

2018

The Legacy of the Diagnostic Evaluation of Language Variation (DELV)

Barbara Zurer Pearson
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/aae_summary



Part of the [Applied Linguistics Commons](#)

Pearson, Barbara Zurer, "The Legacy of the Diagnostic Evaluation of Language Variation (DELV)" (2018).
T.O.M and grammar (Thoughts on Mind and grammar). 2.
Retrieved from https://scholarworks.umass.edu/aae_summary/2

This Article is brought to you for free and open access by the NIH Working Groups on African American English (AAE) at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Summary Documents by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

2018

The Legacy of the Diagnostic Evaluation of Language Variation (DELV©)

Barbara Zurer Pearson

The Legacy of the Diagnostic Evaluation of Language Variation (DELV®)

Barbara Zurer Pearson

University of Massachusetts Amherst

1. Introduction

The Diagnostic Evaluation of Language Variation (DELV) project comprises three tests—the DELV Screening Test (Seymour, Roeper, and de Villiers 2003a, 2018), the DELV-Criterion Referenced (Seymour, Roeper, and de Villiers 2003b), and the DELV-Norm Referenced (Seymour, Roeper, and de Villiers 2005, 2018). They are the first, and so far, the most comprehensive linguistically sophisticated and culturally fair standardized language assessments on the market. They allow all speakers of English, including speakers of the AAE variety, to demonstrate competence in fundamental language elements and processes, and they provide linguistically-appropriate markers of impairment. Thanks to the unique team and the diverse expertise that contributed to their design and implementation, they stand out as ground-breaking both scientifically and socially. In this paper, I reflect on the confluence of ideas and circumstances that brought the DELV team together and created the tests. I then survey the continuing legacy of the tests in the first decade after they were published.

Without sacrificing their very practical goals, the development and implementation of the DELV tests was first and foremost a research endeavor which addressed issues in theoretical and applied linguistics in the service of social justice. Indeed, the DELV's creation allowed more comprehensive coverage of existing research topics and extensions to new questions about typical and disordered acquisition of different varieties of English. Once published, the DELV tests became both a tool and a focus for further research in the associated disciplines of linguistics, psychology, and communication disorders.

1.1 The development of the DELV initiative

The DELV initiative combined the lived experience and ambitious vision of Harry Seymour, the passion and explosive intellect of Tom Roeper, the unique blend of brilliance, creativity, and extraordinary efficiency of Jill de Villiers, and the gentle wisdom and attention to detail of Peter de Villiers. It also relied on a talented generation of graduate students in communication disorders and linguistics to keep it grounded in the twin realities

Barbara Zurer Pearson

of scientific pursuit and clinical practice. Harry excelled at recruiting and mentoring graduate students from underserved groups. He was also able to support and nurture them through a series of Department of Education training grants, starting in the 1980s (Seymour and Seymour 1988) and culminating in two rounds of National Institutes of Health (NIH) funding which carried through to the publication of the DELV-NR in 2005 (R01-DC-02172 and N01 DC-2104). Harry's early background with a B.S. in business administration, plus connections through the enterprising intercession of ASHA colleague Juanita Doty, led him to make a shrewd subcontract with The Psychological Corporation (TPC), a major test publisher with the national reach and production infrastructure that made publication possible as soon as the research was accomplished. From TPC president Aurelio Prifitera to Lois Ciolli, their dedicated and professional Research Director throughout the 10 "NIH years," the unusual academic-industry partnership helped guide the project to its timely completion (on time and under the NIH budget). National experts hosted by the project and a steady stream of international visitors continually brought new viewpoints to the project and insured that the ideas incorporated in the tests represented the cutting edge of language science (de Villiers 2003). It was my good fortune, on the basis of a random job posting on the CHILDES listserv, to find a niche in that nurturing community and I have enjoyed an inside seat as the DELV project unfolded. As Project Manager (and now co-author), I have the honor of having been acknowledged as the "dedicated, imaginative, and energetic glue" (Seymour 2003:iv) that kept everyone going in the same direction through the NIH contract and beyond.

Harry says the DELV tests were 30 years in the making. I count 29. As best I could find, the names of the four Principal Investigators first appeared in the same volume in 1976. Harry and Tom, one a new assistant professor in Communication Disorders and the other a new assistant professor in Linguistics, were on the dissertation committee for Linda Gillum, an African-American graduate student in Harry's department. Her project was perhaps the team's first of many studies on the developing syntax and morphology of children ages 3 to 6 years old, racially balanced between black and white (Gillum 1976). The topic leaned heavily on the work of Harvard psycholinguist Roger Brown, so it is no surprise that de Villiers and de Villiers (1973), as star graduates in Brown's lab, were cited heavily in the rationale and method¹. Still, it took 10 years or more for the four scholars to actually work together.

