

PREVIC: An Item Response Theory based parent report measure of expressive vocabulary  
in children between 3 and 8 years of age

Manuel Bohn<sup>1,2</sup>, Julia Prein<sup>1</sup>, Tobias Koch<sup>3</sup>, Maximilian Bee<sup>3</sup>, Daniel Haun<sup>2</sup>, & Natalia  
Gagarina<sup>4</sup>

<sup>1</sup> Institute for Psychology, Leuphana University Lüneburg, Germany

<sup>2</sup> Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary  
Anthropology, Leipzig, Germany

<sup>3</sup> Institut of Psychology, Friedrich-Schiller-University Jena, Germany

<sup>4</sup> Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin, Germany

11 We thank Susanne Mauritz for her help with the data collection.

12 The authors made the following contributions. Manuel Bohn: Conceptualization,  
13 Formal Analysis, Writing - Original Draft Preparation, Writing - Review & Editing; Julia  
14 Prein: Conceptualization, Software, Writing - Original Draft Preparation, Writing - Review  
15 & Editing; Tobias Koch: Formal Analysis, Writing - Review & Editing; Maximilian Bee:  
16 Formal Analysis, Writing - Review & Editing; Daniel Haun: Conceptualization, Writing -  
17 Review & Editing; Natalia Gagarina: Conceptualization, Writing - Original Draft  
18 Preparation, Writing - Review & Editing.

19 Correspondence concerning this article should be addressed to Manuel Bohn, Max  
20 Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig,  
21 Germany. E-mail: manuel\_bohn@eva.mpg.de

Abstract

lore ipsum.

*Keywords:* language development, vocabulary, individual differences, Item Response  
Models

Word count: X

PREVIC: An Item Response Theory based parent report measure of expressive vocabulary  
in children between 3 and 8 years of age

## Introduction

Learning language is one of the key developmental objectives for children. This learning process is highly variable and leads to persistent individual differences which are related to a wide range of outcome measures later in life (Bleses, Makransky, Dale, Højen, & Ari, 2016; Bornstein, Hahn, Putnick, & Pearson, 2018; Roberta Michnick Golinkoff, Hoff, Rowe, Tamis-LeMonda, & Hirsh-Pasek, 2019; Marchman & Fernald, 2008; Morgan, Farkas, Hillemeier, Hammer, & Maczuga, 2015; Pace, Alper, Burchinal, Golinkoff, & Hirsh-Pasek, 2019; Pace, Luo, Hirsh-Pasek, & Golinkoff, 2017; Schoon, Parsons, Rush, & Law, 2010; Walker, Greenwood, Hart, & Carta, 1994). For example, in a longitudinal study spanning 29 years, Schoon et al. (2010) found that relatively poorer language skills at age five were associated with lower levels of mental health at age 34. Given their high predictive validity, high-quality measures are needed to assess early language abilities.

Child language measures can be broadly categorized into two types: direct and parent report measures. Direct measures are generally used with children of three years and older. Direct expressive language assessments involve prompting children to generate words or sentences in response to a given stimulus, such as a picture depicting a scene or an object. Direct receptive language assessments require children to match a verbal prompt with a corresponding picture of a scene or object. Various direct measures tailored to different languages and age groups have been developed, including measures for English and German (Dunn & Dunn, 1965; Dunn, Dunn, Whetton, & Burley, 1997; Glück & Glück, 2011; Roberta M. Golinkoff et al., 2017; Kauschke & Siegmüller, 2002; Kiese-Himmel, 2005; Lenhard, Lenhard, Segerer, & Suggate, 2015). Additionally, standardized cognitive ability tests frequently incorporate direct language measures (e.g., Bayley, 2006; Gershon et al., 2013; Wechsler & Kodama, 1949).

Parent report measures are widely utilized in psychological research. They are particularly popular as screening methods to identify developmental delays (Diamond & Squires, 1993; Pontoppidan, Niss, Pejtersen, Julian, & Væver, 2017). However, it is important to acknowledge that parent reports come with certain caveats, including the potential for selective reporting and social desirability bias. As a consequence, providing a comprehensive assessment of the overall quality and usefulness of these measures is challenging (Morsbach & Prinz, 2006). Nonetheless, some parent report measures have been found to be both reliable and valid (Bodnarchuk & Eaton, 2004; Hornman, Kerstjens, Winter, Bos, & Reijneveld, 2013; Ireton & Glascoe, 1995; Macy, 2012; Saudino et al., 1998).

In child language research, parent report measures are often utilized with very young children when direct assessment is challenging. One widely used measure is the MacArthur-Bates Communicative Development Inventories (CDI) (Fenson et al., 2007). The CDI asks parents to check those words from a checklist that they believe their child produces and/or understands. This measure has been adapted for a wide range spoken and signed languages (see Frank, Braginsky, Yurovsky, & Marchman, 2021 for an overview), with various versions available (e.g., Makransky, Dale, Havmose, & Bleses, 2016; Mayor & Mani, 2019), including an online version (DeMayo et al., 2021). Collaborative efforts have facilitated the pooling of data from thousands of children learning different languages into centralized repositories (Frank, Braginsky, Yurovsky, & Marchman, 2017; Jørgensen, Dale, Bleses, & Fenson, 2010). Importantly, the CDI exhibits validity as parental reports align with direct observations and assessments of child language (Bornstein & Haynes, 1998; Dale, 1991; Feldman et al., 2005; Fenson et al., 1994).

