Measures of Early Vocabulary and Gestural Typicality

Philip S. Dale[1], Matthew Nichols, Rhiannon Luyster[2], & Nicole Harlaar[3]

[1]University of New Mexico

[2] Emerson College

[3] University of Colorado, Boulder

Address for correspondence:

Philip S. Dale
Speech & Hearing Sciences
University of New Mexico
1700 Lomas Blvd NE
Albuquerque, NM 87131
email: dalep@unm.edu

Please cite, consistent with APA/6 Publication Manual, as:

Dale, P. S., Nichols, M., Luyster, R., & Harlaar, N. (2012). Measures of early gestural and gestural typicality. Unpublished manuscript, Department of Speech & Hearing Sciences, University of New Mexico, Albuquerque, NM.

Abstract

Although discussions of stylistic differences in early language development often focus on vocabulary composition, little work has been done on quantifying this variation, other than with respect to the proportion of nominals in early vocabulary. In this paper we introduce a new measure of vocabulary typicality, independent of developmental level, and examine its psychometric properties, association with the clinical condition of autism, and etiology. An analogous measure for early gestural typicality is also introduced. The measures are first developed and normed on the norming sample for the MacArthur-Bates Communicative Development Inventories. Although independent of total vocabulary level, the typicality measurs are generally higher for girls than for boys, and CDI:WS Vocabulary Produced typicality is positively related to rate of development. In a second study of 116 children with Autism Spectrum Disorders, mean CDI:WG Words Produced and Gestures Produced typicality measures are significantly below norms for typically developing children. In a third, twin study of 2929 2- and 3-year old pairs, a shortform Words Produced measure was found to have significant genetic and even larger shared environmental influence. These first findings are encouraging with respect to the validity and utility of measures of vocabulary and gestural typicality.

Introduction

Early language development is remarkable both for the commonalities observed across children and across languages, and for the differences observed among children (Bates, Dale & Thal, 1995). In addition to differences in rate of development (Fenson, Marchman, Thal, Dale, Reznick & Bates, 2007), several well-defined dimensions of qualitative differences have been identified (Shore, 1995). The focus of the present project is the most intensively studied of these, vocabulary composition. This line of research began with Nelson's (1973) study of early vocabulary development, in which the balance of nominal versus non-nominal words was used to characterize referential versus expressive vocabularies. Numerous other studies (see Shore, 1995, for a review) have examined this and related measures, particularly the proportion of nominals but also predicates and closed class words, in early vocabularies. Correlations of these measures with aspects of the linguistic environment, and with other dimensions of verbal and nonverbal development illustrate the potential value of such measures for investigating proposed mechanisms of language development (Shore, 1995).

Few of the dimensions of individual differences have been well-quantified, with established psychometric properties, though characterization of the distribution of any measure is essential for researching influences on it. The proportion of nominals has been the most often investigated. Bates, Marchman, Thal, Fenson, Dale, Reznick, Reilly and Hartung (1994) developed norms for the proportion of common nouns (on the MacArthur-Bates CDIs) relative to total vocabulary size, that is, a measure of nominal style uncorrelated with overall developmental level. These norms implicitly define a typical range for proportion of nouns at a given vocabulary level, and extreme values in both directions. Several researchers (see Bates et al., 1995;

Charman, Drew, Baird & Baird, 2003) have investigated this and other measures of lexical composition across clinical classifications such as Down Syndrome and focal brain lesion. However, these studies have all operated at the level of predefined categories such as nouns, predicates (verbs and adjectives), and closed-class words.

In the present study, we extend this approach to studying variation in vocabulary composition beyond the proportion of nominals and other broad categories, in two directions. First, at a micro-level, we assess the typicality of individual lexical items at each developmental level without assuming *a priori* the relevance of any particular set of categories. Second, at a macro-level, our goal is to evaluate the typicality of a child's early vocabulary as a whole relative to other vocabularies at that developmental level. This second goal might be thought of as an assessment of the horizontal dimension of vocabulary development, independent of the vertical level. (Note that both types of assessment are based on word types, not frequency of use.) Both types of assessment are potentially useful for exploring correlates of specific aspects of development in the environment and other aspects of the child=s development, behavior, or clinical classification. Typical vocabulary composition is the result of multiple factors of input, conceptual development, and language learning skills and preferences. Deviations from typicality are likely to highlight the influence of specific factors.

The starting point for this research is the norming dataset for the MacArthur-Bates Communicative Development Inventories (CDIs). These parent report measures provide a relatively comprehensive assessment of vocabulary. For children between 8 and 18 months, they provide measures for both receptive and expressive vocabulary, which is highly desirable given the evidence that vocabulary composition may differ somewhat in these two domains (Benedict,

1979). The CDI:Words & Gestures also provides an assessment of early gestures, which appear to serve an important bridging function into language (Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994). Although the issue of individual differences in patterns of gestural development has not been previously studied, similar questions may be asked in this domain. The list of early gestures included in the CDI:WG is heterogenous with respect to cognitive and social requirements, and variations, and for this reason children=s inventories of gestures may vary as a function of personality and/or clinical classification. For example, Charman, Drew, Baird and Baird (2003) noted that pointing and showing in children with autism were delayed relative to actions with objects and imitation of adult actions. For older children (16 to 30 months), with larger vocabularies, the CDIs provide a measure of expressive vocabulary only.

STUDY 1

Construction and Psychometric Properties of the Measure

The first step in this research program was to define the typicality measure for individual words, and for the vocabulary as a whole, and establish the basic psychometric properties particularly for the latter. For this purpose, a norming sample was required. We used the updated (Fenson et al., 2007) norming sample for the MacArthur-Bates Communicative Development Inventories. These include two instruments. The first is the CDI: Words and Gestures (CDI:WG) for children between 8 and 18 months. It yields as main measures a Gestural Production score (max = 63) and Words Produced and Words Understood scores (max = 396 for each). The second is the CDI: Words and Sentences (CDI:WS) for children between 16 and 30 months. It includes several measures of grammatical development, but for the present project, only the

Words Produced score (max = 680) was used (the CDI:WS does not include a Words Understood measure).

METHOD

Participants

The sample is more fully described in Fenson et al. (2007). Following exclusions for major medical/genetic reasons, and for families in which English was not the primary language, a total of 1089 CDI:WG forms (544 girls, 545 boys) were available for 8-18 month olds, and a total of 1461 CDI:WS forms (728 girls, 733 boys) for 16-30 month olds. It should be noted that the updated Fenson et al. (2007) sample utilized in the present study is more representative of the US population than original Fenson et al. (1994) sample with respect to ethnicity and maternal education, though it continues to be somewhat above average for maternal education relative to the 2000 US Census figures.