For Harry and Tom, their experience on Linda Gillum's committee was the start of a friendship and an ongoing tennis match lasting more than 20 years. Their personal interactions turned increasingly professional as the students Harry recruited began to incorporate more and more linguistics coursework in their degree programs—co-supervised by Tom. It is likely that the de Villiers and Tom were known to each other as they came out of the same cradle of modern psycholinguistic research centered around Harvard and MIT. But the de Villiers did not come to Smith and the Five College consortium until 1979, and it took a few more years before Tom pulled Jill and Peter into the nascent DELV circle. Tom and Jill have no significant joint publications until they wrote a successful grant proposal on *wh*-acquisition together, funded by the National

¹ Albeit with first initials R and E (?) instead of J.G. and P.A.

Science Foundation (NSF) from 1989 to 1992. Their first publication together was the seminal article with Vainikka in 1990 (de Villiers, Roeper, and Vainikka).

Further, another collaboration around a student was instrumental in cementing the PIs' creative and productive relationship. Toya Wyatt, Harry's student, also mentored by Tom, spent the year before defending her dissertation in 1991 at Smith College as a Mendenhall Fellow. Toya's dissertation is signed by Harry Seymour as chair, Tom Roeper and Peter de Villiers as committee members (and Charlena Seymour as department head). Jill de Villiers figures prominently in the acknowledgment and we see there, too, that many of the individuals who will become associated with the DELV project were already in place: notably Jessica Harris, Ovetta Harris, Linda Bland, Lisa Green and Eliane Ramos, whom Toya singled out, and who in their turn were mentored by what would become the DELV team. They paved the way for other minority candidates to follow, Tempii Champion, Janice Jackson, D'Jaris Coles, Valerie Johnson, Lamya Abdul-Karim, J. Michael Terry, Frances Burns, Tim Bryant and several masters-level students.

1.2 Historical background: The national context

The fundamental insight of the DELV authors was to bring a comprehensive linguistic perspective to clinical language assessment. The DELV tests respond to the need made urgent by the seminal research of the 1960s when researchers like Labov (1970) and Wolfram (1969) showed the rule-governed, linguistic basis for what they called African-American Vernacular English (AAVE), or alternately Black English, African-American English (AAE), or Ebonics (Williams 1975). Thirty years of research and advocacy, beginning with the Black Caucus of the American Speech-Language-Hearing Association (ASHA) led by Orlando Taylor (and including a young Harry Seymour, Charlena Seymour, Ida Stockman and others), showed the short-comings of existing assessments based on developmental patterns of speakers of mainstream American English (MAE).

The bedrock of the DELV rationale was the *contrastive* versus *non-contrastive* distinction (Seymour and Seymour 1977, Seymour, Bland, and Green 1998, Green 2002) applied to elements of the two varieties of English, AAE and MAE. Areas that diverge between the varieties, for example the present tense verb paradigm, are *contrastive* and not suitable for inclusion in diagnostic materials, especially for children with AAE background below 8 years of age (Jackson and Pearson 2010). Rather, DELV diagnoses of risk rely only on *non-contrastive* elements, those that are shared between AAE and MAE and represent a similar challenge to the child, regardless of dialect. Academic activists advocating for fair testing for speakers of AAE emphasized the need for a new approach, not just adaptations of current materials, such as questionable score modification with tests already deemed to be insufficient (See Stockman 2010 and Pearson, Jackson and Wu 2014, for reviews).

Co-PIs Tom Roeper and Jill de Villiers are internationally recognized leaders in the empirical project to bring insights of theoretical linguistics to bear on child language acquisition and vice versa (de Villiers 2003, Roeper and Seymour 1994, Roeper et al. 2001, Roeper 2007). The close collaboration with Seymour, a leading figure in the field of Communication Disorders, brought a modern linguistics approach to the practical problem of fair language assessment. Their partnerships, embodied in the NIH Working Groups, focused on areas from their psycholinguistics research in the theoretical framework of

Universal Grammar—e.g. *wh*-barriers, quantification, and functional categories—amenable to assessment within known research and clinical paradigms (de Villiers and Roeper 1996, de Villiers 2003). With an NIH R01 grant followed by the NIH contract to create the DELV, the test development team also turned to other less universal facets of language—such as passive constructions, narrative cohesion, and process measures of lexical and phonological development—to find elements which could be shown empirically not to differ between the better-understood mainstream variety of English and AAE (Seymour and Pearson 2004). The tests do not so much capture linguistic features unique to the AAE variety—for example its rich verbal aspect and negative concord systems. Rather, focus on AAE-specific elements is confined to researching the most effective identifiers of AAE speech, which are incorporated in the DELV screening tool to establish Language Variation Status (LVS), but *not* for diagnosis of risk. The LVS screener emphasizes the clear disjunction between identifier elements and those suitable for diagnosis of impairment.

