Research Article

Hammill Institute on Disabilities

Communication Disorders Quarterly
2017, Vol. 38(4) 231–241

DOI: 10.1177/1525740116679886
cdq.sagepub.com
SAGE

Another Look at the Influence of Maternal Education on Preschoolers' Performance on Two Norm-Referenced Measures

Alyson D. Abel, PhD[1], C. Melanie Schuele, PhD, CCC-SLP[2], Karen Barako Arndt, MS, CCC-SLP[3], Marvin W. Lee, PhD[4], and Kathryn Guillot Blankenship, PhD, CCC-SLP[5]

Abstract

The purpose of this study was (a) to describe the performance of preschool children from families with college-educated mothers on two norm-referenced measures, the Preschool Language Scale–4 (PLS-4) and Peabody Picture Vocabulary Tests–III (PPVT-III), and (b) to compare the findings with Qi and colleagues who reported PLS and PPVT scores for children from lower income families. The study involved a secondary analysis of previously collected PLS-4 and PPVT-III data. Participants included 146 typically developing preschoolers who attended preschools serving primarily children from college-educated mothers. Mean standard scores on both measures were at the upper end or exceeded one standard deviation above the normative mean with distributions that approximated normal. Means also greatly exceeded the lower socioeconomic status (SES) group means reported by Qi and colleagues. These results suggest that subsample norms, based on SES, yield multiple distinct but overlapping distributions. Thus, test developers should consider providing subsample norms in addition to traditional population-based norms.

Keywords

SES, language, assessment, preschool

School speech–language pathologists (SLPs) use norm-referenced measures for many purposes, including eligibility decisions for the disability category of *speech or language impairment*. The Individuals With Disabilities Education Improvement Act of 2004 (IDEA) defines conditions that must be met when conducting full and individual evaluations (see Appendix A). As such, SLPs must assure that chosen measures are "technically sound" and "selected and administered so as not to be discriminatory on a racial or cultural basis" (Sec. 614). Because the federal definition of *speech or language impairment* is broad, individual state departments of education publish guidelines or requirements for eligibility decisions. Typically, these guidelines give prominence to a child's performance on norm-referenced measures in determining eligibility as a child with speech or language impairment. For example, Missouri and Tennessee specify a standard score cutoff of 1.5 standard deviations below a measure's norming sample mean (see Appendix B) to meet eligibility requirements.

The study reported herein contributes to the discussion on identification of and eligibility determination for language impairment within the IDEA category of *speech or language impairment*. We address the challenges that ensue from the single population-based norms provided in the majority of, if not all, commercially published norm-referenced measures. We argue that norms based on a single aggregated norming sample can lead to discrimination on a cultural basis because family socioeconomic status (SES), often indexed by maternal education (Bradley, Corwyn, McAdoo, & Coll, 2001; Hoff, 2006), is a critical component of culture (see Note 1), particularly as it relates to school achievement (Waldfogel, 2012). As a result, we argue further that a well-supported evidence-based practice is consideration of family SES in the interpretation of a child's

[1]San Diego State University, CA, USA
[2]Vanderbilt University, Nashville, TN, USA
[3]Radford University, VA, USA
[4]Tennessee State University, Nashville, USA
[5]Middle Tennessee State University, Murfreesboro, USA

Corresponding Author:
Alyson D. Abel, San Diego State University, 5500 Campanile Dr. San Diego, CA 92182, USA.
Email: alyson.abel@mail.sdsu.edu

performance on norm-referenced measures when making eligibility decisions.

In the current study, a large group of preschool children with college-educated mothers were administered two widely used oral language norm-referenced measures. We compared the means for our sample with the single population-based norms provided by the publishers in the test manuals and with the group means reported by Qi and colleagues (Qi, Kaiser, Milan, & Hancock, 2006; Qi, Kaiser, Milan, Yzquierdo, & Hancock, 2003; Qi & Marley, 2009) for samples of children from lower SES families attending Head Start or other community preschools serving low-income families (see original studies for additional participant details). Following the common practice of using parent education level (typically maternal education) to designate SES, our sample of children with college-educated mothers is termed *higher SES*, whereas a sample of children with high school– or less than high school–educated parents is termed *lower SES*. We base this use of *higher* and *lower* on consideration of educational achievement in the U.S. adult population (above the age of 25) wherein approximately 30% have graduated from college and 15% did not complete high school (U.S. Census Bureau, 2014).

Findings from the current study, taken together with the extant literature, demonstrate that single population-based norms set up a situation in which lower SES children are highly likely to fall below a typical cutoff on a norm-referenced measure whereas higher SES children are not. We consider the absolutely critical need for SLPs to consider family SES in the interpretation of child performance on norm-referenced measures of oral language when making eligibility decisions, and more broadly when differentiating instruction/intervention for individual children. To meet this need, one that has been discussed in the literature for decades, publishers could enhance evidence-based assessment decisions if they provide data that allow not only for population-based normative comparisons but also for subsample normative comparisons with relevant population subsamples defined by parental education (i.e., disaggregation of the population).

To orient the reader to the motivation for and importance of our study, we review three pieces of relevant information: (a) typical norming samples, (b) the clinical challenges presented by the use of single population-based norms, and (c) the extant evidence that parental education influences the distribution of scores on norm-referenced measures of oral language.

Typical Norming Samples

Test publishers commonly report that norming studies' participants were selected to construct a sample of children who mirror the U.S. Census for gender, race, ethnicity, and, increasingly, family SES (e.g., indexed by income or education). Doing so, they argue that the norming sample is representative of the children to whom the measure will be administered (i.e., schoolchildren in the United States) thereby purportedly establishing one aspect of IDEA's requirement of "technically sound." The performance of this population-based norming sample is used to construct population-referenced norms that reflect an estimation of the population distribution indexed typically by age and/or grade. Notably, the U.S. population is quite diverse and potentially encompasses subgroups with distinct yet overlapping distributions; we return to this point below. Given the practice of constructing a norming sample that mirrors the U.S. Census, the comparison of an individual child's performance with the population-based norms provides an estimate of where the individual child performs compared with the national population of same-age children or same-grade students. The interpretation of the child's rank within this population is not driven by the measure itself but rather involves clinical judgment. Clinical judgment is influenced, in part, by information in the manuals provided by the test publisher, by peer-reviewed studies on a measure, by an individual SLP's experience, and by local, state, and federal regulations and policies. A critical question is how comparison with population-based norms influences IDEA eligibility decisions.

