, we measured vocabulary size in a sample of four hundred fifty-four 2- to 7-year-old children to examine the variability in DVAP scores and to ask whether caregivers' reports of children's vocabulary increase as a function of age. In Experiments 2 and 3, we assessed the concurrent validity of the DVAP against the MacArthur-Bates CDIs (Fenson et al., 2007), a widely used questionnaire for early vocabulary size and communication skills (Experiment 2), and the PPVT-4 (Dunn & Dunn, 2007), an experimenter-administered test of receptive vocabulary (Experiment 3). Finally, in Experiment 4, we assessed the predictive validity of the DVAP during a 7-month delay.

EXPERIMENT 1: CROSS-SECTIONAL EVALUATION OF THE DVAP

Method

Participants. Caregivers of 454 children ($M_{age} = 4;3$; $SD = 12$ months; range = 2;2–9;5; 231 girls) completed the DVAP during a laboratory visit in which children participated in an unrelated study. Information about the highest level of parental education was available for 74.98% of the mothers and 68.5% of the fathers. Less than 1% of mothers and fathers did not obtain a high school diploma; 2.06% of mothers and 2.89% of fathers had obtained only a high school diploma; 16.47% of mothers and 10.93% of fathers had completed some college but had not obtained a college degree; and 80.88% of mothers and 85.21% of fathers had obtained college or advanced degrees beyond college. Information about the amount of children's exposure to English relative

[1]We obtained permission from the publisher to use these words for our research purposes.

to other languages was available for 403 children (88.77%). For 90.58% of these children, caregivers reported that English comprised at least 90% of the total language exposure. Out of the 38 remaining children, only 2 were exposed to English less than 50% of the time.

The questionnaire was always administered prior to the start of the other study or during the other study if the caregiver did not accompany the child into the testing room. Data from 16 children were excluded from the final sample because the caregiver reported that the child had cognitive deficits ($n = 1$), that the child did not speak enough English ($n = 1$), that the child was older than 7 years of age ($n = 4$), because the caregiver did not complete the entire form ($n = 2$), or the person completing the form was not a primary caregiver and therefore may not have provided an accurate assessment of the child's abilities ($n = 8$). None of the children included in the final sample were reported to have developmental delays. Prior to completion of the questionnaire, parents of all children provided informed consent based on a protocol approved by a university institutional review board.

Materials and procedure. The DVAP consists of a list of the 8 practice items and first 204 words from Form A of the PPVT-4 (Dunn & Dunn, 2007). We used the first 204 words because pilot testing using the PPVT-4 indicated that most children in our targeted age group did not understand the remaining, more difficult words. We decided to exclude these words on the DVAP to shorten the time required to complete the questionnaire. Words included nouns, verbs, and adjectives in order of increasing difficulty. Caregivers were asked to mark all and only the words in the list that they had ever heard their child say. Different pronunciations or parts of speech were considered proper use of the word and caregivers were instructed to mark the words in these cases. In addition, caregivers reported whether they were (one of) the child's main caregiver(s) and indicated the percentage of the child's total language exposure composed of English. Caregivers took about 5 to 10 min to complete the DVAP.

Results and Discussion

DVAP scores were computed as the sum of all marked words. The first eight items (i.e., the original practice items on the PPVT-4) were included to provide caregivers with a few easy words that almost all children, including the youngest ones, would say. However, these were excluded from the final DVAP score to allow for direct comparison between DVAP and PPVT-4 performance in Experiment 3, as not all of the practice items are used during standard administration of the PPVT-4. A second independent rater scored a random sample of approximately 25% of all questionnaires, and interrater reliability was extremely high ($r > .99$).

To determine the extent to which age (in days) correlated with parental report of children's vocabulary, both with and without controlling for the amount of exposure to English, we conducted two linear regressions with age (and percentage of English exposure) as possible predictors for DVAP scores. Thirty-seven caregivers did not report the amount of their child's English exposure. Data from these children were excluded from the analysis that included English exposure as a variable. All R^2 values reported in this and all the following experiments are nonadjusted.

Caregivers reported that their children said an average of 92.54 ($SD = 32.03$) of the 204 words on the DVAP and that English comprised an average of 95.64% ($SD = 11.72\%$) of children's total language exposure. There were only two children for whom English exposure was less than 50%.