2. The Publication of the DELV

The “deliverables” of the NIH Working Groups’ contract accomplished the triple goals of 1) quantifying development between ages 4 and 9 years for a novel set of language components with a population only marginally included in prior language norms, 2) distinguishing clinical status, i.e. typical development versus impairment, but 3) showing no performance difference between speakers of the mainstream dialect and AAE. A collateral benefit of the validation process for the DELV is the spotlight of validation and respect it shines on AAE and its users. The DELV’s publication reinforces the team’s fundamental objective to achieve social justice. It gives children who speak a stigmatized variety of English a means to show their excellence as language learners and users and to have their developmental progress judged by dialect-sensitive standards.

The tests are based on extensive piloting, first in the region around the University of Massachusetts Amherst, then around the nation, using the preliminary Dialect Sensitive Language Test (DSLIT, Seymour, Roeper, and de Villiers 2000). This superset of the DELV had 350 candidate items that were administered to over 1,500 children, ages 4 to 12 years, 2/3 African-American and 1/3 European-American matched for age, gender, parent education, region, and clinical status. Children receiving speech-language services were over-sampled (30%, compared to the less than 10% expected in any population to show language delay, [Tomblin et al. 1997]) so that meaningful statistics could be done to compare clinical groups. Children with lower socio-economic status (SES) from both ethnicities were also oversampled to maximize the number of participants who spoke AAE and to avoid SES asymmetries in the data analyses. Based on outcomes on the DSLIT, a selection of 32 items were adopted for the Screening Test (ST), 15 dialect identifiers and 17 diagnostic items for preliminary screening for risk, and 125 items in four domains--syntax, pragmatics, semantics and phonology--were chosen for the full DELV diagnostic instrument. The key to item selection, except for the items chosen for the LVS screener, was the “3 Ds,” that is, their superiority in showing Development and Disability, with no Difference between AAE and MAE participants.

The DELV-ST was published in 2003² with separate criteria for Language Variation Status and Diagnostic Risk Status established from the DSLT data. The full DELV was also published first as a criterion reference test (CR) based on the DSLT. The DELV-CR (Seymour, Roeper, and de Villiers 2003b) subsequently underwent two further rounds of nationwide standardization testing for publication as the DELV-Norm Referenced (NR, Seymour, Roeper, and de Villiers 2005, 2018). The NIH contract called for norming the DELV-NR on a 100% African-American representative norming sample. Then, to emphasize the suitability of the DELV-NR for speakers of MAE, a second norming sample was recruited to match the general American population according to the 2000 U.S. census. When it was confirmed that the two sets of norms did not differ, the test was published with the norms based on the general American population.

2.1 Supplementary materials produced from the process of creating the DELV

2.1.1 Data from the stages of DELV development

The program of empirical research that suggested the content of the tests and then developed the norms for their use created a mountain of data, for which access is available through the authors. Broadly speaking, it consists of two extensive coded corpora of AAE child language samples and a database of responses and response profiles from the three standardization phases. One set of transcribed videotapes records 24 typically-developing African American five-year-olds from Hartford, CT in three settings, adult-child, peer-to-peer, and narrative, that were collected for the preliminary research during the initial NIH grant. The second set of 78 audiotapes and transcripts with concurrent DELV Screening Tests and the Norm-Referenced tests, were recorded for the final report to NIH: the children were African-American 5- and 6-year-olds from around the country. Twenty were receiving services for language impairment at the time of the recording (but a slightly different set were evaluated by the DELV-NR as impaired). The database of responses from TPC, scored and coded by the authors, records all short-answer response data for each participant, plus a supplementary corpus of approximately 120 phrasal and sentential responses to open-ended questions from the 1,500 children, ages 4 to 12 years who took the DSLT. These “verbatim” responses were written on the record forms on the spot by the examiners. Although there were fewer than 100 tapes of test administrations for reliability purposes, nonetheless, the verbatims have been found to give a useful, fine-grained snapshot of how the children reacted to the test stimuli.

The files, in text and excel formats (and mp4 for the original Hartford tapes), can be read with ordinary software or analyzed with CHILDES CLAN programs (MacWhinney 2016), and are available from the authors.

2.1.2 Subsequent studies

² We list The Psychological Corporation (TPC) as the publisher who saw the project through the extended development phase, but by 2003, TPC had merged with its parent company and so the title page also lists Harcourt Assessments. In 2008, Pearson Inc. acquired Harcourt and maintained the copyright through 2017, when copyright was returned to the authors. As of 2018, the publisher is Ventris Learning of Sun Prairie WI. P. de Villiers and Pearson were added as co-authors in 2018.

The research undertaken for the creation of the DELV, some of it in the dissertations associated with the project and in associated experiments, did not end with the publication of the tests. Research with and about the DELV continues, by members of the original Working Groups and others around the world, and adds to the legacy of the DELV.

Studies following the publication of the DELV can be loosely divided into three broad categories: 1) further validation of the DELV tests; 2) expanding the lessons learned about the test elements, such as double-*wh* questions, articles, and fast-mapping; and 3) new uses for the DELV, for example to inform test development in other languages or with other populations, or in the creation of new corpus tools designed to incorporate the insights of the DELV project. The bibliography of DELV-inspired research is also archived in Scholarworks.