A Single Population-Based Norming Sample and Population-Based Norms: Potential Problems

The most common norm-referenced decisions use cutoff scores; the state, school district, or an individual SLP specifies a cutoff score to demarcate normal and not normal. Common cutoffs are 1.0, 1.5, or 2 standard deviations below the population mean (see Fey, 1986). The logic of making a comparison with population-based norms for interpreting the performance of an individual child appears at first glance reasonably straightforward: Those children who score below the cutoff have the least proficiency in language and are the least able to learn language. From this perspective, it is reasonable to conclude that children below the cutoff have language impairments. More careful consideration of this logic may lead to questions, however. A child's language-learning outcome can be viewed as the product of the child's ability to learn language given adequate input and the environment's capacity to deliver adequate input (quantity as well as quality) considering that, unlike academic skills such as word decoding, spelling, or math calculations, oral language is not explicitly taught to children. Under circumstances where we (society) can be fairly certain that the environment (e.g., home, child care, preschool) provided sufficient language-learning opportunities, oral language performance that is substantially below the mean of same-age peers might be construed as indicative of impaired language-learning ability.

There is substantial evidence, however, that individual differences in language outcomes relate to variability in maternal education and in associated quantity and quality of language-learning opportunities (for a review, see Hoff, 2006). Thus, the use of a single population-based norm to inform eligibility decisions may conceal meaningful subsample distributions that differ from the population distribution. Hutchinson (1996) warned that the wide-angle view of population-based norming samples is insufficient if the performance of subsamples is incongruent with the population distribution. In this case, although the population-based norms can identify those children with the least absolute oral language proficiency from a population perspective, these norms provide no information on how the child performed relative to subsample peers. The lack of subsample-referenced norms for comparison can lead to over- or under-identification of children with language impairment.

If, as the literature suggests (see "The Evidence: Language Outcomes in Subgroups Defined by Maternal Education" section), SES subsamples have distinct yet overlapping distributions of performance on commonly used norm-referenced measures, the proportion of children within each SES subsample who fall below a specified cutoff may vary widely when using the population-based norms to make eligibility decisions. For example, if the mean of a subgroup is substantially greater than the population mean, then far fewer children within that subgroup will fall below the cutoff as compared with a subgroup whose mean equals the population mean. Figure 1 illustrates the challenges of using a cutoff such as 1 standard deviation below the population mean when distinct yet overlapping subgroup distributions [A] underlie the population distribution [B]. If we hypothesize that [A] illustrates three population subgroups defined on Characteristic *x* (e.g., family income, maternal education, home zip code), the −1 standard deviation cutoff (i.e., −1 on the *x* axis) will identify half of Distribution 1, about 15% of Distribution 2, and less than 3% of Distribution 3. The question becomes whether all children below the cutoff are truly language impaired and whether all children with language impairment will fall below the cutoff.

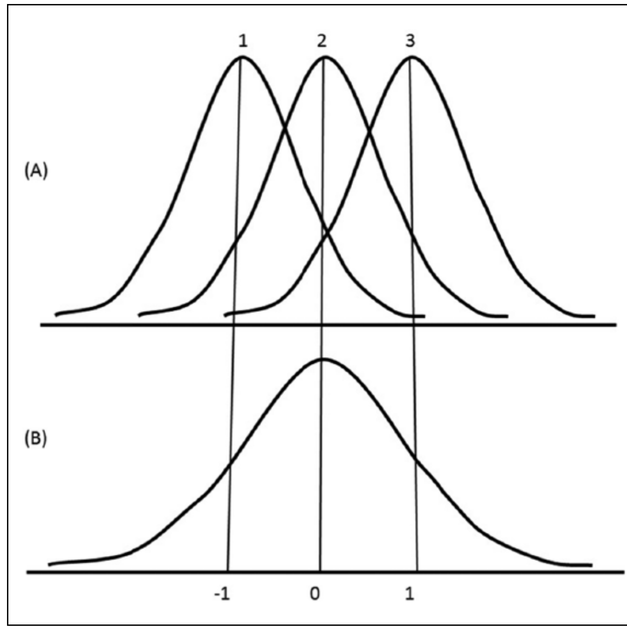

Figure 1. Distribution B illustrates the single norming sample with demarcations at the population mean (0), and one standard deviation above (1) and below (−1), a typical cutoff, the population mean.

Note. The distributions in A illustrate three hypothesized subsample distributions that underlie Distribution B. The extension of the −1 standard deviation cutoff illustrates how this typical cutoff would lead to identification of language impairment within each subsample.

The Evidence: Language Outcomes in Subgroups Defined by Maternal Education

There is substantial evidence that maternal education relates to language outcomes for preschool and early school–age children (e.g., Hoff, 2006). The general trend reveals that as maternal education increases, so do children's language scores. When participants are grouped by maternal education, group means reliably differ. Table 1 summarizes reported group means for research participants grouped by maternal education; SES group differences are evident on single language domain measures of vocabulary as well as omnibus language measures.

In Table 1, group means for *less than high school* maternal education range from standard scores of 76 to 90.57 (seven reported findings), means for *high school diploma* range from 76 to 95.46 (12 reported findings), means for *some college* range from 81 to 101 (seven reported findings), and means for *college graduate* range from to 87 to 110.74 (12 reported findings). When participants were not exclusively recruited from low SES preschools, the means for *college graduate* ranged from 101.11 to 110. For the most part, the sample *SD*s in these studies approximate the population *SD*. The data indicate a trend wherein subsample means increase as maternal education rises. The group means for children from families with less than high school or only high school graduation parental (see Note 2) education are below the population-referenced normative mean, and typically remarkably so. In contrast, the group means for children drawn from higher SES families, college-educated parents, approximate or exceed the population-based normative mean and far exceed the means of children from lower SES families.

