Long-Term Consistency in Speech/Language Profiles: I. Developmental and Academic Outcomes

JOSEPH H. BEITCHMAN, M.D., BETH WILSON, B.Sc., E.B. BROWNIE, B.Sc., HEATHER WALTERS, B.Sc., AND WILLIAM LANCEE, Ph.D.

ABSTRACT

Objective: This study examined the 7-year developmental and academic outcome of speech/language-impaired and control children selected from a community sample. Method: Speech/language and psychiatric measures were administered to the children at ages 5 and 12.5 years. Using children's age 5 speech/language test results, a cluster analysis was performed to ascertain whether specific linguistic subgroups would emerge. The long-term consistency of these subgroups was explored. The association between time 1 speech/language clusters and linguistic, cognitive, and academic measures at time 2 were examined. Results: Four groups were identified in the cluster analysis: high overall, poor articulation, poor comprehension, and low overall. Children with pervasive language problems continued to perform poorly on linguistic, cognitive, and academic measures, while those with comprehension problems fared slightly better but still had more difficulties than those with normal language. The poor articulation cluster had few articulation errors at follow-up. Conclusions: Empirically supported speech/language classifications identified as early as age 5 continued to be relevant into late childhood. Pervasive speech/language impairment in early childhood was associated with increased risk of poor linguistic and academic outcome at follow-up, while isolated articulatory problems improved over time. These findings reveal the urgent need for early intervention among children with pervasive speech/language impairment. J. Am. Acad. Child Adolesc. Psychiatry, 1996, 35(6):804-814. Key Words: speech/language impairment, empirical classification, longitudinal study, academic outcomes, speech/language outcomes.

Given the heterogeneous nature of speech/language disorders (Beitchman et al., 1989b), establishing an empirically based classification system is an important initial step toward a conceptualization of such disorders. An empirical classification system of linguistic profiles will provide a means for testing the associations between clinically relevant variables and various speech/language competencies.

While many researchers have demonstrated the association between demographic, cognitive, behavioral, and academic measures and psychiatric status, few attempts have been made to examine the relationship between these variables and specific speech/language competencies. Beitchman et al. (1989b), Fundudis et al. (1979), and Wolfus et al. (1980) have examined the association between empirically derived speech/language profiles and concurrent cognitive, developmental, and academic measures. However, with the exception of Beitchman et al. (1989b), no investigators have explored the long-term outcome of early childhood speech/language competencies in terms of late childhood speech/language, cognitive, and academic measures.

In the initial study, Beitchman et al. (1989b) identified four speech/language clusters in a community sample of speech/language-impaired 5-year-old children and normal controls. On the basis of the children's scores on measures of articulation and expressive and receptive language and tests of auditory comprehension and auditory memory, four clusters were identified: high overall, poor comprehension, poor articulation, and low overall. Children in the high overall cluster...
scored significantly higher than the other clusters on all speech/language measures. Children in the low overall cluster scored significantly lower than the other clusters on all speech/language measures except the test of articulation. Children in the poor articulation cluster scored significantly lower than all other groups on the articulation measure but had higher scores than the poor comprehension and low overall clusters on the other measures. Children in the poor comprehension cluster scored in the low range on most of the tests, especially the auditory comprehension measure.

Significant differences were found among the clusters in IQ and visual-motor integration scores. Children in the high overall cluster scored highest on the IQ measure, followed by the poor articulation, poor comprehension, and low overall clusters. All groups scored significantly higher than the low overall cluster on the Beery test of visual-motor integration.

As part of a 7-year follow-up study, this report compares the age 12.5 speech/language, cognitive, and academic functioning of children in the four speech/language clusters identified at age 5 years. The results of these comparisons will address a number of theoretical issues pertaining to the nature of speech/language delays. By determining whether these groups still distinguish children on speech/language and other variables, the stability of the original profiles can be assessed. The analysis of changes in speech/language functioning across time allows us to understand better the development of speech/language functioning. Do specific speech/language skills develop at different rates? Do delayed groups “catch up” in speech/language functioning? To date these remain unanswered questions.

In practical terms, this study will provide information about the outcome of specific speech/language problems and related factors. This information can be used to better inform parents and children not only of the outcomes of speech/language disorder but also how other factors such as academic performance are likely to be affected. By providing parents and other caregivers with a full account of the likely effects of children’s speech/language problems, practitioners can improve the chances that children’s deficits in related areas will be understood and not be misattributed to other factors such as lack of motivation. Understanding the long-term outcome of specific speech/language deficits will also provide practitioners with empirically based rationales for a more effective use of speech/language intervention. If children with particular speech/language profiles recover by late childhood from their early childhood deficits, the need for speech/language intervention may be less urgent than for those children whose deficits remain in late childhood.