2.2.1 Further test validation research

The final report for the NIH contract (2004) and the text of the Technical Manuals created by the publisher are by no means the last work to establish the validity and efficacy of the DELV tests. In 2004 in a special issue of *Seminars in Speech*, the [original] DELV authors—Seymour, Roeper, and J. de Villiers--and colleagues P. de Villiers, Pearson, and Ciolli provided a comprehensive overview to understand the concepts and context of the DELV project (Seymour and Pearson 2004). Two later studies, de Villiers and de Villiers 2010 and Pearson, Jackson, and Wu 2014, correct a significant error in the NR Technical Manual (which arose because the authors did not control the final content in the manual).

The Manual presents an accessible rationale for laypeople about the design of the DELV-NR's components and effectively helps clinicians administer and interpret the test according to the intentions of the authors. However, the psychometric analyses by the Harcourt statisticians in Section 7 of the Manual were based on a logical fallacy that failed to acknowledge the very basis for the test. The calculations were based on evaluation of how well DELV scoring recapitulated the prior diagnoses of clinicians using assessments that predated the DELV and that the DELV was deliberately designed to supersede. The solution arrived at by the Harcourt team declared the DELV to be most accurate identifying children as impaired with a criterion that does not match most state-issued guidelines for providing services to children: that is, scoring only 1 standard deviation (SD) below the mean. The NIH Working Groups' final report, the basis for de Villiers and de Villiers 2010 and Pearson, Jackson, and Wu 2014, showed good to excellent diagnostic accuracy relative to concurrent language samples, using the more conventional 1.5 SD below the mean as the criterion for impairment. Thus, the newer analyses bring the DELV into stronger alignment with current best practice.

Articles by Horton and Apel (2014), Petscher, Connor, and Al'Otaibi (2012), Terry, Petscher, and Rhodes (2017), and Gregory and Oetting (2018) further validate the scoring of the DELV tests.

2.2.2 Lessons learned about the components of the tests.

Well after the NIH funding ended, we continue to study of the trove of data generated by the standardization process, and new experimentation on the topics continues. This work

uses the DELV not just for the binary diagnostic question—impaired or not impaired—but explores the empirical data developed for the diagnostic purpose to tell about the test components across age, gender, dialect, socio-economic status (SES), and clinical groups. These articles are also useful to situate the DELV findings within the context of related research. In addition, explorations of the data probed for linkages between the elements of the different subtests, like double-*wh* questions and quantifiers (Strauss et al. 2003), Theory of Mind and *wh*-movement across embedded clause boundaries (P. de Villiers, Burns, and Pearson 2003), or morphosyntax and passives (Pearson and Roeper 2004).

In sum, many other studies and teaching exercises were developed from DSLT data and they continue to be a rich source for both purposes.

3. Broadening the reach of the DELV

Now that the DELV tests and a large part of the data from the test development have been available for a dozen years, the DELV authors, along with other scholars and practitioners have used DELV principles and materials for new purposes, especially 1) creating similar tests for other languages, 2) using the DELV with new populations, and 3) incorporating DELV data and insights in the creation of new corpus tools. (These materials are also referenced in Scholarworks.)

3.1 The DELV as a model for other linguistically-informed tests and tools

The DELV has provided the framework and the example for linguistically-informed test development in several other languages and for younger ages. Jill de Villiers, in particular, has brought her expertise from the DELV to many other assessment initiatives. She has worked closely with researcher-clinicians of Arabic and Romani (Abdul-Karim and de Villiers 2017, Kyuchukov, de Villiers, and Takahesu-Tabori 2016) to find and incorporate linguistic constructions in those languages that would be effective for clinical assessment. She has also participated with teams at Temple University and University of Delaware to adapt many of the DELV paradigms and principles for use with younger children (QUILS, Golinkoff et al. 2017, de Villiers et al. 2017) and bilingual children (Aravind et al. 2013), and her deep knowledge of the DELV and its psychometrics contributed to the development of a preschool assessment for Mandarin (Liu et al. 2017, Ning, Liu, and de Villiers 2014).

Likewise, researcher-clinicians Southwood and van Dulm of Stellenbosch University adapted the DELV for South African English. They also experimented with using some of the universal elements and processes in the English DELV with their Afrikaans-speaking preschool and elementary students for whom they had no culturally and linguistically sensitive assessments (van Dulm and Southwood 2008, Southwood 2013). They published several easel books of practice items (Southwood and van Dulm 2012) and have created ingenious, inexpensive practice worksheets for some of the items which fold into little 4-page “books” that even the most underserved children can take home and share with their families. At the time of this writing, Southwood is hopeful that the new status of the copyright, now with the authors, will permit the publication of an English-Afrikaans DELV.