However, the use of the CDI – in typically developing children – is limited to 37 months of age. Consequently, there is a need for a comparable measure that can be applied to older children, as parental reports offer a convenient and comprehensive means of assessing children’s language abilities and provide a complementary perspective on development. Existing instruments focusing on general cognitive development often include

language scales; however, these scales lack detailed information and fail to capture individual differences effectively (Ireton & Glascoe, 1995). For example, the Ages and Stages Questionnaire (ASQ) at 36 months comprises only six items that encompass general communicative behavior, such as whether the child can say their full name when prompted (Squires, Bricker, Twombly, et al., 2009). One notable example of a dedicated language measure for older children is the Developmental Vocabulary Assessment for Parents (DVAP, Libertus, Odic, Feigenson, & Halberda, 2015). The DVAP is derived from the words used in the Peabody Picture Vocabulary Test [PPVT; Dunn and Dunn (1965)], a widely used direct measure of receptive vocabulary. As perhaps expected, the DVAP demonstrates high validity, as evidenced by its strong correlation with the PPVT. However, the proprietary nature of the PPVT limits the utility of the DVAP for researchers.<sup>1</sup> As a consequence, it is unlikely that comparable “success story” – as observed with the CDI – will emerge where researchers have adapted the original English form to different languages and more efficient forms.

More general issue with language measures is lack of psychometric considerations during development. The developmental process is not transparent IRT provides a toolkit that is very useful for development. Each item is assessed in its usefulness to measure a latent construct at different ages

### The current study

Our goals was to develop a high-quality and easy-access vocabulary checklist for children between three and eight years of age. The measure and all associated materials should be openly available for other researchers to use. To ensure the quality of the items, we created a large initial item pool from which we selected high-quality items. To ensure

---

<sup>1</sup> When the first author approached the license holder of the PPVT in Germany to ask if we could use the words to build a parental report measure, we were told that we would have to pay for every administration of the new measure and we would not be allowed to openly share the materials.

103 easy-access, we implemented the checklist as an interactive web-app.

104 **Task design an implementation**

105 **Item pool generation**

106 **Item selection**

107 **Participants**

108 **Descriptive results**

109 **Automated item selection**

110 **Differential item functioning**

111 **Psychometric properties of new checklist**

112 **Reliability**

113 **Convergent validity**

114 **Discussion**

115 Measure is likely different in type from CDI. Especially early in life, CDI likely  
116 captures the entire vocabulary.

117 **Limitations**

118 Not a representative sample

## Conclusion

### Open Practices Statement

The task can be accessed via the following website:

<https://ccp-odc.eva.mpg.de/orev-demo/>. The corresponding source code can be found in the following repository: <https://github.com/ccp-eva/orev-demo>. The data sets generated during and/or analysed during the current study are available in the following repository: <https://github.com/ccp-eva/orev/>. Data collection was preregistered at: <https://osf.io/qzstk>.



## References

- Bayley, N. (2006). *Bayley scales of infant and toddler development—third edition*. San Antonio, TX: Harcourt Assessment.
- Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics*, 37(6), 1461–1476.
- Bodnarchuk, J. L., & Eaton, W. O. (2004). Can parent reports be trusted?: Validity of daily checklists of gross motor milestone attainment. *Journal of Applied Developmental Psychology*, 25(4), 481–490.
- Bornstein, M. H., Hahn, C.-S., Putnick, D. L., & Pearson, R. M. (2018). Stability of core language skill from infancy to adolescence in typical and atypical development. *Science Advances*, 4(11), eaat7422.
- Bornstein, M. H., & Haynes, O. M. (1998). Vocabulary competence in early childhood: Measurement, latent construct, and predictive validity. *Child Development*, 69(3), 654–671.
- Dale, P. S. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech, Language, and Hearing Research*, 34(3), 565–571.
- DeMayo, B., Kellier, D., Braginsky, M., Bergmann, C., Hendriks, C., Rowland, C. F., . . . Marchman, V. (2021). Web-CDI: A system for online administration of the MacArthur-bates communicative development inventories. *Language Development Research*.
- Diamond, K. E., & Squires, J. (1993). The role of parental report in the screening and assessment of young children. *Journal of Early Intervention*, 17(2), 107–115.
- Dunn, L. M., & Dunn, L. M. (1965). *Peabody picture vocabulary test*.
- Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). British picture vocabulary scale 2nd edition (BPVS-II). *Windsor, Berks: NFER-Nelson*.
- 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(4), 856–868.
- Fenson, L. et al. (2007). *MacArthur-bates communicative development inventories*. Paul H. Brookes Publishing Company Baltimore, MD.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., . . . Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, i–185.
- Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, 44(3), 677–694.
- Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: The wordbank project*. MIT Press.
- Gershon, R. C., Slotkin, J., Manly, J. J., Blitz, D. L., Beaumont, J. L., Schnipke, D., et al.others. (2013). IV. NIH toolbox cognition battery (CB): Measuring language (vocabulary comprehension and reading decoding). *Monographs of the Society for Research in Child Development*, 78(4), 49–69.
- Glück, C. W., & Glück, C. W. (2011). *Wortschatz-und wortfindungstest für 6-bis 10-jährige (WWT 6-10)*. Urban & Fischer.
- Golinkoff, Roberta M., De Villiers, J. G., Hirsh-Pasek, K., Iglesias, A., Wilson, M. S., Morini, G., & Brezack, N. (2017). *User’s manual for the quick interactive language screener (QUILS): A measure of vocabulary, syntax, and language acquisition skills in young children*. Paul H. Brookes Publishing Company.
- Golinkoff, Roberta Michnick, Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child Development*, 90(3), 985–992.
- Hornman, J., Kerstjens, J. M., Winter, A. F. de, Bos, A. F., & Reijneveld, S. A. (2013). Validity and internal consistency of the ages and stages questionnaire 60-month version