This is the first of two articles reporting on the 7-year outcome of children’s empirically classified speech/language functioning. The present article describes the association between children’s speech/language classification at age 5 years and late childhood outcome at age 12.5 years. Associated cognitive variables such as IQ and visual-motor skills are examined, as well as demographic factors such as socioeconomic status (SES). The second article (Beitchman et al., 1996) explores the association between the speech/language groups based on children’s age 5 linguistic functioning and their behavioral, emotional, and social outcomes at age 12.5 according to teacher, mother, and self-reports.

**METHOD**

**Subjects**

A detailed account of subject selection procedures is outlined elsewhere (Beitchman et al., 1994). Briefly, in 1982 the first stage of a three-stage language-screening procedure was administered to a one-in-three random sample (N = 1,655) of all 5-year-old English-speaking kindergarten children in the Ottawa-Carleton school region. Children who passed the screening participated in further speech/language testing, and children who failed the second stage of speech/language testing participated in the third stage of testing, which included cognitive, development, and psychiatric assessments. A total of 169 children, including a subsample of children who passed the speech/language tests, participated in the three stages of testing and were included in the cluster analysis. For a full description of the 1982 study methodology, see Beitchman et al. (1989a,b). Subjects were recontacted in 1989 and invited to participate in follow-up language, cognitive, and psychiatric assessments. Of the 169 time 1 participants (110 boys, 59 girls), 124 (85 boys, 39 girls) participated in speech/language assessments at follow-up.

**Attrition**

To examine the continuity of speech/language performance over time, it was necessary to ensure that the sample discussed in this article did not represent a radically different group from that discussed in Beitchman et al. (1989b). A total of 45 children (26.7%) who participated in the time 1 study did not participate in the follow-up speech/language assessment. The attrition group was compared with the 124 continuing participants on the time 1 language, cognitive, and developmental measures reported in Beitchman et al. (1989b). No differences were found between the groups on any of these measures. No differences in time 1 SES were found between the attrition and participating groups; however, children who participated at time 2 were statistically more likely to have had two-parent homes at time 1.
Measures

Table 1 shows the speech/language, cognitive, developmental, and academic measures used at time 1 and time 2.

Speech/Language Measures. Speech functioning was measured at time 1 and time 2 using the Photo Articulation Test (PAT) (Pondergast et al., 1969) and a nonstandardized clinical assessment of voice and fluency problems. Since no articulation errors are expected in normally developing speech by age 12.5 years, and norms for the PAT were unavailable for children older than age 11 years, raw error totals were used to measure articulation at time 2. For repeated-measures analyses, raw error totals were also used for the time 1 PAT.

Language functioning was measured at time 1 and time 2 using several standardized instruments. The Peabody Picture Vocabulary Test-Revised (PPVT-R) (Dunn and Dunn, 1981) and the Content (C) and Sequence (S) subscales of the Goldman-Fristoe-Woodcock Auditory Memory Test (GFW) (Goldman et al., 1974) were used to measure receptive comprehension and memory at time 1 and time 2. The two language-screening instruments—Screening Test for Auditory Comprehension of Language (STACL) (Carrow, 1973) and Bankson Language Screening Test (BLST) (Bankson, 1977)—were used to measure auditory comprehension and expressive language at time 1 only. At follow-up, auditory comprehension was measured by the Clinical Evaluation of Language Fundamentals-Revised (Semel et al., 1986). Expressive and receptive language were measured by the Test of Language Development (TOLD) at both time 1 and time 2. At time 1, the TOLD (Primary) (Newcomer and Hammill, 1977) was used and at time 2 the TOLD-2 (Intermediate) (Newcomer and Hammill, 1988), which is designed for older children, was used. The TOLD Listening Quotient subtest measured receptive language, while the TOLD Speaking Quotient subtest measured expressive language. The TOLD Spoken Language Quotient subtest measured both receptive and expressive language. In addition, a 10-minute conversational sample was scored using a nonstandardized checklist (P-K) to assess pragmatic skills (Prutting and Kirchner, 1987).

Statistical Classification of Speech/Language Categories. Cluster analysis was used at time 1 to determine whether homogeneous groups of children with similar linguistic profiles could be identified. Speech/language measures used for these analyses were STACL and BLST percentiles, PAT percent scores, PPVT standard scores, and GFWC and GFWS scores. The McQueen's k-means clustering method was used for this analysis, and a four-cluster solution was indicated.