With respect to the relation between family SES and race/ethnicity and norm-referenced measures, Qi and colleagues' research (Qi et al., 2006; Qi et al., 2003; Qi &

Table 1. Reported Group Means (Standard Deviations) for Preschool and Kindergarten Children's Performance on Norm-Referenced Measures of Oral Language Indexed by Family SES.

		Groups defined maternal education			
Study and measure(s)	Race/ethnicity	Less than high (subgroup %)	High school diploma (subgroup %)	Some college (subgroup %)	Bachelor's degree (subgroup %)
Dollaghan et al. (1999; *n* = 240) PPVT-R	AA and EA 3-year-olds	90 (15) [10%]		101 (14) [70%]	110 (14) [20%]
Horton-Ikard and Ellis Weismer (2007, *n* = 30) PPVT-III EVT PLS-4	AA toddlers		92.73 (7.24) 91.93 (5.36) 93.27 (9.63)		104.00 (12.05) 102.60 (9.06) 107.13 (10.18)
Lonigan, Burgess, Anthony, and Barker (1998, *n* = 356) Receptive Language Expressive Language	No info on middle + upper; lower mostly AA		79.09 (17.66) 86.21 (12.59) [33%]		101.11 (14.25) 110.74 (21.12) [67%]
Qi, Kaiser, Milan, Yzquierdo, and Hancock (2003, *n* = 590) PLS-3 Auditory PLS-3 Expressive	AA Head Start, low-income neighborhood child care programs	84.1 (12.3) 87.7 (12.6) [30%]	86.4 (12.0) 87.7 (11.0) [29%]	88.20 (13.70) 89.40 (12.90) [36%]	92.80 (16.50) 97.20 (18.20) [5%]
Qi, Kaiser, Milan, and Hancock (2006, *n* = 482) PPVT-III	AA Head Start, low-income neighborhood child care programs	76 [31%]	76 [35%]	81 [29%]	87 [5%]
Restrepo et al. (2006, *n* = 210) PPVT-III EVT	AA and EA public prekindergarten	77.95 (13.02) 90.57 (7.22) [10%]	89.54 (13.64) 95.46 (11.23) [57%]	91.63 (18.00) 95.17 (14.83) [11%]	103.52 (15.18) 105.50 (11.79) [22%]
Thomas-Tate, Washington, Craig, and Packard (2006, *n* = 165) EVT	AA preschool and kindergarten		94.38 (9.87) [42%]	97.90 (12.31) [58%]	
Washington and Craig (1999, *n* = 55) PPVT-III	AA attending public at-risk preschool	77.30 (10.70) [14%]	93.20 (8.80) [71%]		94.00 (12.30) [18%]

Note. The list studies are alphabetically arranged. Means and standard deviations reported as whole numbers or decimals as in publication. Not all studies reported number of participants or percentage of participants in each subgroup; where numbers were provided, we calculated and reported percent. Qi Kaiser, Milan, and Hancock (2006) provided only whole number means in a bar graph. In Qi Kaiser, Milan, Yzquierdo, and Hancock (2003), only 590/701 participants were included in subgroup means. The Bachelor's Degree subgroup would include those families with parents who have completed graduate or professional doctorate education. The participants overlap in Qi et al. (2003) and Qi et al. (2006) studies, but specific overlap was not specified by authors. Horton-Ikard and Ellis Weismer (2007) designated two groups by Hollingshead as low and middle SES; specific information on maternal education was not provided. Horton-Ikard and Ellis Weismer and Thomas-Tate et al. used Hollingshead for SES classification; we place in our maternal education classification based on the information provided in publications. Lonigan Burgess, Anthony, and Barker (1998) used a combination of the PPVT-R and Grammatical Understanding subtest of the TOLD-P (Newcomer & Hammill, 1988) as receptive language measures and a combination of the EOWPVT and Grammatical Closure subtest of the ITPA as expressive language measures. PPVT-R = Peabody Picture Vocabulary Tests–Revised; AA = African American; EA = European American; EVT = Expressive Vocabulary Test (Williams, 1997); PLS = Preschool Language Scale; PPVT = Peabody Picture Vocabulary Tests; SES = socioeconomic status; TOLD-P = Test of Language Development–Primary; ITPA = Illinois Test of Psycholinguistic Ability.

Marley, 2009) is particularly informative because they disaggregated lower SES populations by race and ethnicity. They reported comparisons for preschoolers who attended Head Start or child care centers in low-income neighborhoods from three racial/ethnic categories, European American, Hispanic monolingual English, and African American, on the Preschool Language Scale–3 (PLS-3; Zimmerman, Steiner, & Pond, 1992) and the Peabody Picture Vocabulary Test–III (PPVT-III; Dunn & Dunn, 1997). The three racial/ethnic groups performed comparably on the measures, and each group mean was substantially below the measure's population-based normative mean (Qi et al., 2006; Qi et al., 2003; Qi & Marley, 2009). The three groups' histograms for the PLS-3 and PPVT-III approximated a normal distribution. Therefore, despite the leftward shift of the group mean, the PLS-3 and PPVT-III arguably still allow for meaningful differentiation of child language proficiency, strong language skills versus weak language skills, within each subsample. That is, the low end of the subsample distribution may still be indicative of language-learning difficulties. Thus, their work suggests distinct subsample distributions as well as family SES as an explanatory factor of lower performance independent of race/ethnicity.

In sum, the performance on norm-referenced measures for child groups defined by family SES suggests subsample distributions approximate the multiple subsample distribution depicted in Figure 1A. Although each subgroup's performance approximates a normal distribution, the group means are substantially different. It follows that diagnostic and eligibility decisions that utilize traditional cutoffs with population-based comparisons (i.e., from a single population-based norming sample) will be biased toward greater identification of language impairment in children from lower SES families and lesser identification in children from higher SES families. For instance, consider the mean and *SD* of Qi et al.'s (2003) African American low SES sample on the PLS-3: $M = 86.09$, $SD = 13.79$. Based on the population distribution, we would expect approximately 7% of preschoolers to fall below the −1.5 standard deviation cutoff of 77 specified by many states for eligibility, but about 17% of Qi et al.'s sample fell below this cutoff.