TABLE 1
Correlational Results Between the Developmental Vocabulary Assessment for Parents (DVAP) Scores and Age (Experiment 1, first row), Scores on Another Caregiver Report Measure (Experiment 2, MacArthur-Bates Communicative Development Inventory [CDI]; Fenson et al., 2007) and One Experimenter-Administered Measure (Experiment 3, Peabody Picture Vocabulary Test [PPVT-4]; Dunn & Dunn, 2007)

DVAP correlation with:	*ß*	*t*	*p Value*	R^2	*F*	*p Value*
Age (Experiment 1)	.49	$t(436) = 11.70$	<.001	.24	$F(1, 436) = 136.87$	<.001
CDIs (Experiment 2)	.79	$t(61) = 10.22$	<.001	.63	$F(1, 61) = 104.34$	<.001
PPVT (Experiment 3)	.69	$t(90) = 8.93$	<.001	.47	$F(1, 90) = 79.71$	<.001

Removing their data from these analyses did not change any of the results reported here. There were no significant differences in DVAP scores between boys and girls ($M_{boys} = 91.94$, $M_{girls} = 93.11$), $t(436) = -0.38$, $p = .70$, and therefore, gender was excluded from all subsequent analyses. As can be seen in Table 1, the linear regression model with age as a possible predictor captured a significant amount of variance in DVAP scores ($R^2 = .24$), $F(1, 436) = 136.87$, $p < .001$. Adding English exposure as a further predictor increased the explained variance in DVAP scores by 2% ($R^2 = .26$), $F(2, 398) = 70.47$, $p < .001$. When controlling for the percentage of English exposure, age was still a significant predictor of DVAP scores ($ß = .47$, $p < .001$). When controlling for age, percentage of English exposure was still a small but significant predictor of DVAP scores ($ß = .17$, $p < .001$).

These results show that 2- to 7-year-old children's expressive vocabulary, measured by the DVAP, reliably increases with age. There were no apparent gender differences in expressive vocabulary, in line with a meta-analysis by Hyde and Linn (1988). Whereas other caregiver report measures such as the CDIs provide separate norms for boys and girls, these differences tend to be small and not reliable at all ages (Fenson et al., 2007).

EXPERIMENT 2: CONCURRENT VALIDITY WITH ANOTHER PARENTAL REPORT OF EXPRESSIVE VOCABULARY

To determine the concurrent validity of the DVAP, we next compared children's DVAP scores to their vocabulary scores as indexed by a well-known measure of expressive language. In Experiment 2, we asked caregivers of approximately one third of the children from Experiment 1, who were 4 years old or younger, to complete the MacArthur-Bates CDIs (Fenson et al., 2007). The CDIs measure expressive vocabulary and communication skills in 8- to 37-month-old children. To increase our sample size, we included some children who were outside of the age range for which the CDIs are normed and used raw scores instead of standardized CDI scores to compare to the DVAP scores.

Method

Participants. Caregivers of 66 children from Experiment 1 ($M_{age} = 3;4$; $SD = 4.78$ months; range = 2;2–4;0; 34 girls) completed the MacArthur-Bates CDI-III (Fenson et al., 2007) in addition to the DVAP. Information about the highest level of parental education was available

for 83.33% of the mothers and 77.27% of the fathers; 1.82% of mothers and 3.92% of fathers had obtained only a high school diploma, 12.73% of mothers and 11.76% of fathers had completed some college but had not obtained a college degree, and 85.45% of mothers and 85.13% of fathers had obtained college or advanced degrees beyond college. Information about children's percentage of exposure to English was available for 57 children (86.36%). Of these children, 89.47% were exposed to English at least 90% of the time. Of the remaining 6 children, only 1 child was exposed to English less than 50% of the time. Data from 3 children were excluded from the final sample because the person completing the form was not one of the child's main caregivers.

Materials and procedure. The CDI-III consists of a checklist of 100 words on which caregivers are asked to mark all and only the words that they have heard their child say. In addition, caregivers are asked to select between different sentences that best reflect their child's grammar and sentence complexity, and they are asked to answer questions about their child's usage of different types of words (e.g., ''Does your child ask questions with more than 1 word that begin 'what' or 'where'?''). Caregivers completed the CDI-III before the DVAP and took about 5 to 10 min to complete each questionnaire. The CDI-III and the DVAP have a total of 4 items in common, making these questionnaires largely independent.