The model of the DELV was also the basis of two projects funded by the European Union (EU) agency for Cooperation in Science and Technology (COST), which encompassed several domains and still more languages. Roeper's many collaborations with EU scholars inspired the first COST action A33, "Cross-linguistically robust stages of children's linguistic performance," that ran from 2006 to 2010 under the direction of Uli Sauerland (2010), and involved 35 participating countries. It organized working groups in five broad areas of syntax and created a framework for expansive cross-linguistic inquiry that could inform assessment. One spin-off was another large COST action, "Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment," funded from 2010 to 2014 and led by A33 participants Armon-Lotem of Israel and de Jong of the Netherlands (Armon-Lotem, de Jong, and Meir 2015). It involved even more EU and partner countries, reaching as far as the Russian Federation, South Africa, and India. The base of cross-linguistic benchmarks of development across languages gathered by A33 created a pathway to bilingual assessments in the form of both parallel monolingual tests, which permitted greater comparability of development across languages, as well as tests for specific language pairs being learned by significant populations of children in EU countries.

3.2 Uses for other purposes and populations

Many of the DSLT and DELV items have found more detailed explanation in the *Prism of Grammar*, a book on language acquisition written by Roeper (2007) for the general public, especially parents and teachers. The whole book is permeated with a sense of respect for language diversity and for the genius all children show in their emerging language. In addition to a chapter on "The Riches of AAE," several item types from the DELV have become "Explorations." These are simple, fun activities to do with children that center around points of abstract grammar that Roeper brings to life in scenarios with everyday sentences and situations.

Although funded for and tested specifically with AAE speakers, the DELV has shown that it is not exclusively an AAE test. Crucially, it was normed on a general American population of mostly MAE speakers, as noted above in section 2. Thus, it is not a surprise that it has been shown to be useful with MAE speakers and other populations, such as Southern White English (SWE) and Cajun English (Riviere, Oetting, and Roy 2018). Similarly, Marinis et al. (2013) and Norbury and Sparks (2012) and others have used its Pragmatics section with autistic children, and it has been adapted for investigating academic readiness in a classroom with diverse African languages in Johannesburg (Kallenbach 2007).

(Note that when using the DELV in new contexts, the results must be interpreted cautiously. If students do well on these challenging items, that is valuable information. However, if the child does not do well, one must probe for other factors that might influence performance, for example with English-Language Learners, possibly insufficient time spent with exposure to English, or particular structures such as double-*wh* questions or quantifiers, that may follow different constraints in the child's first language.)

3.3 Use in the development of new corpus tools

The impact of the DELV data has been expanded by Pearson's participation in the KIDEVAL project led by Bernstein Ratner of the University of Maryland and MacWhinney of Carnegie Mellon, funded by NIH in Fall 2018. Kideval's goal is to integrate the tools for collection, analysis, and normative interpretation of preschool language samples (through age 5;11), aiming to make them user-friendly enough to be incorporated into clinicians' normal routines. To date, language sample analysis (LSA), often recommended as a supplementary assessment for underrepresented children (Pearson, Jackson, and Wu 2014), is prohibitively labor-intensive and has only rudimentary psychometric data to use in distinguishing typical from impaired performance. There have been isolated attempts to adapt existing tools for LSA for varieties like AAE (e.g. for IPSyn, Oetting et al. 2010, for DSS, Nelson and Hyter 1990), but nothing that can be easily integrated into the general process nor that helps speed transcription. Kideval will generate age-graded reference values for major clinical LSA measures and is committed to identifying possible dialect, race, or SES performance differences in LSA measures across early childhood that often influence referral rates for services. Like the DELV project, Kideval targets the gap between language research and clinical practice. Its foundation in the CHILDES system (www.childes.psy.cmu.edu) offers both the data and the computational resources to address this critical gap. The added DELV language sample data made available for its development help expand the impact of both Kideval and the DELV.

4. On the DELV's contribution to social justice

Four scientists committed to linguistic equality, a funding agency giving up the wait for an omnibus "fair test," and a large corporation willing to undertake a risky project because it was the right thing to do, together, took a *revolutionary* step forward in combatting language prejudice. They have given professionals tangible, user-friendly tools to join the challenge. The tests represent a practical application of the abstract ideas of language science to help people recognize the integrity and beauty of AAE as a variety of English, and they emphasize that AAE has well-formedness principles analogous to rules and constraints they may be familiar with from descriptions of the mainstream dialect.

The change in perspective about AAVE from the 1960s to the present, from deficit to difference, is astounding. (See Stockman 2010 for a review.) However, despite official recognition of the legitimacy of AAE as more than "slang" or a pidgin form of mainstream English (NCTE 1974, ASHA 1983, 2003, APA 2002), linguistic prejudices remain strong (Pearson, Conner, and Jackson 2013, Norton 2008). The DELV still stands as an outlier in the field of Communication Disorders (Stockman, Boulton, and Robinson 2008).