The goal of our study was to consider further the issues of family SES in making eligibility decisions based on population-based norms supplied by publishers of norm-referenced measures, norms that were constructed from an aggregated sample whose participants were selected to mirror the U.S. population distribution. Toward this goal, we examined the performance of a large group of children from higher SES families, defined as parental completion of college, on an omnibus measure of language, the Preschool Language Scale–4 (PLS-4; Zimmerman, Steiner, & Pond, 2002), and a single-domain measure of vocabulary, PPVT-III. We made comparisons with the normative samples provided by test publishers and prior work that includes samples of children from lower SES families (Qi et al., 2006; Qi et al., 2003; Qi & Marley, 2009).

Method

Data were extracted from an archival database of preschoolers recruited to participate in language-development studies (Abel & Schuele, 2013; Schuele, 2006). The institutional review board at Vanderbilt University approved the study protocols.

Participants

The sample included 146 monolingual English preschool children (75 boys) recruited from six Nashville, Tennessee, preschools, housed on church grounds, and, with one exception, affiliated with a community church. Mean child age was 53.32 months (*SD* = 8.59 months, range = 36–74 months). Nearly all children were Caucasian and not Hispanic (99%, 99%). The preschool directors reported that the families served in the participating preschools overwhelmingly included college-educated parents; of families reporting maternal education, 97% had at least a bachelor's degree.

Measures

Child performance on two widely used measures of language performance was extracted from the database. The published manual for each measure presents normative tables based on age, derived from a single population-based norming sample that reflected the composition of the U.S. Census at the time of the norming study. Table 2 describes the characteristics of the norming samples, as reported in the published manuals.

Table 2. Characteristics of the Norming Sample for the PPVT-III and PLS-4.

	PPVT-III		PLS-4	
Sample characteristics	*n*	% of sample	*n*	% of sample
Race/ethnic origin				
White	1,753	64	954	62
African American	494	18	232	15
Hispanic	352	13	265	17
Other	26	5	83	5
Primary caregiver's years of education				
11 or less	465	17	266	17
12	854	31	488	32
13–15	852	31	432	28
16 or more	554	20	438	23

Note. PPVT-III = Peabody Picture Vocabulary Test, Third Edition (Dunn & Dunn, 1997); PLS-4 = Preschool Language Scale, Fourth Edition (Zimmerman, Steiner, & Pond, 2002).

The PLS-4 is an individually administered, norm-referenced, omnibus measure designed to assess the comprehension and expressive oral language proficiency of children from birth to 6 years, 11 months. The PLS-4 has two parts, Auditory Comprehension (AC) and Expressive Communication (EC). Standard scores were derived from each child's raw score on the AC and EC parts based on the normative tables reported in the manual. Per manual guidelines, the AC and EC scores (standard scores) were summed and that sum yielded a Total Language score (TL; standard score), derived from normative tables in the manual. The PLS-4 was administered only to Schuele (2006) participants ($n = 98$).

The PPVT-III is an individually administered, norm-referenced measure of single-word receptive vocabulary (i.e., single-domain measure). PPVT-III raw scores were converted to standard scores based on the normative tables in the manual. All participants completed the PPVT-III.

Procedures. The PLS-4 and PPVT-III were administered to participants by trained examiners following the administration procedures described in their respective manuals (see Note 3). For Schuele (2006), the PPVT-III typically was administered prior to the PLS-4. The PPVT-III was completed within one session. The PLS-4 subscales, however, were administered to many children on 2 separate days. For Abel and Schuele (2013), only the PPVT-III was

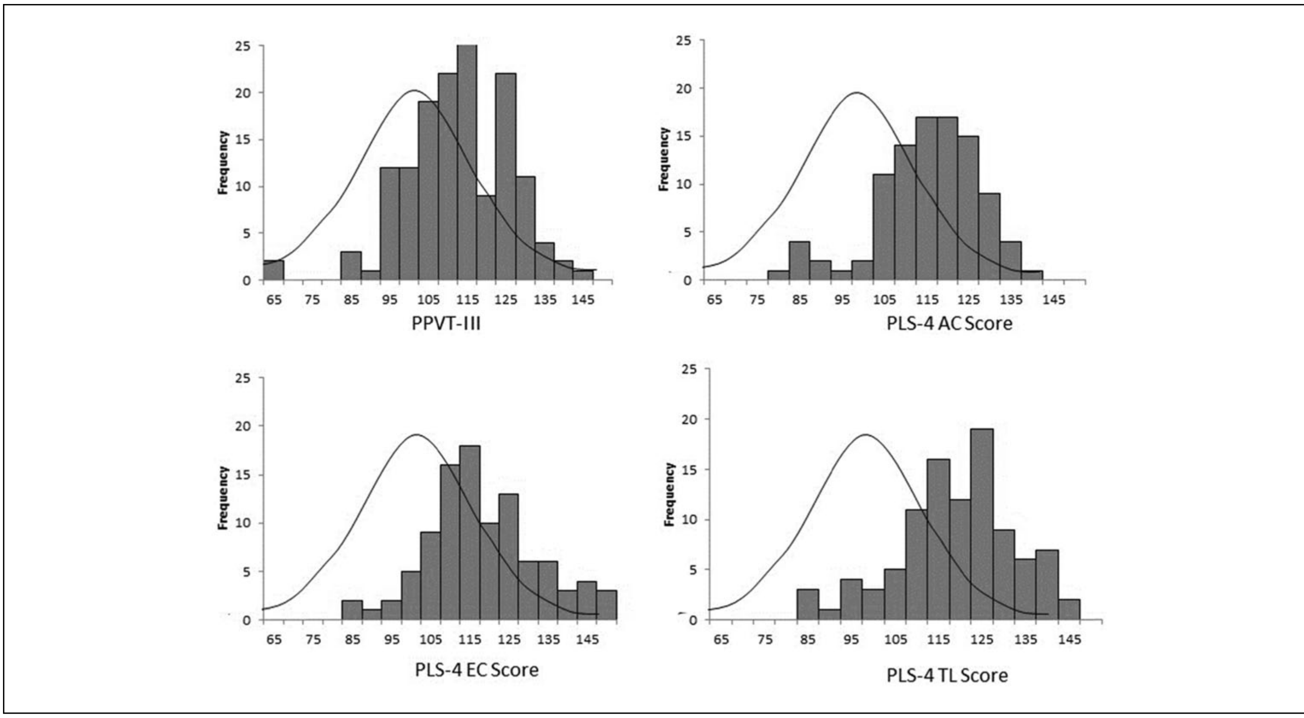

Figure 2. Distribution of standard scores on the PPVT-III, PLS-4, AC Score, EC Score, and TL Score.*Note*. PPVT-III = Peabody Picture Vocabulary Test, Third Edition (Dunn & Dunn, 1997); PLS-4 = Preschool Language Scale, Fourth Edition (Zimmerman, Steiner, & Pond, 2002); AC = Auditory Comprehension; EC = Expressive Communication; TL = Total Language.

administered, again typically in the morning and always within one session.