Results and Discussion

DVAP scores were calculated as described in Experiment 1. A second independent rater scored a random sample of approximately 25% of all questionnaires, and interrater reliability was extremely high ($r > .99$). For the purposes of the present experiment, only CDI-III scores from the vocabulary checklist were included. To obtain these CDI-III scores, the total number of words that the caregiver marked was calculated. Note that we used raw CDI-III scores instead of standardized scores to allow for comparisons between DVAP and CDI-III scores and because many of the children in our sample were older than the age for which the CDI-III is normed.

To determine the extent to which caregivers' reports of children's vocabulary correlated between the two measures, we conducted a linear regression model with DVAP scores as a possible predictor for CDI-III scores. We also conducted a second linear regression model in which we added age as a further possible predictor of CDI-III scores.

Caregivers reported that their children said an average of 75.27 ($SD = 32.37$) words on the DVAP and that their children said an average of 70.48 ($SD = 25.98$) words on the CDI-III. As can be seen in Table 1, the regression model with DVAP scores as a possible predictor captured a significant amount of variance in children's CDI-III scores ($R^2 = .63$), $F(1, 61) = 104.34$, $p < .001$. Adding age as a further predictor increased the amount of variance captured in the model by 2% ($R^2 = .65$), $F(2, 60) = 54.76$, $p < .001$. Most importantly, DVAP scores explained a significant amount of variance in children's CDI-III scores over and above age ($\beta = .77$, $p < .001$), whereas age did not explain a significant amount of variance in children's CDI-III scores over and above DVAP scores ($\beta = .12$, $p = .12$). A median split by age revealed a significant correlation between DVAP scores and CDI-III scores for both younger ($R^2 = .63$, $p < .001$) and older children ($R^2 = .58$, $p < .001$). Fisher's r-to-z transformation indicated no significant differences in the correlations of these two age groups ($z = .31$, $p = .76$).

The findings of Experiment 2 show that the DVAP is a valid measure for indexing children's expressive vocabulary. As a caveat to this general finding, we also note that as children got older

and caregivers reported that children's knowledge of the queried words on the CDI-III reached ceiling (maximum score of 100), variability in DVAP scores increased. This presumably was because the DVAP contains a larger list of more advanced words (thus making it more applicable for a wider age range). Nevertheless, we found significant correlations between DVAP and CDI-III scores for both younger and older children.

EXPERIMENT 3: CONCURRENT VALIDITY WITH AN EXPERIMENTER-ADMINISTERED MEASURE OF RECEPTIVE VOCABULARY

In Experiment 2, we found significant concurrent agreement between the DVAP and another measure that uses caregiver reports to measure children's expressive vocabulary. However, as noted earlier, caregivers' reports can be limited by underestimation or overestimation of children's abilities. Therefore, to further ascertain the concurrent validity of the DVAP, we tested a subset of children from Experiment 1 on the PPVT-4 (Dunn & Dunn, 2007), an experimenter-administered standardized test of receptive vocabulary. This allowed us to investigate the extent to which a) children's expressive vocabulary as measured by the DVAP correlated with their comprehension as measured by the PPVT-4, and b) caregivers' reports correlated with children's laboratory performance.

Method

Participants. A subset of 101 children from Experiment 1 ($M_{age} = 4;3$; $SD = 12$ months; range = 2;10–6;2; 52 girls) participated in Experiment 3. Most of these children came into the laboratory to participate in an unrelated study that was brief enough to allow children to be tested on the PPVT-4 during the same visit. Whether these children were tested on the PPVT-4 was based on age and the availability of a trained experimenter. Information about the highest level of parental education was available for 81.19% of the mothers and 73.27% of the fathers. None of the mothers and 1.35% of fathers had not obtained a high school diploma, 2.44% of mothers and 2.70% of fathers had only obtained a high school diploma, 15.85% of mothers and 10.81% of fathers had completed some college but had not obtained a college degree, and 81.71% of mothers and 85.14% of fathers had obtained college or advanced degrees beyond college. Information about children's percentage of exposure to English was available for 84 children (83.17%). Of these, 88.09% were exposed to English at least 90% of the time. Out of the 10 remaining children, only 2 were exposed to English less than 50% of the time. All children were able to hear sufficiently well to participate in the present experiment. Data from 9 children were excluded from the final sample because the DVAP was not completed by a primary caregiver ($n = 3$), because children failed to complete the experimenter-administered portion of the experiment ($n = 4$), because of parental interference ($n = 1$), or because the child was older than 7 years ($n = 1$).