Measures

Four typicality measures were computed, three for measures based on the CDI:WG (Gestures Produced, Words Understood, and Words Produced), and one for the CDI:WS (Words Produced). The basic method was the same in each case. First, we used item-level data from the norming sample to compute a word typicality matrix WordTyp($word_i$,$voctot_j$), that is, the probability of each item i being checked for each possible vocabulary total score j (other than 0, or scale maximum). Because the data are cross-sectional, and the number of children with each possible total score varies greatly (from zero to a large value), we smoothed the typicality curves for each individual word by fitting a cubic equation predicting the probability of that word, i.e., its typicality, as a function of the total vocabulary size. As well as smoothing the curve and

removing nonmonotonicities, this step permitted the determination of WordTyp scores for future data in which a vocabulary total occurred which did did not occur in the norming data. As Fenson et al. (1994) noted, the cubic function fits the developmental trends well.

In the next step, we computed the mean typicality VocTyp for the CDI vocabulary (or gestures) as a whole, that is, all items checked positively for each child on that scale. However, this first estimate of scale-level typicality is not useful, because it is substantially correlated with the total score on the measure, whereas the potential applications for a typicality measure require that it is uncorrelated with total score, that is developmental level. The correlation results from the fact that as more words have been learned, there are consequently fewer 'degrees of freedom' for each word, and therefore words are more highly typical. For example, consider a child with just one word checked on the CDI:WS vocabulary list: there are 680 possibilities, most of low probability. Contrast this vocabulary with that of a child with 679 words checked - almost all of them must have a high frequency of occurring. Empirically, the correlations of mean typicality with total scale score are r = .258, .109, .242, and .978 for CDI:WG Gesture Production, CDI:WG Words Understood, CDI:WG Words Produced, and CDI:WS Words Produced, respectively.

In order to remove the correlation with total vocabulary size, we empirically standardized the distribution of typicality scores, that is, determined the mean and standard deviation for each level of total vocabulary development. Because there may be few, or no, children with a specific total scale score, we classified children into 'bins' for each total scale score, and then determined the mean and standard deviation of the typicality measure within each bin. For CDI:WG Gestures Produced, the bins were 1-5, 6-10, ... 56-60, 61-63. For CDI:WG Words Understood,

the bins were 1-5, 6-10, ... 96-100, 101-110, ... 191-200, 201-220, ... 361-380, 381-396. For CDI:WG Words Produced, the bins were 1-5, 6-10, ... 96-100, 101-110, ... 191-200, 201-240, ... 280-320, 321-396. For CDI:WS Words produced, the bins were 1-5, 6-10, ... 96-100, 101-110, ... 191-200, 201-220, ... 661-680. The bin sizes were selected to be as narrow as possible (to minimize the effect of the correlation of typicality with total size) while maintaining an adequate sample within the bin (> 20 in almost all cases) for a robust estimate of mean and standard deviation. Then each child=s typicality score was converted to a z-score based on the mean and standard deviation of the typicality scores for children in that bin. Extreme scores were trimmed to a maximum of +/- 3 SD.

The computer program that was developed to perform these analyses also identified individual items for each child which were exceptional by their presence (checked by parents with probability less than .1 in the norming sample at this developmental level, i.e., WordTyp < .1) or by their absence (checked by parents with probability greater than .9 in the norming sample at this developmental level). The probability cutoffs for these purposes are user-selectable.

RESULTS

Table 1 presents a comparison of three children at the same total vocabulary size on CDI:WG Words Produced (12 words) at three levels of VocTyp: average, somewhat atypical, and highly atypical. For child I114200 in particular, both nominals and non-nominals characteristic of later stages are present in the vocabulary.

Table 2 presents the means and standard deviations for the four typicality measures for the sample as a whole, and divided by gender. Due to the small numbers of children at each total score value and the consequent need for binning, the means and standard deviations are not

precisely 0 and 1, respectively, as would be expected from the empirical standardization described above, but they are close. Figure 1 presents a representative histogram, for CDI:WS Words Produced. The distribution is close to normal, but negatively skewed (-.841), with evidence for a subset of children with extremely low typicality scores, which have been trimmed to -3.00 in the typicality calculation. As discussed above, the typicality measures have near-zero correlations with scale total scores, as desired.

Table 1 also presents the means and standard deviations for girls and boys separately. For the three vocabulary measures, typicality is consistently higher for girls than for boys. However, the effect sizes for the gender differences were small.

Although the measures have been constructed specifically to eliminate correlation with overall level of development, it is also possible that there is a relationship between rate of development and typicality, holding constant level of development. For example, faster developing children might have higher or lower typicality scores at a specific developmental level than children developing more slowly. To evaluate this possible, we conducted partial correlation analyses between age and typicality scores, controlling for total vocabulary (or gestures). For CDI:WS Words Produced, the partial correlation was negative and significant, though modest (r = -.171, df = 1458, $p < .001$). In other words, faster children B children who are younger at a given developmental level - have higher typicality scores than slower children, that is, older children at that developmental level. For the CDI:WG measures, however, the results were weaker and inconsistent in direction, though still significant (Words Understood: r = .087, p = .004; Words Produced: r = -.062, p = .04; Gestures Produced: r = .076, p = .012; df = 1086 for all correlations).

DISCUSSION

The four typicality measures have the desired property of little or no correlation with developmental level and relatively normal distribution, albeit with some negative skew. Although typicality is not correlated with level of development, it is positively correlated with rate of development. This finding is somewhat unexpected; it might have been hypothesized that children developing more rapidly are exposed to a linguistically more diverse environment and therefore show less typicality. We return to this result in the conclusions of the paper.

There is a consistent finding across the vocabulary measures of lower typicality, i.e. greater variability in lexical composition, for males than for females. This does not hold for gestures, where there is a nonsignificant trend in the opposite direction. It is possible that the sex difference in vocabulary typicality is related to the well-established but small sex difference favoring girls and the positive relationship of rate of development to typicality, at least for vocabulary production on both the CDI:WG and CDI:WS. It is also possible that boys are simply more variable than girls. If the subsamples of boys and girls showed equivalent degrees of similarity across words, but differed in their central tendencies (the >most typical= vocabulary), the relative equality of sample sizes would have lead to equivalent mean typicality scores. Thus, we conclude that the lower mean typicality scores for males reflect greater variability within that subsample.