Language prejudice is especially entrenched, as language is so deeply embedded in people's minds and emotions. The DELV is a valuable tool in the continuing effort to educate educators and other language professionals, and perhaps especially the general public—as deeply-rooted ideas require *evolutionary* change for the sweeping cultural change needed. The goal has been envisioned and articulated. The DELV authors ask impatiently, "What are we waiting for? We know what we need to do; let's do it."

Reference List

- Abdulkarim, Lamya, and Jill G. de Villiers. 2017. Development and validation of a speech and language test for Saudi Arabic children. Paper presented at the American Speech-Language-Hearing Association, Los Angeles, CA, 9-11 November.
- American Psychological Association (APA). 2002. Guidelines on multicultural education, training, research, practice, and organizational change for psychologists. Available on the APA website, <http://www.apa.org/pi/oema/resources/policy/multicultural-guideline.pdf>.
- American Speech-Language-Hearing Association (ASHA). 1983. Social dialects [Position statement]. Available at www.asha.org/policy.
- American Speech-Language-Hearing Association (ASHA). 2003. American English Dialects [Technical Report]. Available at www.asha.org/policy.
- Aravind, Athulia, Jill de Villiers, M. Freeman, and Aquiles Iglesias. 2013. Assessing bilingual 3-5-year-olds. Paper presented at UConn UMass Smith Language Acquisition Workshop (UUSLAW), Smith College, April.
- Armon-Lotem, Sharon, Jan de Jong, and Natalia Meir. 2015. *Assessing multilingual children: Disentangling bilingualism from language impairment*. Bristol, UK: Multilingual Matters.
- de Villiers, Jill G. 2003. Defining SLI: A linguistic perspective. In *Language competence across populations: Toward a definition of SLI*, ed. by Yonata Levy and Jeanette Schaeffer, 425-447. Mahwah, NJ: Erlbaum.
- de Villiers, Jill G., and Peter A. de Villiers. 1973. Development of the use of word order in comprehension. *Journal of Psycholinguistic Research* 2 (4), 331-341.
- de Villiers, Jill G., M. Gonzalez, E. Jackson, T. Saunders, R. Luo, R. Alper, Roberta M. Golinkoff, Kathy Hirsh-Pasek, and Aquiles Iglesias. 2017. Innovations in a touchscreen test of language comprehension for 2-year-olds. Paper presented at the American Speech-Language-Hearing Association (ASHA). Los Angeles, CA, 9-11 November.
- de Villiers, Jill G., and Thomas Roeper. 1996. Questions after stories: Supplying context and removing it as a variable. In *Methods for assessing children's syntax*, ed. by Dana McDaniel, Helen Cairns, and Cecile McKee, 163-188. Cambridge MA: MIT Press.
- de Villiers, Jill G., Thomas Roeper, and Anne Vainikka. 1990. The acquisition of long-distance rules. In *Language processing and acquisition*, ed. by Lyn Frazier and Jill G. de Villiers, 257-297. Dordrecht: Kluwer.
- de Villiers, Peter A., Frances Burns, and Barbara Z. Pearson. 2003. The role of language in the theory of mind development of language impaired children: Complementing theories. In *Proceedings of the 27th Boston University Conference on Child Development*, ed. by Barbara Beachley, Amanda Brown, and Frances Conlin, 232-242. Somerville, MA: Cascadilla Press.
- de Villiers, Peter A. and Jill G. de Villiers. 2010. Assessment of language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science* 1 (2), 230-244.
- Gillum, Linda H. 1976. Children's performance on judgements of grammaticality in word order, morphology, and sentence structure. Doctoral dissertation, University of Massachusetts Amherst, Amherst, MA.