Child responses to the PLS-4 and PPVT-III were recorded online and standard scores were derived. Subsequently, research assistants checked all protocols for scoring accuracy (item accuracy, raw scores, standard scores). Discrepancies were resolved through mutual consensus between two research assistants; the mutual consensus scoring data were analyzed.

Data Analysis

The dependent variables of interest were standard scores. Group means, *SD*s, ranges, and distributions for the standard scores on the PLS-4 (AC Score, EC Score, and TL Score) and PPVT-III were calculated. Descriptive statistics and distribution of scores were compared with Qi and colleagues' (2006, 2003) studies and the PLS-4 and PPVT-III published normative tables. Cohen's *d* is reported as an effect size to indicate the clinical relevance of the differences between the standard scores from our sample, Qi and colleagues' standard scores, and the normative information (see Schuele & Justice, 2006, for an explanation of the use of effect sizes in the interpretation of research).

Results

Descriptive statistics for the two measures, the PLS-4 (three variables) and the PPVT-III (one variable), are presented in Table 3. The standard score group means for the PLS-4 AC Score and for the PPVT-III are at the upper end of the average range (standard scores of 85–115) as defined by population-based norms. The standard score group means for the PLS-4 EC Score and TL Score exceed 1 standard deviation above the normative mean. The *SD* for each variable is similar to the normative *SD* (i.e., 15). Variability of standard scores for the study sample for each variable was quite broad, ranging from as low as 65 on the PPVT-III to as high as 151 on the PLS-EC. Notably, there were no outlying participants, such that no one child contributed the lowest or highest standard score on all variables.

Figure 2 displays the distributions of participants' standard scores for each variable. For each variable, a large percentage of children scored more than 1 standard deviation above the normative mean: PLS-4 AC Score = 46.9%, EC Score = 45.9% and TL Score = 56.1%, PPVT-III = 33.6%. In comparison, in a population-based distribution, 15% of the scores fall more than 1 standard deviation above the normative mean. Visual inspection of the current data indicated that group performance for each variable

Table 3. Means, *SD*s, Range, and Distribution Statistics for Each Dependent Measure.

Measure	*M*	*SD*	Range	Skewness Statistic	Skewness SEM	Kurtosis Statistic	Kurtosis SEM
PPVT-III (*n* = 146)	110.58	13.24	65–143	−0.42	0.20	0.92	0.40
PLS-4 (*n* = 98)							
AC score	113.57	12.29	76–139	−0.70	0.24	0.94	0.48
EC score	116.88	14.19	83–150	0.30	0.24	0.11	0.48
TL score	116.97	13.35	80–141	−0.40	0.24	0.11	0.48

Note. SEM = standard error of measurement; PPVT-III = Peabody Picture Vocabulary Test, Third Edition (Dunn & Dunn, 1997); PLS-4 = Preschool Language Scale, Fourth Edition (Zimmerman, Steiner, & Pond, 2002); AC = Auditory Comprehension; EC = Expressive Communication; TL = Total Language.

Table 4. Comparison of Performance on the PPVT-III and the PLS: Current Study, Qi and Colleagues, Standardization Studies.

Measure	Current study M (SD)	Qi and colleagues[a] AA sample M (SD)	Qi and colleagues[a] Cohen's *d*	Qi and colleagues[a] EA sample M (SD)	Qi and colleagues[a] Cohen's *d*	Normative distribution Population M (SD)	Normative distribution Cohen's *d*
PPVT-III	110.58 (13.24)	77.87 (13.1)	2.48	81.9 (16.0)	1.95	100 (15)	0.75
PLS							
AC score	113.57 (12.29)	86.17 (12.67)	2.20	88.62 (11.41)	2.10	100 (15)	0.99
EC score	116.88 (14.19)	88.61 (12.58)	2.11	89.96 (14.3)	1.89	100 (15)	1.16
TL score	116.97 (13.35)	86.09 (12.79)	2.36	88.2 (13.24)	2.16	100 (15)	1.20

Note: The participant pool across the two studies overlaps, but the number of participants who overlap was not made clear by the authors. PPVT-III = Peabody Picture Vocabulary Test, Third Edition (Dunn & Dunn, 1997); PLS = Preschool Language Scale (Zimmerman, Steiner, & Pond, 1992, 2002); AA = African American sample; EA = European American sample; AC = Auditory Comprehension, EC = Expressive Communication, TL = Total Language.
[a]Qi, Kaiser, Milan, Yzquierdo, and Hancock (2003) reported on the PLS, and Qi, Kaiser, Milan, and Hancock (2006) reported on the PPVT-III.

approximated a normal distribution. Skewness and kurtosis statistics did not exceed 1.0, supporting the observation of a normal distribution of standard scores (see Table 3). Therefore, we conclude that, although the group means for the higher SES group are shifted to the right, distribution of standard scores around the elevated group mean is nonetheless normally distributed, allowing for differentiation of proficiency across children.

Table 4 presents the comparison of group means on the four variables from the current study's higher SES sample with the normative means published in the test manuals and with those reported for the lower SES samples in Qi and colleagues' (2003, 2006) studies. Following the convention of classifying Cohen's *d* values as small (0.2), medium (0.5), and large (>0.8), all comparisons yielded large effect sizes, with the exception of the PPVT-III comparison between the current study sample and the normative mean (M = 110.58 vs. 100, medium effect size).