A logical next step would be to identify the correlates of typicality, particularly very low scores. However, in the present sample, very limited information was available about the participants. In the next two studies, with different samples, we will examine the reliability and validity of the typicality measures; first, by exploring the sensitivity of the measures to the

clinical classification of autism spectrum disorders, and second, by examining their relationship to genetic and environmental factors, that is, their etiology.

STUDY 2

Clinical Sensitivity to Autism Spectrum Disorder

Delay, differences, and even the absence of language comprise one of the three main indicators of autism and the related autism spectrum disorders (ASD) of Pervasive Developmental Disorder - Not Otherwise Specified (PDD-NOS) and Asperger Disorder. As a parent report measure which does not require child compliance or other skills inherent in structured testing, the CDIs have proven valid and useful tools for the assessment of children with ASD (Charman et al., 2003; Luyster, Lopez & Lord, 2007).

To a greater extent than seen in most forms of developmental disorders, autism spectrum disorders are likely to manifest themselves in qualitatively different profiles of language development. Because of these differences, matching of children to younger, typically developing children is not easily done; and more generally, tests for younger typically developing children are also not fully satisfactory for children with ASD. Among the qualitative differences that have been noted are (1) a relative delay for gestures which have a conventional, communicative function such as pointing and showing relative to actions with objects and imitating adult actions (Charman et al., 2003); (2) a greater delay for pragmatics than for vocabulary and grammar (Tager-Flusberg, Paul & Lord, 2005); (3) a relatively smaller discrepancy between receptive and expressive vocabulary than for typically developing children (Charman et al., 2003; Luyster et al., 2007); (4) relative delay for words for body parts, games and routine words in receptive vocabulary; (5) relative delay for sound effects, animal names, toys, and deictic terms (Charman

et al., 2003; Tager-Flusberg, 1994), the last also manifested in pronoun reversal errors (Lee, Hobson & Chiat, 1994); and (6) evidence based on intrasubtest scatter analysis that children with autism are less likely to follow the developmental sequence of items on the Index of Productive Syntax than typically developing children or developmentally delayed children matched for receptive vocabulary (Eigsti, Bennetto & Dadlani, 2006). However, two systematic studies of vocabulary composition utilizing the Bates et al. (1994) broad categories of nominals, predicates, and closed class items failed to find group difference in the proportions of these categories for children with ASD (Charman et al., 2003; Luyster et al., 2007).

Because one of the major motivations for the development of a typicality measure was to focus on individual items rather than pre-defined broad categories, the early language of children with ASD, often viewed as idiosyncratic and not typical in nature, is an especially appropriate context for evaluating the usefulness of the new measure. In addition, the analysis procedure provides not only a global measure of typicality; it also generates a list of words and gestures which are unusual by their presence or absence at a given overall level of Words Understood and Words Produced, relative to the norming sample for the CDIs. Finally, the present data made possible a preliminary examination of the validity of the measure for sub-typing within ASD by comparing the children with classical autism with those with other forms of ASD, such as PDD-NOS or Asperger Disorder.

METHODS

Participants

The sample for the present study is a subsample of 116 of the 164 toddlers with autism spectrum disorders investigated in Luyster, Kadlec, Carter, and Tager-Flusberg (2008). Participants

included 32 females and 84 males between the ages of 18 and 33 months, recruited through collaboration with early intervention providers in the state of Massachusetts. Within the present sample, 59 participants were assigned a diagnosis of autism (AUT) and 57 were assigned a classification of the broader autism spectrum disorder (ASD) because they did not meet strict criteria for autism but showed a profile consistent with the broader category, a category that often includes schildren who are diagnosed with PDD-NOS or Asperger Disorder. These classifications were made utilizing standard techniques including the Autism Diagnostic Interview B Revised (Lord, Rutter & Le Couteur, 1994) and the Autism Diagnostic Observational Scheduled-Generic (Lord, Risi, Lambrecht, Cook, Leventhal, DiLavore, Pickles & Rutter, 2000). Demographic information on the larger sample is provided in Luyster et al., but due to the anonymous nature of the data shared with the present project which included only gender, age, and diagnostic classification, comparable figures are not available for the present sample. The 116 participants in the present project were those for whom the original CDI forms could be located, and were without missing, inconsistent, or otherwise problematic data at the individual item level.

Measures

Three typicality measures were computed for each participant, all based on the CDI:WG: Gestures Produced; Words Understood, and Words Produced. These measures were computed using the program developed in Study 1, and the typicality measures for each child were standardized with reference to the means and standard deviations estimated from the CDI norming study sample. Because of the size and representativeness of the CDI norming sample, the parameters estimated from that sample were taken as population values. Results from the

present ASD sample were evaluated by comparison with them similar to comparing measures from a sample to a normed test score, i.e., sample-to-population rather than sample-to-sample. Consequently, the most important question is the evaluation of the null hypothesis that the present ASD sample was drawn from a population with a mean of zero and standard deviation of one for each of the three measures. If vocabulary composition for this population is less typical than a normally developing population, the means should be negative. (Note that while negative typicality scores are interpretable as atypical development, it is not clear how to interpret unusually positive scores.) Furthermore, the observed z-score means are effect sizes in themselves.

RESULTS

Table 3 provides a summary of the three developmental measures and the corresponding three typicality measures for the full ASD sample, and for separate analyses by gender and by diagnostic category within that sample. The mean levels of the three measures correspond to median performance (Alanguage age@) at 12, 13, and 15 months for Gestures Produced, Words Understood, and Words Produced Respectively. Boys have significantly higher Words Understood scores, and that is the only significant gender difference. Children with autism spectrum disorders other than autism have higher Gestures Produced scores, and that is the only significant diagnostic classification difference.

The lower half of the tables presents the results of the analyses concerning the typicality measurs. The mean typicality for the full ASD sample (first numerical column) is significantly below that for the (typically developing) norming sample of Study 1, as hypothesized, for Gestures Produced (t(116)=-9.22, $p < .001$) and for Words Produced ($t(116)=-13.72$, $p < .001$). In

both cases, the effect sizes are large. However, for Words Understood, there was a significant effect in the opposite direction ($t(116)=3.25$, $p < .01$). There are no significant differences for either gender or diagnostic classification for the typicality measures.