The Legacy of the DELV

- Golinkoff, Roberta M., Jill G. de Villiers, Kathy Hirsh-Pasek, Aquiles Iglesias and Mary Wilson. 2017. *QUILS: Quick Interactive Language Screener*. Baltimore, MD: Brookes Publishing.
- Green, Lisa. 2002. *African American English: A linguistic introduction*. Cambridge: Cambridge University Press.
- Gregory, K.-D. and Janna B. Oetting. 2018. Classification accuracy of teacher ratings when screening nonmainstream English-Speaking kindergartners for Language Impairment in the rural South. *Language, Speech, and Hearing Services in Schools* 49 (2), 218-231. Doi: 10.1044/2017_LSHSS-17-0045
- Hao, Y., Li Sheng, Yiwen Zhang, Fan Jiang, Jill G. de Villiers, Wendy Lee, Xueman Lucy Liu. 2018. A narrative evaluation of Mandarin-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research* 61, 345-359. doi:10.1044/2017.
- Horton, Ramonda M. and Ken Apel. 2014. Examining the use of spoken dialect indices with African American children in the southern United States. *American Journal of Speech-Language Pathology* 23, 448-460.
- Jackson, Janice E. and Barbara Z. Pearson. 2010. Variable use of features associated with African American English by typically developing children ages 4 to 12. *Topics in Language Disorders* 30 (2), 135-144.
- Kallenbach, Amy. 2007. Language for academic purposes: Performance of grade one English second language (ESL) learners on the Diagnostic Evaluation of Language Variation. Masters thesis, University of Witwatersrand, Johannesburg, South Africa.
- Kyuchukov, Hristo, Jill G. de Villiers, and Andrea Takahesu-Tabori. 2016. Assessing Roma children in their native language of Romani. In *Lincom studies in language acquisition and bilingualism. Volume 2: New trends in the psychology of language*, ed. by Hristo Kyuchukov.
- Labov, William. 1970. *The study of nonstandard English*. Urbana, IL: National Council of Teachers of English.
- Liu, X., Jill G. de Villiers, C-Y. Ning, E. Rolffhus, W. Lee, T. Hutchings, F. Hsiang, and Y. Ziang. 2017. Research to establish the validity, reliability and clinical utility of a comprehensive language assessment of Mandarin. *Journal of Speech, Language, and Hearing Research* 60 (3), 592-606.
- MacWhinney, Brian. 2016. The CHILDES Project. Parts 1 & 2. Available at www.childes.psy.cmu.edu.
- Marinis, Theodoros, Arhonto Terze, Angeliki Kotsopoulou, and Konstantinos Francis. 2013. Pragmatic abilities of high-functioning Greek-speaking children with autism. *Psychology* 20 (3), 321-337.
- National Council of Teachers of English (NCTE). 1974 (Fall). Students' right to their own language. Special Issue of *College Composition and Communication*, Vol. XXV.
- Nelson, Nikola W., and Yvette D. Hyter. 1990. How to use Black English Sentence Scoring (BESS) as a tool of non-biased assessment. Short course presented at the American Speech-Language-Hearing Association Annual Meeting, Dallas, TX. November.
- NIH Working Groups on AAE. 2004. Final Report: NIH Contract N01-DC8-2104: Development and validation of a language test for children speaking non-standard English: A study of children who speak African American English. 3/1/98 to 8/29/05. Ms., University of Massachusetts Amherst.

- Ning, C.Y., X.L. Liu, and Jill G. de Villiers. 2014. The Diagnostic Receptive and Expressive Assessment of Mandarin (DREAM). Dallas, TX: Bethel Hearing and Speaking Training Center.
- Norbury, Courtenay and Alison Sparks. 2012. Difference or disorder? Cultural issues in understanding neurodevelopmental disorders. *Developmental Psychology* 40 (1), 45-58.
- Norton, Pamela C. 2008. Speech-language therapists' perception of dialect and risk for disorder in African-American-English-speaking children. Doctoral dissertation, University of California Berkeley with San Francisco State.
- Oetting, Janna B., Brandi L. Newkirk, L. R. Hartfield, C. G. Wynn, Sonja L. Pruitt, and A. W. Garrity. 2010. Index of Productive Syntax for children who speak African American English. *Language, Speech, and Hearing Services in Schools* 41(3), 328-339.
- Pearson, Barbara Z. 2004. Learnability and triggers: Obligatory versus optional triggers for the passive in two dialects of English and in language impairment. In *BUCLD 28, The Proceedings of the 28th Annual Boston University Conference on Language Development*, ed. by Alejna Brugos, Linnea Micciulla, and Christine E. Smith, 447-460. Somerville, MA: Cascadilla Press.
- Pearson, Barbara Z., Tracy Conner, and Janice E. Jackson. 2013. Removing obstacles for African American English-speaking children through greater understanding of language difference. In special issue, *Deficit or difference: Interpreting diverse developmental paths*, ed. by Nameera Akhtar and Vikram Jaswal, *Developmental Psychology* 49 (1), 31-44.
- Pearson, Barbara Z., Janice E. Jackson, and Haotian Wu. 2014. Seeking a valid gold standard for an innovative, dialect-neutral language test. *Journal of Speech, Language, and Hearing Research* 57 (2):495-508.
- Petscher, Yaacov, Carol M. Connor, and S. Al'Otaiba. 2012. Psychometric analysis of the Diagnostic Evaluation of Language Variation assessment. *Assessment for Effective Intervention* 37 (4), 243-250. [doi:10.1177/1534508411413760](https://doi.org/10.1177/1534508411413760)
- Rivière, Andrew M., Janna B. Oetting, and Joseph Roy. 2018. Effects of Specific Language Impairment on a contrastive dialect structure: The case of infinitival TO across various nonmainstream dialects of English. *Journal of Speech, Language, and Hearing Research* 61, 1989-2001.
- Roeper, Thomas. 2007. *Prism of grammar: How child language illuminates humanism*. Cambridge, MA: MIT Press.
- Roeper, Thomas, Eliane Ramos, Harry N. Seymour, and Lamya Abdul-Karim. 2001. Language disorders as a window on a universal grammar: An abstract theory of agreement for IP, DP, and V-PP. *Brain and Language* 77 (3), 378-397.
- Roeper, Thomas and Harry N. Seymour. 1994. The place of linguistic theory in the theory of language acquisition and language impairment. In *Other children, other languages: Issues in the theory of language acquisition*, ed. by Yonata Levy, 305-330. Hillsdale, NJ: Lawrence Erlbaum.
- Sauerland, Ulrich. 2010. Cross-linguistically robust stages of children's linguistic performance. Final Evaluation Report, COST Action A33. Ms., Zentrum für allgemeine Sprachwissenschaft, Berlin.
- Scholarworks archive: <https://scholarworks.umass.edu/aae>.