Cognitive, Developmental, and Academic Measures. Cognitive functioning was measured at time 1 and time 2 using the age-appropriate Wechsler intelligence scale: WPPSI (Wechsler, 1967) at time 1 and WISC-R (Wechsler, 1981) at time 2. The Developmental Test of Visual-Motor Integration (VMI) (Beery and Buktenica, 1967) was administered at time 1 and the revised version (Beery, 1982) was administered at time 2. Academic functioning was measured at time 2 using the Kaufman Test of Educational Achievement (KTEA) (Kaufman and Kaufman, 1985). Items from a teacher report form measured use of special education services. To be considered a consumer of special education services, a child's teacher had to report that the child had been referred for special class placement, services, or tutoring. To verify that placements reflected remedial aid services rather than enhancement programs for gifted children, teachers were asked to describe the kind of services rendered and their responses were coded accordingly. Unfortunately, only a determination of the general category of special education services was obtained rather than a more specific measure.

Demographics. Demographic measures included marital and common law status from the parent's interview and Blishen ratings of SES. The Blishen Scale is a 7-point ordinal coding scale of employment status with 1 as unemployed or never employed and 7 as the highest status occupation. The scale was developed using census data based on persons in the male labor force. Employment status is based on prestige scores, income, and educational requirements (Blishen and McRoberts, 1976). SES was assessed for each parent. The occupation of the parent with the highest SES was used to calculate the Blishen code. Because of the variation of students' economic positions, parents with the highest SES in the family who also identified as students were omitted from this analysis (n = 3). Blishen ratings below 3 were classified as low, Blishen ratings at or above 5 were classified as high, and the remaining ratings were classified as medium SES. Issues of gender with reference to the linguistic clusters will be explored in a future publication. Since there were very few nonwhite participants in this study, race was not recorded.

Analyses

Analyses of variance (ANOVAs) and t-tests were conducted to ascertain whether time 2 demographic, linguistic, cognitive, developmental, and educational outcomes were associated with the children's time 1 linguistic profiles. Where possible, repeated-measures analyses were conducted with demographic, linguistic, cognitive, and developmental measures to determine the extent to which the children's scores remained stable over time within their respective clusters and the extent to which changes between clusters

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Tests and Measures at Time 1 and Time 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1 Speech/Language Tests</td>
<td>• Photo Articulation Test</td>
</tr>
<tr>
<td>• Peabody Picture Vocabulary Test-Revised</td>
<td></td>
</tr>
<tr>
<td>• Goldman-Fristoe-Woodcock Auditory Memory Test-Content</td>
<td></td>
</tr>
<tr>
<td>• Goldman-Fristoe-Woodcock Auditory Memory Test-Sequence</td>
<td></td>
</tr>
<tr>
<td>• Screening Test for Auditory Comprehension of Language</td>
<td></td>
</tr>
<tr>
<td>• Bankson Language Screening Test</td>
<td></td>
</tr>
<tr>
<td>• Test of Language Development-Primary</td>
<td></td>
</tr>
<tr>
<td>Time 2 Speech/Language Tests</td>
<td>• Photo Articulation Test</td>
</tr>
<tr>
<td>• Peabody Picture Vocabulary Test-Revised</td>
<td></td>
</tr>
<tr>
<td>• Goldman-Fristoe-Woodcock Auditory Memory Test-Content</td>
<td></td>
</tr>
<tr>
<td>• Goldman-Fristoe-Woodcock Auditory Memory Test-Sequence</td>
<td></td>
</tr>
<tr>
<td>• Test of Language Development-2-Intermediate</td>
<td></td>
</tr>
<tr>
<td>• Clinical Evaluation of Language Fundamentals-Revised</td>
<td></td>
</tr>
<tr>
<td>• Prutting &amp; Kirchner Pragmatics Scale</td>
<td></td>
</tr>
<tr>
<td>Time 1 Cognitive and Developmental Measures</td>
<td>• Wechsler Preschool and Primary Scale of Intelligence</td>
</tr>
<tr>
<td>• Beery Developmental Test of Visual-Motor Integration</td>
<td></td>
</tr>
<tr>
<td>• Audiology screening</td>
<td></td>
</tr>
<tr>
<td>Time 2 Cognitive, Developmental, and Academic Measures</td>
<td>• Wechsler Intelligence Scale for Children-Revised</td>
</tr>
<tr>
<td>• Beery Developmental Test of Visual-Motor Integration</td>
<td></td>
</tr>
<tr>
<td>• Kaufman Test of Educational Achievement</td>
<td></td>
</tr>
<tr>
<td>• Audiology screening</td>
<td></td>
</tr>
</tbody>
</table>

* The test was used in time 1 cluster analysis.
occurred. Tukey Honestly Significant Difference post hoc comparisons with t values set at .01 and .05 were used to test for significant differences among specific clusters when significant cluster effects were found. It was not possible to conduct repeated-measures analyses using educational tests since the WCTA was administered at time 2 only, and no comparable educational measure was taken when the children were 3 years old.