Because typicality scores for Gestures Produced and Words Produced were significantly and substantially lower for the ASD sample as a whole compared to typically developing children, it is of interest to examine individual words and gestures. For each child, words and gestures were identified as unusual ($p<.1$) by their presence in a repertoire of this size; the words and gestures which emerged most frequently are listed in Table 4. Conversely, words and gestures were identified as unusual ($p> .9$) by their absence in a repertoire of this size; the words and gestures which emerged most frequently are listed in Table 5. (Note that these lists aggregate across the full developmental level; a word which is quite typical later in development will occur in Table 4 if a number of children with ASD produce it early in development; conversely, a word which is unusual early in development will occur in Table 5 if a number of children with ASD fail to produce it later in development.) In both Tables, the CDI category to which the item belonged is listed. In Table 4, the words which are relatively early appearing are most often in the categories of furniture and rooms (3), sound effects and noises (3), action words (2), animal names (2), toys (2), and people (2). Although the appearance of words for people on this list is at first surprising, it is notable that the specific words are for 'teacher', with a well-defined role and location, and 'uncle', a relatively distant relationship. In contrast, as Table 5 illustrates, the primary relationship words of 'mommy' and 'daddy' are relatively late acquisitions. Other words that are acquired late tend to be toys (5), games and routines (3), clothing (2), and action words (2).

In the domain of gestures, early-acquired items are most frequently from the categories of pretending to be a parent (3) and actions with objects (2). Late-acquired items are most frequently from the categories of actions with objects (4), imitating other adult actions (3), and first communicative gestures(2). On the whole, this pattern is consistent with expectations for children with ASD. An examination of the specific items in the category 'actions with objects' provides interesting evidence for the value of examining individual items. The two items from this category in Table 4, putting on a shoe or sock, and stirring pretend liquid, are nonsocial activities, whereas the four late-acquired items in Table 5 include two that are quite social (put telephone to ear and throw a ball) and a third that has a dramatic element (push toy car or truck). To be sure, on the basis of parental checking of a behavior without additional contextual information, interpretation of the behavior as social or nonsocial can only be very tentative.

DISCUSSION

As hypothesized, children with autism spectrum disorders have significantly and substantially lower typicality scores for Gestures Produced and Words Produced. This finding is consistent with a result of Study 1, that more slowly developing children show lower typicality, though the effect size is much larger here, as would be expected for a clinical defined sample with known language impairments. The result suggests that typicality measures provide a valid means of assessing disruption in vocabulary, a core component of language, at a very early stage of development for children with ASD. For Words Understood, however, there is a smaller but significant difference observed in the opposite direction. Whether this finding reflects the difficulty parents of children with ASD have in assessing comprehension due to reduced use of social cues of comprehension such as orientation (Bruckner, Yoder, Stone & Saylor, 2007), or a

genuine qualitative difference in receptive vs. expressive lexical composition will require examination of evidence using other methods of vocabulary study. Nevertheless, the new measure appears to be more sensitive to the clinical classification of ASD within the domain of production than the category-based analyses (nominals, predicates, closed-class items) of Charman et al. (2003) and Luyster et al. (2007). An intriguing question for further research is whether these differences in the qualitative nature of early vocabulary reflect specifically linguistic aspects of ASD, more general differences in interest and motivation for children with ASD, or features of the linguistic environment and educational program provided for them. Fine-grained analyses of this type might be used to select stimuli for experimental training studies which could explore this question.

In this first exploration, although typicality measures were sensitive to the presence of ASD, they did not appear to be sensitive to the distinction between autism and other forms of ASD. The utility of the measure for this distinction and other proposed categories of communicative disorder merits further investigation, particularly with longitudinal designs, given the instability of early ASD subclassifications.

Fine-grained analysis of lexical and gestural composition has at least two potential applications for work with children with autism spectrum disorders. First, it might provide basic information for the development of an autism-specific modification of the CDI. The unevenness of language development in this population means that neither the CDI:Words & Gestures nor the CDI:Words & Sentences is completely appropriate. There may be ceiling effects for some categories of words on the CDI:WG, e.g., furniture and rooms, animal names, but moving to the CDI:WS may be too big a leap for other categories, such as games and routines, and clothing, as

well as giving up assessment of comprehension. A combination or hybrid instrument, based on observed developmental patterns, would be better suited for this purpose. Bruckner et al. (2007) came to somewhat similar conclusions based on a different analytic approach, differential item functioning, which identified a number of items that appeared to operate differently as indices of development for children with ASD and typically developing children. (Their analysis, however, examined only the Words Understood scale.) Second, item-specific information may provide useful insights for the selection of intervention targets. For example, priority might be given to those items which are particularly difficult for children with ASD, as illustrated in Table 5.

STUDY 3

Etiology of Typicality

What is the source of early variability in vocabulary and gestural composition? As typicality is a new measure, there is no prior literature on it. However, there is a rich and diverse literature on other dimensions of language acquisition style, particularly those involving proportion of nominals vs. other broad categories (referential/expressive vocabulary, etc.) and also those involving the analytic vs. holistic/imitative/rote dimension. A wide variety of explanatory factors have been suggested (cf. Shore, 1995 for a thorough and insightful review). Some of the putative explanations are directly environmental in nature, such as those that emphasize mothers= use of nouns or object naming as a determinant of the balance of nominals in early child language. These behaviors in turn may be related to social class differences in parenting style. However, the assumption that such differences reflect environmental effects may be unwarranted; such parent-child correlations are highly ambiguous. Mothers may be responding to child interest and

communication, or the behavior of both mother and child may reflect common genetic endowment. A more frequent category of explanations are based on relating language style to cognitive, social and personality dimensions, e.g., preferred types of play, imitativeness, and temperament. Still others are language-specific, typically emphasizing differential sensitivity to language input or aspects of language. Most of these latter two types of explanation have not been explicit about etiology, but the general tone seems to be one of assuming that the differences are organismic, which we take to be genetic. And occasionally this possibility is explicitly stated, e.g., Hardy-Brown (1983). Still another perspective on variation in vocabulary composition comes from cross-linguistic studies, which have identified both commonalities in vocabulary composition and its development (e.g., Bornstein, Cote, Maital, Painter, Park, Pascual, Pecheux, Ruel, Venuti & Vyt, 2004) and differences (Tardif, Fletcher, Liang & Kaciroti, in press). The latter in particular presumably reflect the influence of linguistic environment on composition, at least with respect to central tendencies across languages.