- and the effect of three scoring methods. *Early Human Development*, 89(12), 1011–1015.
- Ireton, H., & Glascoe, F. P. (1995). Assessin children’s development using parents’ reports: The child development inventory. *Clinical Pediatrics*, 34(5), 248–255.
- Jørgensen, R. N., Dale, P. S., Bleses, D., & Fenson, L. (2010). CLEX: A cross-linguistic lexical norms database. *Journal of Child Language*, 37(2), 419–428.
- Kauschke, C., & Siegmüller, J. (2002). *Patholinguistische diagnostik bei sprachentwicklungsstörungen: Diagnostikband phonologie*. Elsevier Urban & Fischer.
- Kiese-Himmel, C. (2005). AWST-r-aktiver wortschatztest für 3-bis 5-jährige kinder (AWST-r-active vocabulary test for 3-to 5-year-old children). *Göttingen: Hogrefe*.
- Lenhard, A., Lenhard, W., Segerer, R., & Suggate, S. (2015). *Peabody picture vocabulary test-4. Ausgabe: Deutsche fassung*. Frankfurt am Main: Pearson Assessment.
- Libertus, M. E., Odic, D., Feigenson, L., & Halberda, J. (2015). 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*, 16(3), 442–454.
- Macy, M. (2012). The evidence behind developmental screening instruments. *Infants & Young Children*, 25(1), 19–61.
- Makransky, G., Dale, P. S., Havmose, P., & Bleses, D. (2016). An item response theory-based, computerized adaptive testing version of the MacArthur-bates communicative development inventory: Words & sentences (CDI: WS). *Journal of Speech, Language, and Hearing Research*, 59(2), 281–289.
- 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(3), F9–F16.
- Mayor, J., & Mani, N. (2019). A short version of the MacArthur-bates communicative development inventories with high validity. *Behavior Research Methods*, 51(5), 2248–2255.
- Morgan, P. L., Farkas, G., Hillemeier, M. M., Hammer, C. S., & Maczuga, S. (2015).

- 24-month-old children with larger oral vocabularies display greater academic and behavioral functioning at kindergarten entry. *Child Development*, 86(5), 1351–1370.
- Morsbach, S. K., & Prinz, R. J. (2006). Understanding and improving the validity of self-report of parenting. *Clinical Child and Family Psychology Review*, 9, 1–21.
- Pace, A., Alper, R., Burchinal, M. R., Golinkoff, R. M., & Hirsh-Pasek, K. (2019). Measuring success: Within and cross-domain predictors of academic and social trajectories in elementary school. *Early Childhood Research Quarterly*, 46, 112–125.
- Pace, A., Luo, R., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Identifying pathways between socioeconomic status and language development. *Annual Review of Linguistics*, 3, 285–308.
- Pontoppidan, M., Niss, N. K., Pejtersen, J. H., Julian, M. M., & Væver, M. S. (2017). Parent report measures of infant and toddler social-emotional development: A systematic review. *Family Practice*, 34(2), 127–137.
- Saudino, K. J., Dale, P. S., Oliver, B., Petrill, S. A., Richardson, V., Rutter, M., . . . Plomin, R. (1998). The validity of parent-based assessment of the cognitive abilities of 2-year-olds. *British Journal of Developmental Psychology*, 16(3), 349–362.
- Schoon, I., Parsons, S., Rush, R., & Law, J. (2010). Children’s language ability and psychosocial development: A 29-year follow-up study. *Pediatrics*, 126(1), e73–e80.
- Squires, J., Bricker, D. D., Twombly, E., et al. (2009). *Ages & stages questionnaires*. Paul H. Brookes Baltimore, MD.
- Walker, D., Greenwood, C., Hart, B., & Carta, J. (1994). Prediction of school outcomes based on early language production and socioeconomic factors. *Child Development*, 65(2), 606–621.
- Wechsler, D., & Kodama, H. (1949). *Wechsler intelligence scale for children* (Vol. 1). Psychological corporation New York.