RESULTS

Four linguistic groups emerged from the cluster analysis: high overall, poor comprehension, poor articulation, and low overall. Children in the high overall cluster were characterized by higher performance on all speech/language tests used in the cluster analysis. Members of the poor comprehension cluster obtained scores that were within the normal range on all of these tests except for the STACL, a measure of auditory comprehension. Children in the poor articulation cluster were impaired on the PAT, a measure of articulation, but they scored within the normal range on all other tests used in the cluster analysis. Finally, children in the low overall cluster obtained low scores on all of these speech/language measures.

The time 1 sample (n = 169) included 16 children from the high overall cluster, 30 from the poor comprehension cluster, 56 from the poor articulation cluster, and 67 from the low overall cluster. The time 2 sample used in the analyses presented in this article (n = 124) were similarly distributed, with 13 children from the high overall cluster, 24 from the poor comprehension cluster, 46 from the poor articulation cluster, and 41 from the low overall cluster. Children who participated at time 2 did not necessarily complete every test in the battery or may have had incomplete demographic information, causing the sample size per measure to vary slightly.

Demographics

No significant differences between linguistic clusters were found when examining parental marital status or marital/common law status at time 2. This result differed from our time 1 finding that linked marital status with children’s linguistic cluster membership (Beitchman et al., 1989b). While post hoc tests were not significant, there appeared to be a higher percentage of unmarried parents among children from the low overall cluster than among those from the other three clusters.

There was a significant relationship between time 2 SES and time 1 cluster membership ($\chi^2 [6] = 12.76$, $p < .05$). Only 7.7% of children from the high overall cluster had families with low SES at time 2, with the remaining high overall children equally divided between medium and high SES families. Similarly, only 13.4% of children from the poor articulation cluster had families with low SES at time 2, with the greatest proportion coming from high SES families. Among children in the poor comprehension cluster, family SES was more evenly distributed at time 2, with more than one fourth from low SES families. Among low overall cluster children, 42.2% had families with low SES at time 2, with the remainder divided between medium and high SES families.

To determine the extent to which family SES changed over time by cluster membership, a repeated-measures analysis was conducted using Blishen codes at time 1 and time 2. No linguistic cluster by time interaction effect was found.

Speech/Language Measures

To ascertain whether children’s time 1 speech/language clusters differed significantly on their time 2 language test scores, one-way ANOVAs were conducted. Results are shown in Table 2.

With the exception of the PAT, all time 2 linguistic test results significantly differentiated among the time 1 speech/language clusters. Post hoc tests revealed that the low overall cluster continued to perform poorly relative to their peers on tests of language proficiency. Children from the poor comprehension cluster also showed continued language deficits compared with children without early receptive language impairment. These children experienced difficulties in expressive and receptive language and pragmatics. In contrast, the poor articulation cluster performed similarly to the high overall cluster on many time 2 linguistic measures.

To determine the extent to which children’s scores changed over time by speech/language cluster, repeated-measures analyses were conducted with the PPVT, PAT, TOLD, and the GFW speech/language tests at time 1 and time 2. Repeated-measures analyses using mean scores from the PAT and the GFW speech/language tests revealed significant linguistic cluster by time interaction effects. These results are graphed in Figure 1 and Figure 2. On the PAT, children in the high overall and poor comprehension clusters had slight decreases in the number of errors made from time 1 to time 2, but there were very large decreases for the poor articulation cluster (21.65 to 3.61) and for the
low overall cluster (14.65 to 2.51). Score decreases on the GFWC were most pronounced for the high overall, poor articulation, and poor comprehension clusters from time 1 to time 2, whereas scores for the low overall cluster stayed at the same very low level.

Cognitive and Developmental Measures

At time 2, the WISC-R and the VMI were used as cognitive and developmental measures. To ascertain whether children’s cognitive and developmental scores at time 2 differed by linguistic cluster, one-way ANOVAs were conducted. The results of these analyses are shown in Table 3. Post hoc tests revealed that the low overall cluster had significantly lower mean scores than the other three clusters on these cognitive and developmental measures. However, no differences were found between mean IQ scores of the poor articulation and high overall clusters.

Repeated-measures analyses were conducted on these cognitive and developmental measures. No linguistic cluster by time interaction effects were found. Cognitive and developmental disadvantages experienced at time 1 by the low overall cluster appeared to continue at time 2. Furthermore, children in the poor comprehension cluster, while obtaining higher scores than the low overall cluster, continued to perform at a significantly lower cognitive level than children in the high overall cluster. The poor articulation cluster performed similarly to the high overall cluster on the WISC-R and the VMI.

At time 1, the results on the audiometry test significantly discriminated among the four clusters. Of the children in the low overall cluster, 25% failed the audiometry test, while fewer than 10% of children in the poor comprehension and poor articulation groups failed, and no children in the high overall group failed. The time 2 audiometry test results, however, did not discriminate among the four time 1 clusters (χ²[3] = 2.72, not significant).