Behavioural genetic analysis based on twin data (Plomin, DeFries, McClearn & McGuffin, 2008) provides a valuable tool for exploring the sources of stylistic differences such as the typicality measure proposed in this paper. Comparisons of correlations between monozygotic (identical) twins, who share 100% of their genetics, and dizygotic (fraternal) twins, who share on average 50%, make it possible to analyze the total variance on the measure of interest into three components:

1. variance due to genetic variance among children, symbolized h^2;

2. variance to shared environmental factors B environmental factors such as neighborhood, SES, and consistent parenting styles that are common to family members and thus tend to make siblings similar to each other, symbolized c^2;

3. variance due to nonshared environmental factors B those that affect individual children, such as illness or changing family circumstances, and thus tend to make siblings less similar to each other, symbolized e^2.

In addition to illuminating the relative role of genetic and environmental factors in typicality, there is another and even more important motivation for conducting such an analysis of a new measure. Measurement error variance will be included in the third category above, nonshared environmental variance (Plomin et al., 2008). Evidence that either or both of the first two sources of variance, genetic and shared environment, is statistically significant and substantial constitutes evidence for the reliability and validity of the measure. In effect, such evidence shows that the measure is reliably related to genetic and shared environmental factors, even though they are not measured directly, and hence provides convergent validity for the measure.

It is by now well-established that variability in the rate of early language development is partially genetic, and more substantially environmental in its source (Dionne et al, 2003), and that genetic influence is generally smaller in early childhood than later. In the present study, we examine typicality, an orthogonal dimension of language development independent of rate. No directional hypothesis about etiology can be motivated at this time. Atypicality could reflect culture- and family-specific themes, leading to higher shared environment influence. Or it could

reflect child interests, which could themselves be partly genetic (Agenes code for appetites, not aptitudes,@ Kovas et al., 2007). Or both processes could be occurring.

METHOD

In this study, we take advantage of vocabulary production data at age 2 and age 3 in a very large, population-based sample of twins, the Twins Early Development Study (Trouton, Spinath & Plomin, 2002; Oliver & Plomin, 2007).

Participants

The participants in this study constituted a subsample of the full TEDS sample. All twin pairs for whom language measures were available at both 2 years and 3 years were identified. As in other TEDS analyses, twins were then excluded for uncertain zygosity, major medical disorders including serious perinatal hazard and genetic syndromes, English as the nonprimary language of the home, and parent report booklets returned outside the appropriate time window. This resulted in a final sample of 2929 pairs, consisting of 1918 MZ twins (846 males, 1072 females), 1936 DZ twins (974 males, 962 females), and 2004 twins in opposite-sex DZ pairs. As discussed in Trouton, Spinath and Plomin (2002) and Dale et al. (2003), the TEDS sample is reasonably representative of the UK population, notably with respect to mothers= educational qualifications.

Measures

In TEDS, shorter vocabulary checklists of 100 words were used, rather than the full CDI instruments. At age 2, the measure was a UK adaptation of the Level II Short Form, based on the CDI:WS (Fenson et al., 2000). At age 3, the measure was a UK adaptation of the CDI-III (Fenson et al., 2007). Only vocabulary production is assessed with these measures. Previous research (Dionne et al., 2003) has shown that these measures of vocabulary are moderately heritable (at

age 2, h^2 = 21% and 17% for the 1994 and 1995 birth cohorts, respectively; at age 3, h^2 = 10% and 14%), with higher values for shared environmental influence (at age 2, c^2 = 75% and 80% for the two cohorts; at age 3, c^2 = 84% and 82%).

Analysis of TEDS vocabulary data to construct the typicality measure followed exactly same procedure as in Study 1 above. The probability of each word occurring at each possible total score (WordTyp) was first calculated, using the TEDS sample as a self-norming study; and a cubic curve for the probability of each word occurring in vocabularies of each possible size was fitted to the empirical data for smoothing purposes. Next the mean typicality for each child=s vocabulary was computed, followed by normalization compared to other children with the same total vocabulary size, yielding a VocTyp measure for each child. Due to the large size of the TEDS sample, and the reduced range of scores, no binning was necessary; normalization could be performed at each possible total vocabulary score.

RESULTS

Phenotypic analyses

For these analyses, in order to preserve independence of data, one twin was chosen randomly from each pair. Because extensive longitudinal information was available on children and families in TEDS, several interesting additional phenotypic questions could be addressed. As shown in Table 5, in contrast to the results of Study 1, typicality was not related to gender.

Table 6 reports correlations of typicality with mothers' education, concurrent nonverbal ability, and two outcome measures at age 7 administered by telephone, vocabulary (from the WISC-III-UK) and word reading (Test of Word Reading Efficiency; TOWRE). See Kovas, Haworth, Dale and Plomin (2007) for additional information on the administration and validation

of these age 7 measures. Those correlations which are significant, if weak, are consistently positive: higher typicality at 2 is associated with higher maternal education and 7-year vocabulary and word reading, and higher typicality at 3 is associated with higher nonverbal ability and 7-year word reading. This pattern is consistent of the finding from Study 1 that typicality is positively correlated with rate of development.

Behavioural genetic modeling

For the behavioural genetic modeling, only same-sex DZ pairs were included, for comparison with MZ pairs, which are necessarily same-sex. All analyses were performed using the Mx statistical software program (Neale, Boker, Xie & Maes, 2002), using full maximum likelihood (FIML) estimation. Table 7 reports intraclass correlations and model-fitting analyses for vocabulary at age 2 and age 3. Vocabulary typicality at both 2 and 3 - independent of developmental level - is modestly but significantly heritable, as shown by higher correlations for MZ than DZ twins, and confirmed by model-fitting analyses which yield heritability estimates for 14% and 19%, respectively. Shared environment influence is considerably larger, with estimates of 49% and 50%, respectively.

DISCUSSION

This study represents the first behavioural genetic analyses of a dimension of individual differences in early language acquisition style. The results provide evidence for the reliability and validity of the measure. The MZ correlations of .65 and .70 at ages 2 and 3, respectively, which reflect both genetic and shared environmental effects, provide a lower-bound estimate of the reliability of the measures. Genetic and shared environmental variance, both significant, sum to

63% and 69% of the variance at ages 2 and 3, respectively. These figures provide evidence for substantial validity of the measure, as described in the introduction to this study.