Materials and procedure. The PPVT-4 (Dunn & Dunn, 2007) consists of a picture book of four colored drawings per page. For each page, the experimenter reads a word aloud and the child is asked to point to the picture that best matches the meaning of the read word. Children were tested on Form A of the PPVT-4, following the official testing protocol. Briefly, the experimenter explained the task using one of two practice pages depending on the child's age. Once the child

responded correctly to 2 practice items, actual testing started at an age-defined entry point and continued until the child missed 8 or more words in a set of 12 words. If the child missed more than 1 word on the first set of 12 words, the experimenter administered easier sets of 12 items until the child made no more than one error. The experimenter generally praised the child but did not provide any feedback as to the correctness of the answers.

Caregivers completed the DVAP either before PPVT-4 testing or while an experimenter administered the PPVT-4 to the child if the caregiver chose not to accompany the child into the testing room. This was done to ensure that caregivers' reports were not biased by observing children's responses on the PPVT-4. Caregivers took about 5 to 10 min to complete the DVAP and children took about 20 min to complete the PPVT-4.

Results and Discussion

DVAP scores were calculated as described in Experiment 1. To obtain PPVT-4 scores, we calculated the sum of all words correctly identified by each child. Practice items were excluded from these scores. Note that we used raw PPVT-4 scores and not standardized scores to allow for comparisons between DVAP and PPVT-4 scores. A second independent rater scored a random sample of approximately 25% of all tests and interrater reliability was perfect ($r = 1.0$). The average PPVT-4 standard score for children in Experiment 3 was 116.16 ($SD = 14.74$), suggesting that our sample performed about 1 standard deviation above the typical performance. Note that previously reported PPVT-4 scores were obtained using a sample of children who were likely from a wider range of socioeconomic backgrounds and parent educational achievement compared with the sample of children in our study (Dunn & Dunn, 2007).

To determine the extent to which parental report of children's vocabulary on the DVAP correlated with children's performance on the PPVT-4, we conducted a linear regression model with DVAP score as a possible predictor for PPVT-4 scores. We also conducted a second linear regression model in which we added age as a further possible predictor of PPVT-4 scores. Lastly, to assess whether the validity of caregivers' reports decreases as children get older (as suggested by Feldman and colleagues [2005]), we split the data by median age and correlated DVAP and PPVT-4 scores separately for younger ($M_{age} = 3;4$) and older children ($M_{age} = 5;1$).

Caregivers reported that their children said an average of 93.12 ($SD = 34.84$) words on the DVAP. We found that children knew an average of 85.30 ($SD = 27.52$) words on the PPVT-4. As can be seen in Table 1, the regression model with DVAP scores as a possible predictor captured a significant amount of variance in children's PPVT-4 scores ($R^2 = .47$), $F(1, 90) = 79.71$, $p < .001$. Adding age as a further predictor increased the amount of variance in children's PPVT-4 scores captured in the model by 14% ($R^2 = .61$), $F(2, 89) = 69.98$, $p < .001$. Most importantly, DVAP scores explained significant variance in children's PPVT-4 scores over and above any age effects ($\beta = .42$, $p < .001$). Age also explained a significant amount of variance in children's PPVT-4 scores over and above DVAP scores ($\beta = .46$, $p < .001$). Finally, correlations between DVAP and PPVT-4 scores were significant both for younger ($R^2 = .35$), $F(1, 45) = 24.69$, $p < .001$, and older children ($R^2 = .21$), $F(1, 43) = 11.08$, $p < .01$. The amount of variance in PPVT-4 scores that was explained by DVAP scores decreased with age, suggesting that caregivers' ability to accurately report their children's expressive vocabulary may decrease as children get older. However, the correlation coefficients between DVAP and PPVT-4 scores were

not statistically different for the younger versus the older children in this sample ($p = .36$, two-tailed, using Fisher's r-to-z transformation).

These findings show that caregivers' reports of children's expressive vocabulary on the DVAP correlate with children's receptive vocabulary skills on a widely used experimenter-administered test. Our results support the concurrent validity of the DVAP as an assessment for early vocabulary size and show that there is a close association between expressive vocabulary as measured by the DVAP and receptive vocabulary as measured by the PPVT-4. Even though the correlation between DVAP and PPVT-4 scores decreased with age, it was significant even in the older age group. Thus, caregivers' reports are a reliable measure of children's vocabulary size across the entire age range tested in this study.