Educational Measures

One-way ANOVAs were conducted to determine whether children’s time 1 linguistic clusters were associated with time 2 performance on the KTEA, an aca-

### TABLE 2
Mean Time 2 Speech/Language Test Scores by Time 1 Speech/Language Cluster

<table>
<thead>
<tr>
<th>Time 2 Speech/Language Test</th>
<th>C1 High Overall (n = 46)</th>
<th>C2 Poor Comprehension (n = 41)</th>
<th>C3 Poor Articulation (n = 24)</th>
<th>C4 Low Overall (n = 13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAT raw score total</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>WISC-R Standard score</td>
<td>116.85</td>
<td>19.05</td>
<td>97.58</td>
<td>13.67</td>
</tr>
<tr>
<td>Goldman-Fristoe-Woodcock Auditory Memory Test, Subscale</td>
<td>52.23</td>
<td>7.75</td>
<td>48.21</td>
<td>7.27</td>
</tr>
<tr>
<td>Clinical Evaluation of Language Fundamentals Revisited</td>
<td>108.62</td>
<td>11.65</td>
<td>43.71</td>
<td>9.15</td>
</tr>
<tr>
<td>GFWC standard score</td>
<td>50.85</td>
<td>11.65</td>
<td>43.71</td>
<td>9.15</td>
</tr>
<tr>
<td>CELF-R receptive language</td>
<td>102.00</td>
<td>8.40</td>
<td>85.33</td>
<td>14.89</td>
</tr>
<tr>
<td>PNC, K &amp; Kirchner’ (n = 121)</td>
<td>1.85</td>
<td>2.41</td>
<td>5.09</td>
<td>4.93</td>
</tr>
</tbody>
</table>

Note: PAT = Phonotactic Analysis Test; PPVT-R = Peabody Picture Vocabulary Test-Revised; GFWC = Goldman-Fristoe-Woodcock Auditory Memory Test, C1, C2, C3, C4 = Significantly different at p = .05; C1, C2, C3, C4 = Significantly different at p = .001.
Results are presented in Table 4. Time 2 mean KTEA subscale scores significantly differentiated among time 1 linguistic clusters. On each KTEA subscale and the overall composite, the low overall cluster had significantly lower mean scores than the other three clusters according to post hoc tests. The poor comprehension cluster had significantly lower mean scores than the high overall cluster on each of the KTEA subscales and the overall composite. While the high overall clusters appeared to have slightly higher mean scores on all educational measures than the poor articulation cluster, these differences were not significant on post hoc tests. Repeated-measures analyses could not be conducted because the KTEA was administered only at time 2.

Children's time 2 use of special education services differed significantly by their time 1 linguistic cluster membership ($\chi^2[3] = 11.84, p < .01$). Children in the high overall cluster differed significantly from children in the other three clusters on special education placement. None of the children from the high overall cluster received special education. In contrast, 64% of the low overall cluster, 50% of the poor comprehension cluster, and 46.3% of the poor articulation cluster received special education.

**DISCUSSION**

These results reveal that childhood speech/language competence assessed as early as age 5 is associated with demographic characteristics, linguistic skills, IQ, visual-motor skills, and academic achievement measures 7 years later. In general, the high overall cluster had the best scores followed by the poor articulation, poor comprehension, and low overall clusters, respectively. This pattern of results was found across language, cognitive, developmental, and academic measures. Language differences among clusters were found on time 2 measures of receptive vocabulary, auditory memory, receptive language, expressive language, and pragmatics.
TABLE 3
Mean Time 2 Cognitive and Developmental Scores by Time 1 Speech/Language Cluster

<table>
<thead>
<tr>
<th>Cognitive and Developmental Measures</th>
<th>C1 High Overall (n = 13)</th>
<th>C2 Poor Comprehension (n = 22)</th>
<th>C3 Poor Articulation (n = 46)</th>
<th>C4 Low Overall (n = 40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WISC-R (n = 121)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full Scale IQ*</td>
<td>115.46 ± 12.89</td>
<td>102.23 ± 7.72</td>
<td>106.65 ± 11.18</td>
<td>86.40 ± 13.37</td>
</tr>
<tr>
<td>Performance IQ*</td>
<td>118.54 ± 10.82</td>
<td>107.50 ± 9.40</td>
<td>108.83 ± 11.72</td>
<td>90.58 ± 14.60</td>
</tr>
<tr>
<td>Verbal IQ*</td>
<td>109.77 ± 13.62</td>
<td>97.77 ± 9.80</td>
<td>103.98 ± 11.14</td>
<td>84.70 ± 13.07</td>
</tr>
<tr>
<td>VMI standard score† (n = 116)</td>
<td>109.85 ± 8.67</td>
<td>104.00 ± 13.24</td>
<td>102.25 ± 9.86</td>
<td>90.16 ± 11.61</td>
</tr>
</tbody>
</table>

Note: WISC-R = Wechsler Intelligence Scale for Children-Revised; VMI = Developmental Test of Visual-Motor Integration.