Two potentially important differences between the current study and Qi and colleagues' studies are noted. First, Qi et al. (2003) used the third edition of the PLS, whereas the current study used the fourth edition of the PLS. Zimmerman and Castilleja (2005) reported similar group means and *SD*s when the PLS-3 and the PLS-4 were administered to the same children. Therefore, differences between our higher SES sample and the lower SES samples are likely not attributable to the varying editions of the PLS. Second, the participants' mean age in the current study sample was approximately 10 months greater than the participants' mean ages in the Qi and colleagues' studies (2006, 2003). The current study group mean age was 53.32 months (SD = 8.59, range = 36–74 months). The Qi et al. (2003) group mean age was 43 months (range = 36–53 months). For Qi et al. (2006), the African American group mean age was 43 months (range = 36–54 months) and the European American group mean age was 42.8 months (range = 36–51 months). Because there is no evidence to suggest that language-learning circumstances differ across the 3- to 6-year-old age period, we hypothesize that the age difference does not influence the interpretation of our findings. In addition, children in both our sample and the Qi samples were all prekindergarten. As all samples encompass children in the prekindergarten year, our assumption is that developmental expectations were similar across the samples (e.g., no one was expected to be learning to read).

Discussion

This study explored the performance of a group of higher SES preschool children on two commonly used norm-referenced language measures, the PLS-4 and PPVT-III, as compared with the single population-based norms provided by the test publishers, and in addition, with a large sample of children from lower SES families (Qi et al., 2006; Qi et al., 2003). Across both measures, the mean standard scores for our higher SES sample (PLS-4: AC = 113.57, EC = 116.88, TL = 116.97, PPVT-III = 110.58) exceeded the test mean (i.e., published standard score mean for the norming sample, 100 on the PLS and PPVT); mean scores on two variables (PLS-4 EC and TL) were greater than 1 standard deviation above the test mean. Standard deviations and skewness and kurtosis measures indicated that relative to the test mean, performance for our study sample on the PLS and PPVT was approximately normally distributed, although the group means were shifted to the right. Our sample means also greatly exceeded the lower SES group means reported by Qi and colleagues (2006, 2003), sometimes by as much 2 *SD*s.

These findings, in conjunction with the extant literature, support our prediction that subsample norms yield multiple distinct but overlapping distributions with approximately 1 standard deviation difference when subsample assignment is based on maternal education. If the low end of the population distribution is taken as evidence of language impairment (Leonard, 1991), then researchers and clinicians alike need to consider carefully what these distinct subsample distributions indicate for who is or will be called language impaired (cf. Lahey, 1990). We argue that there are sufficient data in the literature to support a call for publishers to provide subsample norms based on maternal education for norm-referenced vocabulary and omnibus language measures. Only with these subsample norms can clinicians make evidence-based interpretations of a child's performance on norm-referenced instruments as compared with his or her peer group. Reference to the performance distribution of all same-age or same-grade children (i.e., the population) will lead to very different classification decisions for some children, as compared with reference to performance of similar SES same-age or same-grade peers.

In talking with school-based clinicians, it is our impression that individual state department of education guidelines for the IDEA eligibility category of *speech or language impairment* for the most part mean that norm-referenced measures function de facto as the gatekeeper for speech–language services. If a child's standard scores fall below the state-designated cutoff for the population-based normative distribution (i.e., tests' published norms; for example, 2 *SD*s below the mean), that child is eligible under *speech or language impairment*; if not, then the child is not eligible. As such, our findings in combination with the extant literature (such as that cited in Table 2) indicate that typical cutoff decisions (see Table 1) using published norms (i.e., single aggregated population) will lead to identification of both (a) a large proportion of children from low SES homes, perhaps as great as 50%, and (b) only a very small proportion of higher SES children, perhaps as little as 1%. Although best practice espoused in clinical guidelines argues that norm-referenced measures are to be only one piece of the diagnostic or eligibility puzzle, these measures nevertheless get priority in state guidelines, a driving force behind special education eligibility decisions. Clinicians and researchers must carefully contemplate the impact of continuing the current practice of interpreting a child's performance relative to norms derived from an aggregated population sample of children. The evidence base provides convincing support for SLPs to use subsample distributions in addition to population-based distributions. Thus, when making eligibility decisions, assessment teams can consider a child's performance not only relative to the population-based distribution but also relative to his or her maternal education peer group. Obviously, to make such practices possible, publishers of norm-referenced measures will need to not only provide consumers norms based on an aggregated population sample, but also disaggregate the population and provide norms for subsamples defined by maternal education.

Admittedly, the decision to make comparisons based on subsample norms as well as single population-based norms is influenced by an individual researcher or clinician's view or definition of developmental language impairment. For SLPs who view language impairment as the low end of normal across the entire population (e.g., Leonard, 1991), irrespective of influences such as SES, population-based norms are not a problem. Any child who falls below an agreed-upon cutoff should be identified as having a language impairment. Alternatively, for SLPs who view language impairment as an inordinate, unexpected difficulty in learning language in light of the available language-learning opportunities (i.e., learning given the available language input and environment), population-based norms create a conundrum. From this alternate perspective, language-learning challenges arise more so from endogenous factors that lie within the language learner, rather than from exogenous factors such as the learner's circumstances. In this view, there is reason to believe that the rate of language impairment would be somewhat similar across SES groups; certainly, this view would argue against a 50% rate of impairment in a lower SES group and an almost nonexistent rate in a higher SES group. Known variability in input and environment across individuals, but with variation that is predictable with specific factors, favors the latter view of language impairment, supporting the call for subsample distributions from test publishers.

Ours is not the first call for subsample norms. In their investigation of low SES populations' performance on the PLS-3, Qi and colleagues (2003) recommended, "future test development efforts should assess the performance of these populations [defined by SES] and create subscale norms" (p. 589). Yet, more than10 years after their call for changes to how test norms are developed and presented, neither the PPVT-4 (Dunn & Dunn, 2006) nor the PLS-5 (Zimmerman, Steiner, & Pond, 2011; or any other norm-referenced language measures known to the authors) includes maternal education subsample norms. Moving forward, clinicians and researchers must recognize what ignoring these evidence-supported different subsample distributions mean.