EXPERIMENT 4: STABILITY AND PREDICTABILITY OF DVAP SCORES

Previous investigations of early vocabulary development have shown that parental report *predicts* later language development (Camaioni, Castelli, Longobardi, & Volterra, 1991; Feldman et al., 2005). Therefore, in our final experiment, we assessed the predictive validity of the DVAP across a 7-month time span and also assessed whether DVAP scores were a significant predictor for PPVT-4 scores 7 months later. To this end, we measured children's vocabulary size twice using the DVAP and PPVT-4 at both times, separated by a 7-month delay.

Method

Participants. A subset of 47 children from Experiment 3 ($M_{age} = 4;11$; $SD = 12$ months; range = 3;6–6;6; 22 girls) participated in Experiment 4. The average delay between testing was 7.34 months (range = 4.7–11.73 months). DVAP scores from 7 children were excluded from the final sample because the DVAP was not completed by a primary caregiver ($n = 3$), because the DVAP was completed incorrectly ($n = 1$), because the DVAP form was lost ($n = 1$), or because an earlier version of the DVAP was mistakenly administered ($n = 2$). PPVT-4 data at Time 1 from 1 child were excluded from the final sample due to external interference.

Materials and procedure. All materials and procedures were identical to those in Experiment 3. DVAP and PPVT-4 scores were calculated as described in Experiment 3.

Results and Discussion

A second independent rater scored a random sample of approximately 25% of all tests and interrater reliability was extremely high ($r > .99$). DVAP scores showed that children said an average of 97.59 ($SD = 32.00$) of the tested words at Time 1 and 105.77 ($SD = 28.15$) words at Time 2. PPVT-4 scores showed that children knew an average of 86.30 ($SD = 29.35$) of the tested words at Time 1 and 97.89 ($SD = 30.48$) words at Time 2.

To determine the predictive validity of the DVAP while controlling for the delay between testing sessions, we conducted a linear regression with DVAP scores at Time 1 and delay between testing sessions as possible predictors of DVAP scores at Time 2. This model captured a significant amount of variance in children's DVAP scores at Time 2 ($R^2 = .48$), $F(2, 37) = 16.72$,

$p < .001$. Most importantly, DVAP scores at Time 1 explained a significant amount of variance in children's DVAP scores at Time 2, over and above any effects of delay ($\beta = .69$, $p < .001$), whereas delay did not explain a significant amount of variance in children's DVAP scores at Time 2 over and above DVAP scores at Time 1 ($\beta = -.14$, $p = .25$). The higher a child's DVAP score at Time 1, the higher their DVAP score at Time 2. Thus, the DVAP appears to have good predictive validity after a delay of 7 months.

However, it is possible that the correlation between DVAP scores at Time 1 and Time 2 primarily reflects aspects of caregivers' reporting (e.g., some caregivers may have consistently underestimated or overestimated their children's vocabulary). Therefore, to assess the DVAP's ability to predict children's receptive vocabulary in an experimenter-administered task 7 months later, we correlated DVAP scores at Time 1 with PPVT-4 scores at Time 2. We found a significant positive correlation ($R^2 = .28$), $p < .001$, suggesting that expressive vocabulary as measured by caregivers' reports on the DVAP is a valid predictor of receptive vocabulary as measured in an experimenter-administered test 7 months later.

Finally, to assess the unique contributions of expressive and receptive vocabulary scores in predicting later expressive vocabulary scores, we conducted a linear regression model with DVAP and PPVT-4 scores at Time 1 as possible predictors for DVAP scores at Time 2. We found that DVAP and PPVT-4 scores at Time 1 captured significant variance in children's DVAP scores at Time 2 ($R^2 = .47$), $F(2, 36) = 15.76$, $p < .001$. Importantly, DVAP scores at Time 1 were a significant, unique predictor of DVAP scores at Time 2 (when controlling for PPVT-4 scores at Time 1; DVAP, $\beta = .76$, $p < .001$), but PPVT-4 scores at Time 1 were not a significant unique predictor of DVAP scores at Time 2 (when controlling for DVAP scores at Time 1; PPVT-4, $\beta = -.14$, $p = .35$).

In sum, the findings of Experiment 4 reveal the predictive validity of the DVAP by showing a strong correlation between DVAP scores during a 7-month delay. Furthermore, our findings show that the DVAP predicts PPVT-4 scores 7 months later. Finally, receptive vocabulary scores did not contribute uniquely to the predictive power of expressive vocabulary for later expressive vocabulary scores.