*F(3,117) = 31.15, p < .0001, C4 < C3, C2, C1, p < .01; C2 < C1, p < .01.
†F(3,117) = 24.65, p < .0001, C4 < C3, C2, C1, p < .01; C2 < C1, p < .01.
‡F(3,112) = 14.79, p < .0001, C4 < C3, C2, C1, p < .01.
§F(3,118) = 23.91, P < .0001, C4 < C2, p < .01; C4 < C1, C3, p < .01; C2 < C1, p < .01.
¶F(3,117) = 24.78, P < .0001, C4 < C3, C2, C1, p < .01; C2 < C1, p < .01.
||F(3,116) = 20.18, P < .0001, C4 < C2, p < .05; C4 < C1, C3, p < .01; C2 < C1, p < .01.
©F(3,116) = 24.55, P < .0001, C4 < C3, C2, p < .05; C4 < C1, C3, p < .01; C2 < C1, p < .01.

In an earlier report, Beitchman et al. (1989b) suggested that neurodevelopmental delay may be the underlying cause of linguistic deficiencies. If this were the case we would expect to see delays or deficiencies in other areas of functioning. In fact, the results demonstrate continuing deficits in IQ, visual-motor skills, and academic achievement among these children. These results are consistent with a neurodevelopmental immaturity model of language impairment.

Powell and Bishop (1992) argued that the motor and perceptual deficits that co-occur with specific language impairment arise from a common etiology, as yet unspecified. Locke (1994) believed that neurodevelopmental delay is responsible for many of the children’s auditory, visual, and tactile perceptual deficits, a view

TABLE 4
Mean Time 2 KTEA Subscale Scores by Time 1 Speech/Language Cluster (n = 120)

<table>
<thead>
<tr>
<th>Time 2 KTEA Subscale</th>
<th>C1 High Overall (n = 13)</th>
<th>C2 Poor Comprehension (n = 21)</th>
<th>C3 Poor Articulation (n = 46)</th>
<th>C4 Low Overall (n = 40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Battery Com SS by age*</td>
<td>116.23 ± 11.68</td>
<td>97.38 ± 12.82</td>
<td>106.72 ± 16.37</td>
<td>84.03 ± 14.54</td>
</tr>
<tr>
<td>Spelling SS by age†</td>
<td>114.77 ± 14.48</td>
<td>99.95 ± 11.77</td>
<td>103.50 ± 15.43</td>
<td>87.35 ± 14.70</td>
</tr>
<tr>
<td>Reading Com SS by age‡</td>
<td>116.54 ± 12.73</td>
<td>98.71 ± 13.66</td>
<td>106.76 ± 15.92</td>
<td>86.43 ± 14.25</td>
</tr>
<tr>
<td>Math Com SS by age§</td>
<td>113.77 ± 9.56</td>
<td>97.05 ± 12.92</td>
<td>107.89 ± 15.27</td>
<td>84.75 ± 14.48</td>
</tr>
</tbody>
</table>

Note: KTEA = Kaufman Test of Academic Achievement; Com = Composite; SS = standard score.
*F(3,116) = 23.91, P < .0001, C4 < C2, p < .05; C4 < C1, C3, p < .01; C2 < C1, p < .01.
†F(3,116) = 15.20, P < .0001, C3 < C1, p < .05; C4 < C2, p < .05; C4 < C1, C3, p < .01; C2 < C1, p < .01.
‡F(3,116) = 20.18, P < .0001, C4 < C2, p < .05; C4 < C1, C3, p < .01; C2 < C1, p < .01.
§F(3,116) = 24.55, P < .0001, C3 < C2, p < .05; C4 < C1, C3, p < .01; C2 < C1, p < .01.
similar to our neurodevelopmental immaturity hypothesis and consistent with the follow-up scores of children from the low overall cluster. The idea that language-delayed children’s brains develop more slowly than others and that language-delayed children perform poorly on tasks that, like language, require cognitive or motor sophistication is also similar to Bishop and Edmundson’s (1987) notion of maturational lag. In their 2-year follow-up study, Bishop and Edmundson reported that while language-delayed and normal-language youngsters developed linguistically at similar rates, language-delayed children maintained their original level of delay.

Repeated-measures analyses on available linguistic, cognitive, and developmental scores for the four clusters demonstrated similar rates of development across clusters over the 7-year follow-up period. Only two significant time by cluster interactions, one involving articulation and the other involving memory for content, were found. No other language or cognitive measure showed a significant time by cluster interaction, indicating that developmental rates were similar across clusters.