Individually and as a profession, we must address what population norms mean for children, for who is found eligible under *speech or language impairment*, and for how language impairment is defined or construed in educational settings. Do we believe language impairment is rare in children from higher SES families and quite common in children from lower SES families? Is speech–language intervention the appropriate remediation model for all children from lower SES families who have limited language skills? Does poor language performance that results from difficulties learning language require the same remediation as poor language that results from limited opportunities to learn language? These are challenging questions that require serious discussions; the generation of new evidence undoubtedly will further the discussion in the future. At present, there is concern that children from higher SES families who have language impairment may not be found eligible for services; the detrimental impact of language impairment on academic, social, and vocational outcomes (Conti-Ramsden & Botting, 2004; Law, Rush, Schoon, & Parsons, 2009; Redmond, 2011; Rost & McGregor, 2012) raises huge concerns when children with language impairment do not receive services. At the same time, educators and society at-large are well aware of the adverse impact that limited language skills have on the academic achievement of children from lower SES families. Remediating these language deficits and narrowing the language gap between SES groups are important educational goals. It is critical, however, to explore whether language therapy is the best option for remediating the deficits of children from lower SES families who not only differ widely from their SES peers but do differ widely from the aggregated population mean.

Our findings quite simply bolster the argument that has been in the literature for quite some time. We believe that the evidence base indicates that as members of Individualized Education Program (IEP) teams, SLPs must frame the identification of language impairment for individual children with comparisons with peers from similar SES families. The literature (with Hart & Risley, 1995, most often cited; see also Burchinal, Peisner-Feinberg, Pianta, & Howes, 2002; Hoff, 2003; Hoff-Ginsberg, 1998) documents multiple ways in which language input and language and academic outcomes differ by family SES and associated educational differences. Assessment data from research participant samples defined by family SES further document variance in language outcomes related to family SES. In 2004, IDEA advocated radically different methods of eligibility determination for specific learning disability, based on the science of learning disabilities. The time has come to utilize new methods of eligibility determination for language impairment, based on the science of language development and disorders. Those new methods must utilize subgroup norms. It is time for publishers to provide disaggregated normative data alongside aggregated normative data that allow for evidence-based assessment.

Appendix A

IDEA Regulations for Full and Individual Evaluations and Eligibility Decisions.

Sec. 614 (b) (2) (A) Use a variety of assessment tools and strategies to gather relevant functional, developmental, and academic information . . . that may assist in determining (i) whether the child is a child with a disability; and . . . (B) not use any single measures or assessment as the sole criterion for determining whether a child is a child with a disability . . . . (C) use technically sound instruments . . . (3) . . . Each local educational agency shall ensure that (A) assessments and other evaluation materials used to assess a child . . . are selected and administered so as to not be discriminatory on a racial or cultural basis; . . . (4) Determination of eligiblity and educational need . . . (A) the determination of whether the child is a child with a disability as defined in section 602(3) and the educational needs of the child shall be made by a team of qualified professionals and the parent of the child . . . .

Sec 602(3)(a)(i) . . . The term "child with a disability" means a child with . . . speech or language impairments . . . (ii) who by reason thereof, needs special education and related services

Source. http://idea.ed.gov/download/statute.html. IDEA = Individuals With Disabilities Education Improvement Act of 2004.

Appendix B

Illustration of Cutoffs Using Norm-Referenced Comparison With Document Eligibility in Two States.

State guidelines	Criteria for language within speech or language impairment
Tennessee (Tennessee Department of Education)	• "Language Impairment—A significant deficiency not consistent with the student's chronological age in one or more of the following areas: (a) a deficiency in receptive language skills to gain information; (b) a deficiency in expressive language skills to communicate information; (c) a deficiency in processing (auditory perception) skills to organize information." • "a significant deficiency in language shall be determined by: (1) an analysis of receptive, expressive, and/or composite test scores that **fall at least 1.5 standard deviations below the mean of the language assessment instruments administered**" http://www.state.tn.us/education/student_support/eligibility.shtml
Missouri (Missouri Department of Elementary and Secondary Education)	**1500.30.a**. The evaluation report documents the results of two (2) norm-referenced and standardized language assessments which measure the same areas of language. **1500.30.b.** The evaluation determines whether the child's language functioning is significantly below the child's cognitive abilities. The following criteria apply: **1500.30.b.(1)** Children ages 3 through 5 years, not kindergarten eligible: 2 standard deviations below peers. **1500.30.b.(2)** Children who are kindergarten age eligible and older: **1.5 standard deviations below cognitive ability.** **OR** Use professional judgment with sufficient data present in the evaluation report to document the existence of a language disorder even though the criterion defined in 1500.30.b.(1) and 1500.30.b.(2) has not been met. NOTE: If unable to obtain the child's full-scale cognitive score, professional judgment must be used. http://dese.mo.gov/sites/default/files/1500-ELIGIBILITY%20CRITERIA-Speech%20and-or%20Language%20Impairment.pdf

Note. Emphasis (bold) added.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The project was supported in part by Office of Special Education Programs (OSEP 325D080075 and 325D140087) and National Institutes of Health/National Institute of Deafness and other Communication Disorders (NIH/NIDCD RO3DC007329).

Notes

1. "Culture can encompass a variety of factors including race, ethnicity, religious beliefs, age, national origin, gender, gender identity/gender expression, sexual orientation, socioeconomic levels, and disabilities" (http://www.asha.org/Practice/ethics/Cultural-and-Linguistic-Competence/)
2. We use *parental* here because not all studies reported exclusively maternal education.
3. Item 66 (two story retell sub-items) on the Preschool Language Scale, Fourth Edition (PLS-4) Expressive Communication Scale was not administered to any child; the time required for this individual item was substantial. Because our interest in the original study was to document typical language performance, we elected to not spend the time on this item. Our analyses indicated that inclusion of this item would not have substantially influenced the derived standard scores.

References

Abel, A. D., & Schuele, C. M. (2013). The influence of two cognitive-linguistic variables on Incidental Word Learning in 5-year-olds. *Journal of Psycholinguistic Research*, *43*, 1–17. doi:10.1007/s10936-013-9264-4

Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & Coll, C. G. (2001). The home environments of children in the United States part I: Variations by age, ethnicity, and poverty status. *Child Development*, *72*, 1844–1867. doi:10.1111/1467-8624.t01-1-00383

Burchinal, M. R., Peisner-Feinberg, E., Pianta, R., & Howes, C. (2002). Development of academic skills from preschool through second grade: Family and classroom predictors of developmental trajectories. *Journal of School Psychology*, *40*, 415–436.