The Legacy of the DELV

- Seymour, Charlena M. and Harry N. Seymour. 1988. Anatomy of an effective recruitment and retention program, In *Communication Sciences and Disorders: Marketing careers to minority students*, ed. by L. Cole. Rockville, MD: American Speech-Language-Hearing Association.
- Seymour, Harry N. 2003. Acknowledgements. In DELV-ST Technical Manual, Harry Seymour, Thomas Roeper, and Jill de Villiers, iii-iv. San Antonio, TX: The Psychological Corporation.
- Seymour, Harry N., Linda Bland-Stewart, and Lisa J. Green. 1998. Difference versus deficit in child African-American English. *Language, Speech, and Hearing Services in Schools*, 29, 96-108.
- Seymour, Harry N. and Barbara Z. Pearson, eds. February 2004. Evaluating language variation: Distinguishing development and dialect from disorder. Special issue of *Seminars in Speech and Language*, New York, NY: Thieme Medical Publishers.
- Seymour, Harry N., Thomas Roeper, and Jill G. de Villiers. 2000. Dialect Sensitive Language Test (DSLIT). San Antonio, TX: The Psychological Corporation. (Unpublished).
- Seymour, Harry N., Thomas Roeper, and Jill G. de Villiers. 2003a. Diagnostic Evaluation of Language Variation: Screening Test (DELV-ST). San Antonio, TX: The Psychological Corporation. [Republished 2018, Sun Prairie, WI: Ventris Learning.]
- Seymour, Harry N., Thomas Roeper, and Jill G. de Villiers. 2003b. DELV-CR (Diagnostic Evaluation of Language Variation) Criterion-Referenced Test. San Antonio, TX: The Psychological Corporation.
- Seymour, Harry N., Thomas Roeper, and Jill G. de Villiers. 2005. DELV-NR (Diagnostic Evaluation of Language Variation) Norm-Referenced Test. San Antonio, TX: The Psychological Corporation. [Republished 2018, Sun Prairie, WI: Ventris Learning.]
- Seymour, Harry N., and Charlena M. Seymour. 1977. A therapeutic model for communicative disorders among children who speak Black English Vernacular. *Journal of Speech and Hearing Disorders* 42, 247-256.
- Southwood, Frenette. 2013. Towards a dialect-neutral assessment instrument for the language skills of Afrikaans-speaking children: The role of socioeconomic status. *Journal of Child Language* 40, 415-437.
- Southwood, Frenette, and Ondene van Dulm. 2012. Practice Books for DELV Test Components. Johannesburg, South Africa: Jopie van Rooyen and Partners.
- Stockman, Ida J. 2010. A review of developmental and applied language research on African American children: From a deficit to difference perspective on dialect differences. *Language, Speech, and Hearing Services in Schools* 41 (1), 23-38. Doi: 10.1044/0161-1461(2009/08-0086)
- Stockman, Ida, Joanna Boulton, and Gregory C. Robinson. 2008. Multicultural/ multilingual educational programs: A survey of perceived faculty practices and outcomes. *American Journal of Speech Language Pathology* 17 (3), 241-64. doi: 10.1044/1058-0360(2008/023).
- Strauss, Uri, Thomas Roeper, Barbara Z. Pearson, Jill G. de Villiers, and Harry N. Seymour. 2003. The acquisition of exhaustivity in *wh*-questions. Paper presented at the General Linguistics in the Old World (GLOW) Annual Meeting, Lund, Sweden, April.

Barbara Zurer Pearson

- Terry, Nicole P., Yaacov Petscher, and Katherine T. Rhodes. 2017. Psychometric analysis of the Diagnostic Evaluation of Variation–Screening Test: Extension to low-income African American pre-kindergarteners. *Assessment for Effective Intervention* 42 (3), 176–185. DOI: 10.1177/1534508416679402
- Tomblin, J. Bruce, N. Records, P. Buckwalter, X. Zhang, E. Smith, and M. O’Brien. 1997. Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research* 40, 1245-1260.
- van Dulm, Ondene, and Frenette Southwood. 2008. Toward a dialect-neutral Afrikaans-medium child language assessment instrument: Test item development. *Language Matters* 39 (2), 300-315.
- Williams, Robert L. 1975. *Ebonics: The true language of black folks*. St. Louis, MO: Institute of Black Studies.
- Wolfram, Walter. 1969. *A sociolinguistic description of Detroit Negro speech*. Washington, DC: Center for Applied Linguistics.