GENERAL DISCUSSION

The four experiments we presented here show that caregiver questionnaires such as the DVAP can provide reliable, valid, and inexpensive measures of 2- to 7-year-old children's expressive vocabulary. The DVAP's concurrent validity was established in two separate experiments using one caregiver report measure (Experiment 2, MacArthur-Bates CDI; Fenson et al., 2007) and one experimenter-administered measure (Experiment 3, PPVT-4; Dunn & Dunn, 2007). We found that DVAP scores were highly correlated with scores on the CDIs and the PPVT-4. Moreover, we established the DVAP's predictive validity by showing a significant correlation between early DVAP scores and DVAP scores 7 months later. Finally, DVAP scores reliably predicted PPVT-4 scores 7 months later. Questionnaires such as the DVAP thus offer a way to measure children's vocabulary that avoids lengthy and labor-intensive testing. The DVAP has the additional advantage of covering a wider age range than previous questionnaires such as the CDIs or the LDS (Rescorla, 1989).

Our study underscores and extends existing work on the reliability of caregivers' reports of children's language abilities. Previously, Reznick and Goldsmith (1989) found that caregiver

report is reliable across five different word lists. Dale and colleagues (1989) showed that caregiver reports on the Early Language Inventory at 20 months correlated with children's performance on the language subscale of the Bayley Scales of Infant Development, and Ring and Fenson (2000) showed that caregiver report on the MacArthur CDIs between 20 and 30 months correlates with children's performance on an expressive vocabulary task. Furthermore, caregiver report on the ''MacArthur CDI: Toddlers'' correlates with children's performance on several experimenter-administered, standardized language tasks at 24 months (e.g., EOWPVT-4, Memory of Sentences subtest of the Stanford-Binet Intelligence Scale; Dale, 1991). Finally, Rescorla and Alley (2001) showed that caregiver report on the LDS correlates with children's scores on the Reynell Receptive and Expressive Language Scales, Bayley Mental Development Index, and Vineland Adaptive Behavior Composite at 24 to 26 months. Even though Dale (1991) claimed that caregivers may be unable to accurately report children's vocabulary beyond 30 months of age, here we have shown that caregivers' reports are a reliable measure of children's vocabulary size until at least 7 years of age (the highest age we assessed here). Our results suggest that reliability remains significant even into the early elementary school years.

In addition to its concurrent validity, caregiver report appears to be *predictive* of later language development. For example, Camaioni and colleagues (1991) showed that caregiver reports at 12 months predict language ability at 20 months, and Feldman and colleagues (2005) showed that scores on the CDIs at 2 and 3 years of age correlate with measures of cognitive development (McCarthy Scales of Children's Abilities; McCarthy, 1972), receptive language (PPVT-R; Dunn & Dunn, 1981), and measures from parent–child conversation at 3 years of age. In the present experiments, we found that caregiver report on the DVAP when children were approximately 4 years of age predicted their receptive vocabulary scores on the PPVT-4 (Dunn & Dunn, 2007) 7 months later.

Despite the multiple successful validation studies presented here, clinicians and experimenters should be cautious when using questionnaires such as the DVAP for clinical or research purposes. Additional validation studies with other lists of words and other experimenter-administered standardized tests of expressive language (such as the EVT-2 [Williams, 2007] or the EOWPVT-4 [Martin & Brownell, 2011a]) are necessary to fully validate caregiver questionnaires as adequate screening tools. Most importantly, larger sample sizes across a wider range of socioeconomic backgrounds and cognitive abilities than those observed in this study are needed to shed light on the validity of caregiver reports. For example, Roberts, Burchinal, and Durham (1999) reported that caregivers, especially in low-income families, may underestimate their children's language abilities, suggesting that caregiver report should be complemented with other assessment tools such as experimenter-administered tests. In contrast, Reese and Read (2000) found that the predictive validity of the New Zealand version of the CDI for the PPVT is higher for children whose mothers have received less education. Finally, in the present study, we did not assess the sensitivity or specificity of the DVAP for detecting language delays. Nonetheless, the DVAP may provide a useful measure for researchers seeking a way to rapidly assess children's vocabulary.

ACKNOWLEDGMENTS

We thank Sean Ostro, Meg Bowen, Andrea Stevenson, Selin Zeytinoglu, and Anna Zamm for help with data collection.