Conti-Ramsden, G., & Botting, N. (2004). Social difficulties and victimization in children with SLI at 11 years of age. *Journal of Speech, Language, and Hearing Research*, *47*, 145–161.

Dollaghan, C., Campbell, T. F., Paradise, J. L., Feldman, H. M., Janosky, J. E., Pitcairn, D. N., & Kurs-Lasky, M. (1999). Maternal education and measures of early speech and language. *Journal of Speech, Language, and Hearing Research*, *42*, 1432-1443.

Dunn, L. M., & Dunn, D. M. (2006). *Peabody Picture Vocabulary Test* (4th ed.). Toronto, Ontario, Canada: Pearson Canada Assessment.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Fey, M. E. (1986). *Language intervention with young children*. San Diego, CA: College-Hill Press.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*, 1368–1378.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*, 55–88. doi:10.1016/j.dr.2005.11.002

Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics*, *19*, 603–630.

Horton-Ikard, R., & Ellis Weismer, S. (2007). A preliminary examination of vocabulary and word learning in African American toddlers from middle and low socioeconomic status homes. *American Journal of Speech-Language Pathology*, *16*, 381–392. doi:10.1044/1058-0360(2007/041)

Hutchinson, T. A. (1996). What to look for in the technical manual: Twenty questions for users. *Language, Speech, and Hearing Services in Schools*, *27*, 109–121. doi:10.1044/0161-1461.2702.109

Individuals With Disabilities Education Act of 2004. 20 U.S.C. § 1400 et seq. (2004).

Lahey, M. (1990). Who shall be called language disordered? Some reflections and one perspective. *Journal of Speech and Hearing Disorders*, *55*, 612–620.

Law, J., Rush, R., Schoon, I., & Parsons, S. (2009). Modeling developmental language difficulties from school entry into adulthood: Literacy, mental health, and employment outcomes. *Journal of Speech, Language, and Hearing Research*, *52*, 1401–1416.

Leonard, L. B. (1991). Specific language impairment as a clinical category. *Language, Speech, and Hearing Services in Schools*, *22*, 66–68.

Lonigan, C. J., Burgess, S. R., Anthony, J. L., & Barker, T. A. (1998). Development of phonological sensitivity in 2- to 5-year-old children. *Journal of Educational Psychology*, *90*, 294–311. doi:10.1037/0022-0663.90.2.294

Newcomer, P. L., & Hammill, D. D. (1988). *Test of Language Development—Primary: Second edition*. Austin, TX: Pro-Ed.

Qi, C. H., Kaiser, A. P., Milan, S., & Hancock, T. (2006). Language performance of low-income African American and European American preschool children on the PPVT-III. *Language, Speech, and Hearing Services in Schools*, *37*, 5–16. doi:10.1044/0161-1461(2006/002)

Qi, C. H., Kaiser, A. P., Milan, S., Yzquierdo, Z., & Hancock, T. (2003). The performance of low-income, African American children on the Preschool Language Scale-3. *Journal of Speech, Language, and Hearing Research*, *46*, 576–590. doi:10.1044/1092-4388(2003/046)

Qi, C. H., & Marley, S. C. (2009). Differential item functioning analysis of the Preschool Language Scale–4 between English-Speaking Hispanic and European American Children from low-income families. *Topics in Early Childhood Special Education*, *29*, 171–180. doi:10.1177/0271121409332674

Redmond, S. M. (2011). Peer victimization among students with specific language impairment, attention-deficit/hyperactivity disorder, and typical development. *Language, Speech, and Hearing Services in Schools*, *42*, 520–535.

Restrepo, M. A., Schwanenflugel, P. J., Blake, J., Neuharth-Prichett, S., Cramer, S. E., & Ruston, H. P. (2006). Performance on the PPVT-III and the EVT: Applicability of the measures with African American and European American preschool children. *Language, Speech, and Hearing Services in Schools*, *37*, 17–27. doi:10.1044/0161-1461(2006/003)

Rost, G. C., & McGregor, K. K. (2012). Miranda rights comprehension in young adults with specific language impairment. *American Journal of Speech-Language Pathology*, *21*, 101–108.

Schuele, C. M. (2006). *Unpublished archival database*. Nashville, TN: Vanderbilt University.

Schuele, C. M., & Justice, L. M. (2006). The importance of effect sizes in the interpretation of research. *The ASHA Leader*, *11*, 14–27.

Thomas-Tate, S., Washington, J., Craig, H., & Packard, M. (2006). Performance of African American preschool and kindergarten students on the Expressive Vocabulary Test. *Language, Speech, and Hearing Services in Schools*, *37*, 143–149. doi:10.1044/0161-1461(2006/016)

U.S. Census Bureau. (2014). *State & County QuickFacts*. Retrieved from http://quickfacts.census.gov/qfd/states/00000.html

Waldfogel, J. (2012). The role of out-of-school factors in the literacy problem. *The Future of Children*, *22*, 39–54.

Washington, J. A., & Craig, H. K. (1999). Performances of at-risk, African American preschoolers on the Peabody Picture Vocabulary Test-III. *Language, Speech, and Hearing Services in Schools*, *30*, 75–82. doi:10.1044/0161-1461.3001.75

Williams, K. (1997). *Expressive Vocabulary Test*. Circle Pines, MN: American Guidance Service.

Zimmerman, I. L., & Castilleja, N. F. (2005). The role of a language scale for infant and preschool assessment. *Mental Retardation and Developmental Disabilities Research Reviews*, *11*, 238–246. doi:10.1002/mrdd.20078

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (1992). *Preschool Language Scale*, 3rd edition. San Antonio, TX: The Psychological Corporation.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool Language Scale*, 4th edition. San Antonio, TX: The Psychological Corporation.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). *Preschool Language Scale*, 5th edition. San Antonio, TX: The Psychological Corporation.