Why did performance on articulation and memory for content measures progress at different rates across the four clusters? Children from the poor articulation cluster had the greatest number of articulation errors at time 1, but their rapid rate of improvement suggests that their articulation skills may have been following a different developmental time course than were the articulation skills of children from the other clusters, with more rapid progress between the ages of 5 and 12.5 rather than prior to the age of 5. This finding of improved articulation at time 2 is not surprising given results from Baker and Cantwell’s (1987) follow-up study of speech/language-impaired youngsters. These researchers reported that the percentage of pure speech problems found in their sample had the most dramatic decline at follow-up, in comparison to pure language and mixed speech and language problems, which appeared resistant to change.

In the case of memory for content, it appears that the low overall cluster had continued deficits over time. The other three clusters showed decreases in their mean scores on auditory memory for content over time. The stability in the performance of the low overall cluster may actually reflect an improvement for this group. Mean scores of the other three clusters decreased by approximately 10 standard score points from time 1 to time 2. These decreases may reflect the increased skill level required by the age 12.5 measure of auditory memory compared with the age 5 measure of auditory memory. If so, we would expect all groups to show a decrease over time. Instead, the mean auditory memory for content scores for the low overall cluster were relatively similar at both time points. However, any improvements these children did experience were not sufficient to allow them to “catch up” with their peers. As demonstrated by the overall results, the low overall cluster continued to have marked deficits in several areas of language competence including auditory memory for content.

These results illustrate how early detection of specific language problems can help to identify children most at risk and in need of interventions. The poor articulation cluster did not appear to be at risk in terms of linguistic and educational performance at follow-up. It is unclear whether their articulation improved spontaneously or through successful speech interventions over the 7-year period. Other study data revealed that half of the children from the poor articulation cluster received some form of speech/language intervention. However, the impact of intervention cannot be adequately assessed owing to the nature of this study. That is, because this longitudinal study did not seek to evaluate the efficacy of speech/language intervention, children were not randomly assigned to treatment and no-treatment control groups.

In contrast to the performance of the poor articulation cluster, the low overall cluster showed the least progress over time and was most at risk educationally. Early intervention for these youngsters must be a priority. Furthermore, it is apparent that their deficits extend beyond simply that of language but include visual-motor skills and educational achievement.

Similar to the low overall cluster, the poor comprehension cluster had significantly lower mean scores on several language, cognitive, developmental, and academic measures compared with the high overall cluster. However, the poor comprehension cluster’s mean scores were within the normal range on most of these measures. Still, differences between the poor comprehension and high overall clusters suggest that children from the poor comprehension cluster may appear somewhat impaired, particularly in academic settings, where linguistic skills are most likely to be taxed. It may be
that specific interventions would improve the skill level and abilities of these children as well. Despite mean scores within the normal range at follow-up, analyses revealed that half of the poor comprehension cluster had participated in some special education service, suggesting some degree of impairment.

It is also of interest that there was a significant association between speech/language cluster and SES. The vast majority of children from the high overall and poor articulation clusters were from middle to high SES families at time 2, whereas more than one quarter of the children from the poor comprehension cluster and almost half of the children from the low overall cluster were from low SES families at time 2. While SES for children in each of the clusters improved over time, there was no significant difference in the amount of change in SES among the clusters.

It appears that children with more pervasive language problems were also more likely to live in economically disadvantaged circumstances. The importance of SES has also been revealed in literature linking child poverty to poor nutrition, inadequate housing, and, in general, restricted family access to necessary resources (Centre for International Statistics on Economic and Social Welfare, 1993). Researchers have also found that children exposed to environments rich in language stimulation acquire language competence earlier than do children deprived of such environments (Puckering and Rutter, 1987). It is clear that economic conditions play a role in children’s speech/language development. The degree of influence social class exerts on children’s speech/language development requires further investigation.

Speech/Language Clusters and the DSM-IV Classification System for Communication Disorders

In addition to the issues discussed in the previous section, these empirically derived clusters provide a basis for critically examining the validity of the DSM-IV classification system for communication disorders (American Psychiatric Association, 1994). The DSM-IV has classified communication disorders under the following categories: mixed receptive-expressive language disorder, phonological disorder, stuttering disorder, expressive language disorder, and language disorder not otherwise specified. A brief examination of the similarities and differences between our empirically based subgroups and the DSM-IV classifications follows.

Characteristics distinct to the low overall cluster appeared most similar to the description of mixed receptive-expressive language disorder from the DSM-IV. Similar to the DSM-IV criterion for the mixed receptive-expressive language disorder, children from the low overall cluster had scores on receptive and expressive language tests that were substantially below those obtained from standardized measures of their nonverbal intellectual capacity. As well, we found evidence of disturbances in their academic achievement and social communication, another DSM-IV criterion.