FUNDING

This work was supported by the National Institute of Child Health and Human Development Grant #R01 HD057258 to LF and JH.

REFERENCES

Bates, E., & Carnevale, G. F. (1993). New directions in research on language development. *Developmental Review*, *13*, 436–470.

Bayley, N. (2005). *Bayley Scales of Infant and Toddler Development* (3rd ed.). San Antonio, TX: Pearson.

Camaioni, L., Castelli, M. C., Longobardi, E., & Volterra, V. (1991). A parent report instrument for early language assessment. *First Language*, *11*, 345–359.

Dale, P. S. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech and Hearing Research*, *34*, 565–571.

Dale, P. S., Bates, E., Reznick, J. S., & Morisset, C. (1989). The validity of a parent report instrument of child language at twenty months. *Journal of Child Language*, *16*, 239–249.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test-Revised*. Circle Pines, MN: American Guidance Service.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (4th ed.). San Antonio, TX: Pearson.

Edwards, S., Letts, C., & Sinka, I. (2011). *The New Reynell Developmental Language Scales*. London, England: GL Assessment.

Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., & Paradise, J. L. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development*, *76*, 856–868. doi:10.1111/j.1467–8624.2005.00882.x

Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, J. S., & Thal, D. (2000). Measuring variability in early child language: Don't shoot the messenger. *Child Development*, *71*, 323–328.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*, 1–173.

Fenson, L., Marchman, V., Dale, P., Reznick, J. S., Thal, D., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories* (2nd ed.). Baltimore, MD: Brookes.

Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Language, Speech, and Hearing Services in Schools*, *30*, 196–206.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, *104*, 53–69.

Ireton, H. (1992). *Child Development Inventory*. Minneapolis, MN: Behavior Science Systems.

Kamphaus, R. W., Petoskey, M. D., & Rowe, E. W. (2000). Current trends in psychological testing of children. *Professional Psychology: Research and Practice*, *31*, 155–164.

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics*, *32*, 69–92.

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*, F9–16. doi:10.1111/j.1467–7687.2008.00671.x

Martin, N. A., & Brownell, R. (2011a). *Expressive One-Word Picture Vocabulary Test* (4th ed.). Novato, CA: Academic Therapy.

Martin, N. A., & Brownell, R. (2011b). *Receptive One-Word Picture Vocabulary Test* (4th ed.). Novarto, CA: Academic Therapy.

McCarthy, D. (1972). *McCarthy Scales of Children's Abilities*. San Antonio, TX: Psychological Corporation.

Newcomer, P. L., & Hammill, D. D. (2008). *Test of Language Development* (4th ed.). Austin, TX: Pro-Ed.

Reese, E., & Read, S. (2000). Predictive validity of the New Zealand MacArthur Communicative Development Inventory: Words and Sentences. *Journal of Child Language*, *27*, 255–266.

Rescorla, L. (1989). The Language Development Survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders*, *54*, 587–599.

Rescorla, L., & Alley, A. (2001). Validation of the Language Development Survey (LDS): A parent report tool for identifying language delay in toddlers. *Journal of Speech, Language and Hearing Research*, *44*, 434–445.

Reznick, J. S., & Goldsmith, L. (1989). A multiple form word production checklist for assessing early language. *Journal of Child Language*, *16*, 91–100.

Ring, E. D., & Fenson, L. (2000). The correspondence between parent report and child performance for receptive and expressive vocabulary beyond infancy. *First Language*, *20*, 141–159.

Roberts, J. E., Burchinal, M., & Durham, M. (1999). Parents' report of vocabulary and grammatical development of African American preschoolers: Child and environmental associations. *Child Development*, *70*, 92–106.

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales* (2nd ed.). San Antonio, TX: Pearson.

Squires, J., Bricker, D. D., Twombly, E., & Potter, L. (2009). *Ages & Stages Questionnaires: A parent-completed child-monitoring system* (3rd ed.). Baltimore, MD: Brookes.

Stiles, J. (1994). On the nature of informant judgments in inventory measures:. . . And so what is it you want to know? *Monographs of the Society for Research in Child Development*, *59*, 180–185.

Wallace, G., & Hammill, D. D. (2002). *Comprehensive Receptive and Expressive Vocabulary Test* (2nd ed.). Austin, TX: Pro-Ed.

Williams, K. T. (2007). *Expressive Vocabulary Test* (2nd ed.). San Antonio, TX.: Pearson.