Although these results demonstrate genetic influence on typicality, perhaps reflecting genetic influence on child interests, the effect of shared environmental influences was considerably stronger. This finding is consistent with the hypothesis that culture- and family-specific themes may influence early vocabulary. However, characterizing the specific environmental factors that are operating remains a challenging task for future research. It is possible that frequency of use of individual words by parents is one important influence (Goodman, Dale & Li, 2008). But it is also likely that factors of patterns of language use (Nelson, 1973), or even nonverbal factors play a substantial role. Detailed information about the language learning environment of individual children is required to identify environmental factors and determine their influence. This will not be easy, as behavioural genetic modeling requires large samples in order to estimate parameters. The problem is magnified by the likelihood of significant gene-environment correlations, in which aspects of the environment are themselves influenced by genetic factors shared by parent and child (Plomin et al., 2008). More complex designs and still larger sample sizes wll be required to disentangle these factors.

CONCLUSION

In this paper, several sources of evidence are provided that the newly defined typicality measures have some essential desired properties: they are relatively independent of level of development (though there is some relationship to rate of development); they show only weak gender differences; they are sensitive to the clinical classification of autism spectrum disorder independent of the overall vocabulary delay; and the vocabulary production measure used in

TEDS shows significant genetic and shared environmental influences, confirming the validity of the measure. The utility of identifying exceptionally early- or late-acquired vocabulary and gestural items is also illustrated by the results from children with ASD in Study 2.

An important limitation to the generalizability of the conclusions of this research is the inherent limitation to words and gestures included in the CDI instruments. It is very likely that many, perhaps most, children have mastered words and gestures that are not included in the CDI. And the more atypical a child=s vocabulary, the more >off-CDI= items are likely to be mastered. It is a plausible assumption that the number of atypical items on the CDI which are checked by the parent is positively correlated with the number of off-CDI items. It is also likely that the number of off-CDI items is positively correlated with the total number of CDI items checked, that is, as children=s language grows, they are more likely to know words not on the CDI. Consequently, the CDI scores for children with atypical vocabularies are likely to underestimate the child=s development. Although this problem does not directly affect the overall validity of the measure - children with more atypical vocabularies based on the measures proposed here have genuinely less typical vocabularies - it may affect some specific comparisons. In particular, the finding in Study 1 that faster children have more typical vocabularies, may be partially due to this effect, because the children with more atypical vocabularies are not, in effect, receiving credit for all the words they know. However, on more detailed examination, the correlation between Words Produced (on both the CDI:WG and CDI:WS) held even for children with relatively small vocabularies, where off-CDI items are less likely.

The program developed for this project is available from the first author, and we would welcome additional research on its use. Among issues to be explored are four that seem

especially interesting to us. First, what are the specific genetic and environmental influences on typicality? Perhaps surprisingly, in Study 3 we found a positive correlation between mother=s education B known to be reflected in a more diverse vocabulary in speech to children B and typicality in the vocabularies of two-year-olds. How do patterns of language use affect typicality? Second, how useful will these measures be for distinguishing various clinical populations and subtypes? The validity and cost-effectiveness of the CDI instruments for assessing early language development facilitates the collection of large samples of item-specific developmental data for this purpose. Third, what is the predictive significance of the measure, beyond overall developmental level, both for normal development and that of clinical populations? And fourth, how might this information inform therapeutic intervention, both in selecting therapeutic targets and for anticipating ease or difficulty of learning? All four of these questions bear on the larger goal of helping us better understand how clinical populations differ from typical development and from each other. We suggest that this 'micro-level' descriptive information about vocabulary and gestural composition provides a within-subject comparative perspective (across items) that can facilitate theory-testing. Finally, we would point out that the analytic technique used for the development of these measures could be used to define parallel measures for other languages, capitalizing on norming studies for adaptations of the CDI into other languages.

References

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., Reilly, J., & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. Journal of Child Language, 21, 85-123.

Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In Paul Fletcher & Brian MacWhinney (Eds.), *Handbook of Child Language*, 96-151. Oxford: Basil Blackwell.

Benedict, H. (1979). Early lexical development: Comprehension and production. Journal of Child Language, 6, 183-200.

Bishop, D. V. M. (2000). Pragmatic language impairment: A correlate of SLI, a distinct subgroup, or part of the autistic continuum? In D. V. M. Bishop & L. B. Leonard (Eds.), *Speech and Language Impairments in Children: Causes, Characteristics, Intervention and Outcome*, 99-113. Philadelphia: Psychology Press.

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S-Y., Pascual, L., Pecheux, M-G., Ruel, J., Venuti, P., & Vyt, A. (2004). Cross-linguistic analysis of vocabulary in Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. Child Development, 75, 1115-1139.

Bruckner, C., Yoder, P., Stone, W., & Saylor, M. (2007). Construct validity of the MCDI-I receptive vocabulary scale can be improved: Differential item functioning between toddlers with Autism Spectrum Disorders and typically developing infant. Journal of Speech, Language, and Hearing Research, 50, 1631-1638.

Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in

pre-school children with autism spectrum disorders using the MacArthur Communicative Development Inventory (Infant Form). Journal of Child Language, 30, 213-236.

Dale, P.S., Price, T. S. Bishop, D. V. M., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. Journal of Speech, Language, and Hearing Research, 46, 544-560.

Dionne, G., Dale, P. S., Boivin, M., & Plomin, R. (2003). Genetic evidence for bidirectional effects of early lexical and grammatical development. Child Development, 74, 394-412.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. (1994). Variability in early communicative development. Monographs of the Society for Research in Child Development, 59, Serial No. 242.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short form versions of the MacArthur Communicative Development Inventories. Applied Psycholinguistics, 21, 95-115.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007) *MacArthur-Bates Communicative Development Inventories: User=s Guide and Technical Manual, Second Edition*. Baltimore: Paul H. Brookes.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. Journal of Child Language, 35, 515-531.

Hardy-Brown, K. (1983). Universals and individual differences: Disentangling two approaches to the study of language acquisition. Developmental Psychology, 19, 610-624.

Kovas, Y., Haworth, C. M. A., Dale, P. S., & Plomin, R. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. Monographs of the

Society for Research in Child Development, 72, Serial No. 288.

Lee, A., Hobson, R. P., & Chiat, S. (1994). I, you, me, and autism: An experimental study. Journal of Autism and Developmental Disorders, 24, 155-176.