In contrast, characteristics of the poor articulation cluster appeared most similar to the DSM-IV phonological disorder and stuttering disorder categories. Similar to the DSM-IV criteria for both stuttering and phonological disorders, these children showed age-inappropriate speech sounds and fluency problems at time 1. However, in contrast to the criteria for both of these disorders, children from the poor articulation cluster showed high academic performance and strong social communication skills. It may be that the poor articulation cluster represents children with the least severe of the phonological or stuttering disorder categories. Their improved articulation scores at follow-up provide support for this notion.

The poor comprehension cluster was characterized by impaired receptive language functioning but had normal scores on the BLST, a measure of expressive language functioning, at time 1. The DSM-IV reports that children with purely receptive language disorders are “virtually never seen.” It is unclear whether children in the poor comprehension cluster had undetected expressive language deficits or whether due to the nature of community samples a group of children were revealed that would be very unlikely to be detected and referred to clinicians for treatment. Evidence presented in this article’s companion paper regarding behavioral, emotional, and social outcomes demonstrates an association between early childhood poor comprehension and behavioral problems (Beitchman et al., 1996). These behavioral problems may overshadow any language deficits children with poor comprehension experience.

It is interesting that none of the speech/language clusters had a linguistic profile which resembled the DSM-IV expressive language disorder classification. The
DSM-IV reports prevalence rates for developmental expressive language disorder between 3% and 5%. Yet no cluster clearly resembled this category despite the utilization of a community sample. In another report (Beitchman et al., 1994), we have described a group of children from this study with expressive language deficits only. The TOLD Spoken Language and Listening Quotients were used to define this group. In terms of distinguishing between expressive language deficits and normal functioning, it is possible that the TOLD Spoken Language Quotient is a more sensitive instrument than the BLST. Because the BLST is intended as a screening instrument, using more sensitive measures of expressive language functioning in similar cluster analyses may result in the identification of a group with purely expressive language disorders.

Bishop and Rosenbloom (1987), using their clinical experience and findings from the relevant literature, proposed a classification system for speech and language disorders. They have described a category of disorder labeled syntactic-pragmatic disorder, which loosely corresponds to delayed or disordered phonological and expressive language. However, Bishop and Rosenbloom reported that studies of these children revealed deficits of receptive language as well. The identification of a category of children with expressive language impairment as described in the DSM-IV is clearly important, but the empirical data remain mixed at this point.

Conclusions

The predictability of late childhood linguistic, cognitive, and academic outcomes based on early childhood speech/language competence supports the utility of these classifications. Many significant associations were found between time 1 speech/language cluster membership and follow-up speech/language, cognitive, developmental, demographic, and academic measures. However, it should be stressed that these findings are correlational and caution must be exercised in interpreting them. Potentially important factors such as speech/language intervention have not been taken into consideration in these analyses. While study data do reveal that a considerable number of children included in these analyses received speech/language treatment \((n = 48)\), treatment was not randomly assigned and consequently its impact cannot be adequately assessed in this study. However, selectivity factors certainly play a role in who receives treatment, since the most severely disordered children are most likely to receive treatment and least likely to improve at follow-up (Beitchman et al., 1993). This study provides one of the first empirically grounded classification systems for early childhood speech/language impairment. While it provides a foundation from which to build, it is important in future work to take into account baseline and intermittent variables in accurately assessing the effect of early childhood speech/language disorders on outcome. Factors such as comprehensive preschool intervention programs and psychosocial factors such as the experience of early trauma have not been incorporated in these analyses. Taking into account such variables will provide a more elaborate model for understanding the association between early speech/language impairment and late childhood linguistic, cognitive, and academic outcome. In addition to these outcomes, early childhood speech/language profiles are discussed in terms of behavioral, social, and emotional outcomes in part II (Beitchman et al., 1996).

REFERENCES


Bankson NW (1977), Bankson Language Screening Test. Baltimore: University Park Press

Beevy KE (1982), The Developmental Test of Visual-Motor Integration. Cleveland: Modern Curriculum Press


Bishop DVM, Edmondson A (1987), Specific language impairment as a maturational lag: evidence from longitudinal data on language and motor development. Dev Med Child Neurol 29:442–459

BEITCHMAN ET AL.

Carrow E (1973), Screening Test for Auditory Comprehension of Language, 5th ed. Boston: Teaching Resources Corp
Fundudis T, Kelving I, Gazide R (1979), Speech Retarded and Deaf Children: Their Psychological Development. London: Academic Press
Newcomer PL, Hammill DD (1977), The Test of Language Development (TOLD). Austin, TX: Empire Press
Wechsler D (1967), Manual for the Wechsler Preschool and Primary Scale of Intelligence. New York: Psychological Corp