Non-contrastive versus Traditional Language Assessment in General-American- and African-American-English Speakers

Robinson and Stockman (2009) and Norton (2008) report that speech-language-pathologists (SLPs) often find social dialects like African-American English (AAE) hard to understand, and express uncertainty about their ability to judge impairment in children who use them. Traditional diagnostic methods, which include *contrastive* forms, i.e. those that differ in General-American English (GAE) and other varieties, are not appropriate for assessing such non-GAE speakers (Craig & Washington, 2006; Seymour, Bland-Stewart, & Green, 1998). Therefore, these authors and others recommend diagnostic methods that target features of English less susceptible to influence from dialectal varieties (*non-contrastive* features) for identifying language impairment in speakers of varieties like AAE. This paper investigates disparities between traditional methods (non-contrastive and contrastive elements combined) and exclusively non-contrastive diagnosis methods for speakers of different varieties of English: AAE as well as GAE. We hypothesized that diagnoses of impairment or typical development derived from traditional versus non-contrastive methods would differ for all children, but would be more disparate for those whose dialect was more different from GAE.

To test these hypotheses, we used a retrospective quasi-experimental design to compare diagnoses made with traditional methods to diagnoses from an assessment comprised of only non-contrastive elements, and confirmed by language-sample analyses. The major independent variable was dialect, AAE or GAE, and within AAE, high or low dialect-density.

The dependent variable was disparity between the two methods of diagnosis in the designation of clinical status.

Participants were 720 AAE-speakers with a range of dialect densities and 280 European-American GAE-speakers, ages 4 to 12, who took part in fieldtesting for the *Diagnostic Evaluation of Language Variation* tests *(DELV*, Seymour, Roeper & deVilliers, 2000, 2003, 2005). Over 400 certified SLPs participated in data-collection, and recruited children according to specific design variables: *dialect*, judged as AAE or GAE by teachers and clinicians; and *clinical status*, typical-development (TD) or language impairment (LI), according to whether the child was receiving speech-language services in their schools. Since data collection took place prior to publication of the *DELV-Norm-Referenced (DELV-NR*, Seymour et al., 2005), the first standardized test based on exclusively non-contrastive elements, it was stipulated that prior instruments used by the SLPs to identify LI children for fieldtesting included some contrastive elements.

Materials: Participants took the 350-item pilot-version of the *DELV* tests. Items eventually selected for inclusion in the *DELV-Screening-Test (DELV-ST) and the DELV-NR* were extracted from the fieldtest database of responses and scored according to the published guidelines for the tests (which were available by the time the present analyses were made).

Procedures: Data were analyzed by cross-tabulation and χ^2-statistics with the SLPs' pre-existing diagnosis of clinical status from traditional methods of their practice ("traditional-TD" or "traditional-LI") along one dimension and clinical status according to non-contrastive methods indicated by scores on the *DELV-NR* and confirmed by language sample analysis along the other dimension. Separate analyses were performed for AAE and

GAE speakers. Among AAE-speaking children, there was further analysis by dialect-density categories as defined by the *DELV-ST* Part 1 ("*no difference*" versus "*some* or *strong difference* from GAE"—here abbreviated, low- or high-density). Transcriptions of concurrent language samples from a subset of participants served as confirmation of the non-contrastive diagnoses (Magaziner, K., Pearson, B., & deVilliers, P., 2008 and in submission).

Results: Overall agreement between the diagnoses yielded by the two methods (traditional versus non-contrastive) was 82.5% for the GAE-speaking group, giving a baseline for the comparison of the methods in the GAE population. In the AAE-speaking group, the percentage of agreement was 73%, a significant difference between dialect groups by χ^2 (1, N=1000) = 9.51, p=.002. Within the AAE group, agreement was higher for low-density than high-density children, 87% versus 70%: χ^2 (1, N=720) = 15.24, p<.0005.

For all children, there were higher rates of "over-identification" (over-ID) by the traditional diagnosis relative to non-contrastive diagnostic methods than under-identification (under-ID) as follows:

OverID:

> Of the 152 AAE- and 61 GAE-speaking children in the sample who were receiving services (traditional-LI), 45% of AAE-speakers and 39% of GAE-speakers scored higher than 1.5 standard deviations below the *DELV-NR* mean, indicating typical development ("noncontrastive-TD"). The difference by dialect group was not significant, p=.473.

UnderID:

By contrast, among the 568 AAE- and 219 GAE-speaking children not receiving services (traditional-TD), 22% and 11% respectively failed the *DELV-NR* ("*non-contrastive*-LI"), indicating possible under-ID. AAE under-ID was significantly different from GAE under-ID: χ^2 (1, N=787) = 11.49, p=.001.

Next, under- and over-ID within the AAE group were analyzed by dialect density. As shown under "overall agreement" above, the diagnoses were less likely to be discrepant overall for the low-density group. Similarly, the AAE low-density traditional-TD experienced significantly less under-ID than the traditional-TD high-dialect group, 11% versus 18% (p=.001). Over-ID between AAE density groups could not be evaluated statistically because there were only five low-density AAE children receiving speech-language services. Nonetheless, since three of those five children passed the *DELV-NR* (i.e. were noncontrastive-TD, or over-identified), identification of impairment appeared equally problematic for low-density speakers.

Therefore, our hypotheses were supported. The two methods gave significantly disparate diagnoses for all children, AAE and GAE groups, but they were more disparate for AAE children than GAE. Within the AAE group, diagnoses for high-density speakers were more disparate than for low density. Crucially, regardless of dialect density, 45% of AAE children in our sample receiving speech-language services in their schools were classified as typically-developing by an assessment based on only non-contrastive elements. Notably, similar over-ID was found for the GAE speakers, who were also over-identified at only slightly lower rates (39%).

Limitations: One might be tempted to conclude from these data that the non-contrastive method differed dramatically from traditional assessment methods because the

DELV-NR might not have been sensitive enough to find language impairment. However, the greater agreement of the *DELV* diagnosis of clinical status with measures from concurrent language samples argues against that conclusion (Magaziner et al., 2008).

Clinical implications: These results provide evidence to support the use of non-contrastive, dialect-neutral measures for all children—GAE speakers and low- and high-density AAE speakers.