Luyster, R. J., Kadlec, M. B., Carter, A., & Tager-Flusberg, H. (2008). Language assessment and development in toddlers with Autism Spectrum Disorders. Journal of Autism and Developmental Disorders, 38, 1426-1438.

Luyster, R., Lopez, K., & Lord, C. (2007). Characterizing communicative development in children referred for Autism Spectrum Disorders using the MacArthur-Bates Communicative Development Inventory (CDI). Journal of Child Language, 34, 623-654.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2002). *Mx: Statistical Modeling (6th Ed.).* Richmond, VA: Department of Psychiatry.

Nelson, K. (1973). Structure and strategy in learning to talk. Monographs of the Society for Research in Child Development, 38, Serial No. 149.

Oliver, B., Dale, P. S., Saudino, K. J., Petrill, S. A., Pike, A., & Plomin, Rt. (2002). The validity of a parent-based assessment of cognitive abilities in three-year- olds. Early Child Development & Care, 17, 337-348.

Oliver, B., & Plomin, R. (2007). Twins= Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behaviour problems from children through adolescence. Twin Research and Human Genetics, 10, 96-105.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2008). *Behavioral genetics, Fifth edition*. New York: Worth.

Saudino, K. J., Dale, P. S., Oliver, B., Petrill, S. A., Richardson, V., Rutter, M., Simonoff, E.,

Stevenson, J., & Plomin, R. (1998). The validity of parent-based assessment of the cognitive abilities of two-year-olds. British Journal of Developmental Psychology, 16, 349-363.

Shore, C. M. (1995). *Individual differences in language development*. Thousand Oaks: Sage Publications.

Tager-Flusberg, H. (1994). Dissociations in form and function in the acquisition of language by autistic children. In *Constraints on language acquisition: Studies of atypical children* (pp. 175-194). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In F. Volkmar, R. Paul, A. Klin & D. Cohen (Eds), Handbook of autism and pervasive developmental disorders. 3rd ed., Vol. I, 335-364. Hoboken, NH: Wiley.

Tardif, T., Fletcher, P., Liang, W., & Kaciroti, N. (in press). Early vocabulary development in Mandarin (Putonghua) and Cantonese. Journal of Child Language.

Trouton, A., Spinath, F. M., & Plomin, R. (2002). Twins Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behaviour problems in childhood. Twin Research, 5, 444-448.

Table 1

Sample vocabularies for 3 children with 12 Words Produced on the CDI:WG, with different typicality scores: Average, Low, and Very Low

Child ID	**Typicality score for Words Produced (Number of low-probability words)**	**Vocabulary composition**[1]
I209033	.002 (2)	baabaa, yumyum, cookie, juice, baby, ***child***, grandma, mommy, ***own-name***, byebye, hi, no
58	-1.00 (6)	***car, truck, block***, book, ***cake***, cracker, ***hat***, daddy, mommy, byebye, hi, ***shh***
I114200	-2.65 (8)	***bread, candy, water, aunt***, bye, hi, no, yes, ***go, down, in, another***

[1] Words with $p < .1$ at this vocabulary level are marked with bold italics

Table 2

Mean typicality measures for the norming sample from Fenson et al. (2007)

Measure	Total Sample M (SD) N	Females M (SD) N	Males M (SD) N	t-test for gender	Cohen=s d
CDI:WG Gestures Produced	.005 (.974) 1089	-.041 (.990) 544	.051(.957) 545	t = -1.556 p = .120	0.08
CDI:WG Words Understood	.004 (.986) 1089	.068 (.966) 544	-.060 (1.001) 545	t = 2.148 p = .032	0.11
CDI:WG Words Produced	.001 (.911) 1089	.081 (.910) 544	-.079 (.907) 545	t = 2.916 p = .004	0.15

CDI:WS Words	.012 (.957)	.074 (.924)	-.051 (.985)	t = -1.556	0.09
Produced	1461	728	733	p = .013	

Table 3

Mean vocabulary and gestural scores, and typicality measure as a function of gender, and of Autism Spectrum Disorder classification, for the ASD study sample

Measure	Full Sample N=116 M (SD)	Females N=32 M (SD)	Males N=84 M (SD)	t-test for gender difference	AUT N=59 M (SD)	ASD N=57 M (SD)	t-test for Dx difference
CDI:WG Gestures Produced score	24.0 (10.3)	23.1 (10.3)	24.3 (10.4)	*t*=.56 *p* = .58	22.0 (10.6)	26.0 (9.7)	*t*=-2.11 *p*=.037
CDI:WG Words Understood score	102.5 (80.2)	78.2 (51.9)	111.8 (87.2)	*t*=-2.54 *p* = .013	92.8 (83.7)	112.5 (75.9)	*t*=-1.33 *p*=.187
CDI:WG Words Produced score	22.8 (35.4)	19.6 (30.6)	24.1 (37.1)	*t*=.61 *p* = .54	20.0 (36.2)	25.8 (34.6)	*t*=-.883 *p*=.38
CDI:WG Gestures Produced typicality	-.90 (1.05)	-1.18 (1.07)	-.79 (1.02)	*t* = -1.80 *p* = .074	-.96 (1.02)	-.83 (1.07)	*t*=-.658 *p*=.51
CDI:WG Words	.43 (1.42)	.23 (1.30)	.51 (1.46)	*t*=.948	.21 (1.36)	.65 (1.46)	*t*= -1.65

Understood typical.				p = .35			p=.102
CDI:WG Words	-1.33 (1.05)	-1.13 (1.96)	-1.41 (1.04)	t = 1.309	-1.27 (1.09)	-1.39 (1.00)	t=.613
Produced typicality				p = .183			p=.54

Table 4

The ten most commonly occurring, atypically early words and gestures relative to developmental level for children with ASD

CDI:WG Words Understood		CDI:WG Words Produced		CDI:WG Gestures[1]	
Word	**Category**	**Word**	**Category**	**Gesture**	**Category**
teacher	People	grr	Sound effects & noises	put on a shoe or sock	C
bathroom	Furniture & rooms	baabaa	Sound effects & noises	put to bed (doll or animal)	D
bedroom	Furniture & rooms	balloon	Toys	play Aso big@	
church	Outside & places to go	down	Prepositions & locations	rock it (doll or animal)	D
uncle	People	vroom	Sound effects & noises	stir pretend liquid in cup or pan with a spoon	C
wet	Descriptive words	go	Action words		
away	Prepositions & locations	cup	Small household items	brush/comb its hair (doll or animal)	D
break	Action words	pig	Animal names		
carrots	Food & drink	bird	Animal names	sweep with broom or mop	E

kitchen	Furniture & rooms	block	Toys

A = First communicative gestures; B = Games & routines; C = Actions with objects; D = Pretending to be a parent; E = Imitating other adult actions

Table 5

The ten most common atypically late words and gestures relative to developmental level for children with ASD

CDI:WG Words Understood		CDI:WG Words Produced		CDI:WG Gestures[1]	
Word	**Category**	**Word**	**Category**	**Gesture**	**Category**
ball	Toys	ball	Toys	Aread@ (open book-turn pg)	E
bye	Games & routines	bye	Games & routines	push toy car or truck	C
daddy	People	dog	Animal names	dance	B
book	Toys	bubbles	Toys	eat with spoon or fork	C
show	Action words	book	Toys	write with pen/pencil/marker	E
bath	Games & routines	cookie	Food & drink	put telephone to ear	C
mommy	People	car	Vehicles	waves byebye on own	A
car	Vehicles	cat	Animal names	throw a ball	C
eat	Action words	hat	Clothing	extends arm to request pickup	A
diaper	Clothing			put on glasses	E

	juice	Food & drink

[1]A = First communicative gestures; B = Games & routines; C = Actions with objects; D = Pretending to be a parent; E = Imitating other adult actions

Table 5

Mean vocabulary typicality measures for the TEDS sample by gender

	Vocabulary typicality at 2		**Vocabulary typicality at 3**	
	M (SD)	**N**	**M (SD)**	**N**
Total sample	0 (1.00)	2921	0 (1.00)	2921
Males	-.02 (1.05	1404	-.02 (1.01)	1402
Females	.02 (.95)	1517	.01 (.99)	1519
t-test for gender difference	$t(2834.21) = 1.05$, $p = .29$		$t(2919) = .97$, $p = .43$	
Cohen's *d* for gender	.04		.03	

Table 6

Correlation of vocabulary typicality measures with mothers= educational level, concurrent nonverbal ability (PARCA), and WISC Vocabulary and Word Reading (TOWRE) at age 7, for TEDS sample

Measure	**Mother=s educ**	**PARCA**	**7yr WISC Vocabulary**	**7yr TOWRE**
Age 2 typicality	.115 ***	-.003	.049 *	.066 **
Age 3 typicality	.017	.049 **	.020	.062 **

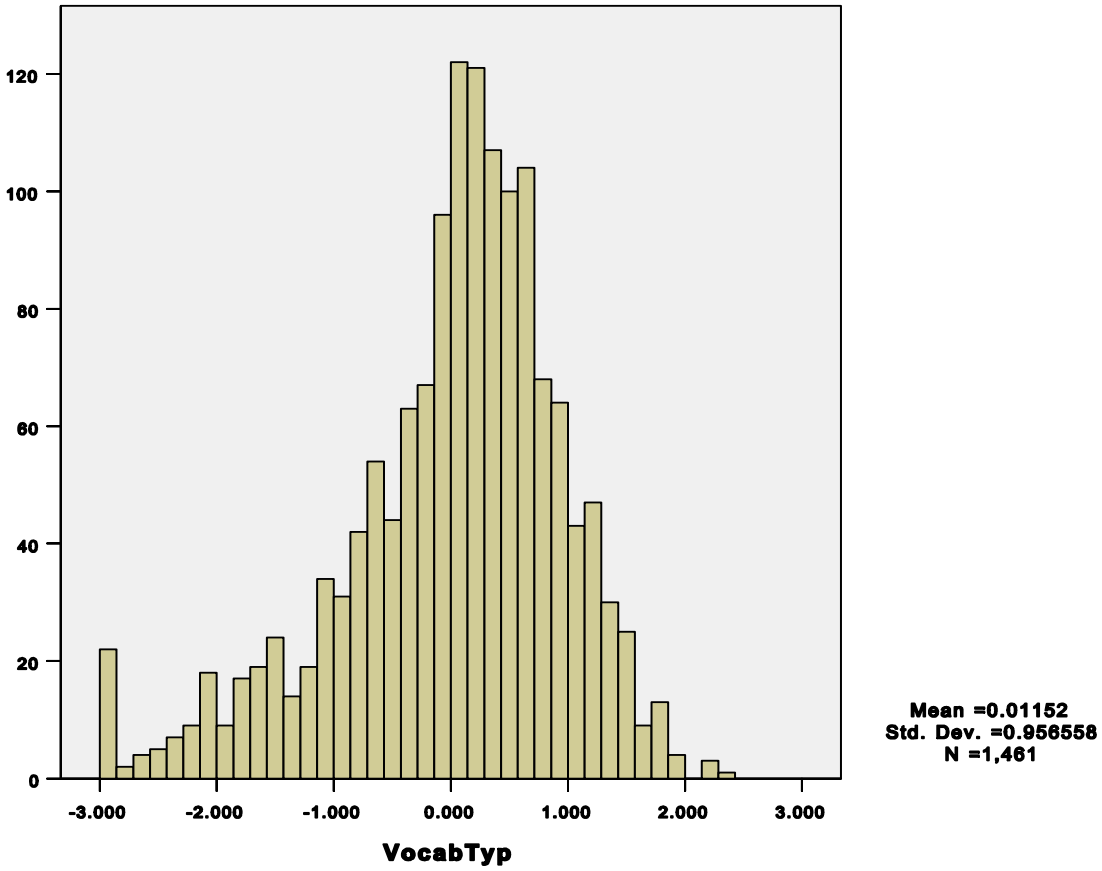
Table 7

Intraclass correlations and model-fitting results for typicality measures (95% confidence intervals in parentheses), for TEDS sample

Measure	**MZ corr**	**DZ corr**	**h^2**	**c^2**	**e^2**
Age 2 typicality	.65 (.91, .69)	.53 (.49, .58)	.14 (.04, .25)	.49 (.40, .58)	.37 (.34, .40)
Age 3 typicality	.70 (.67, .73)	.57 (.53, .62)	.19 (.10, .28)	.50 (.41, .57)	.31 (.28, .35)

Figure Captions

Figure 1. Illustrative histogram of the typicality measure for CDI:WS Words Produced


120
100
80
60
40
20
0
-3.000
-2.000
-1.000
0.000
1.000
2.000
3.000
VocabTyp
Mean =0.01152
Std. Dev. =0.956558
N =1,461