3-26-2021

Reliability of speech-language pathologists' categorizations of preschoolers' communication impairments in practice

BJ Cunningham
bj.cunningham@uwo.ca

Janis Oram Cardy
janis.cardy@uwo.ca

1 Reliability of speech-language pathologists' categorizations of preschoolers'
2 communication impairments in practice
3
4
5
6
7 Barbara Jane Cunningham, PhD, SLP(C), PhD[a]
8 Western University
9
10 Janis Oram Cardy, PhD, S-LP(C)[a]
11 Western University
12
13   a) School of Communication Sciences and Disorders, Western University, Elborn
14   College, 1201 Western Road, London, ON N6G 1H1
15
16
17
18
19
20
21
22
23
24
25
26
27 **Corresponding author**: Barbara Jane Cunningham PhD., SLP(C)
28 Assistant Professor
29 School of Communication Sciences and Disorders, Western University
30 Elborn College, Room 2516
31 1201 Western Road
32 London, ON N6G 1H1
33 Email: bj.cunningham7@uwo.ca
34 Phone: (519) 661-2111 x88179
35 Fax: (519) 850-2369
36

43

44 **Abstract**

45 **Purpose**. An efficient and reliable way to categorize children's communication

46 impairments based on routine clinical assessments is needed to inform research and

47 clinical decisions. This preliminary study assessed interrater reliability of speech-

48 language pathologists' categorization of preschoolers' speech, language, and

49 communication impairments using a clinical consensus document.

50 **Method**. Six speech-language pathologists at three community sites worked in pairs to

51 assess 38 children aged 1-5 years, then used the clinical consensus document to

52 categorize children's communication impairments broadly. Identified language and

53 speech sound impairments were further sub-categorized.

54 **Results**. Speech-language pathologists had substantial to almost perfect agreement for

55 three broadly-focused impairment categories. Agreement for whether language

56 difficulties/disorders were developmental or associated with a biomedical condition was

57 almost perfect, but moderate for whether difficulties impacted receptive or expressive

58 language, or social communication skills. Agreement was fair for rule-based speech

59 delays/disorders, but low for motor-based or mixed speech impairments.

60 **Conclusions.** Results support use of the clinical consensus document to collect data

61 for reliable categories. Additional work is needed to confirm reliability for some broadly-

62 focused impairment categories and for sub-categorization of speech impairments.

63   In pediatric speech-language pathology research, methods for identifying children with

64   different communication impairments are usually standardized across children, and

65   assessments are typically completed by trained research personnel. This approach is

66   used in epidemiological studies that estimate the prevalence of communication

67   difficulties and disorders in children (Broomfield & Dodd, 2001; Eadie et al., 2015;

68   Korpilahti et al., 2016; Nelson et al., 2006; Tomblin et al., 1997). For example, Tomblin

69   et al. (1997) screened 7,218 children using a brief language test, and conducted a

70   comprehensive battery of diagnostic language and IQ tests on 2,084 children. Eadie et

71   al. (2015) completed parent report and standardized tests for 1,494 children at 8, 12,

72   24, and 48 months. Using a standard research approach for identifying impairments in

73   large samples of children is not only useful for estimating prevalence, but also for

74   achieving other scientific and clinical aims such as evaluating costs and benefits of

75   interventions, justifying fiscal support for programs, quantifying risk factors, and

76   supporting planning for community prevention programs (Antoniadis & Lubker, 1997).

77   The standard research approach to identifying communication impairments in

78   children is less feasible for researchers seeking to use real-world clinical samples, and

79   for systems conducting program evaluations. In practice, speech-language pathologists

80   (SLPs) do not typically follow a standard assessment protocol across children, but

81   rather use a unique combination of standardized tests, observation, parent/teacher

82   report, and informal assessment for each child they see (Caesar & Kohler, 2009;

83   Dockrell et al., 1997; Skahan et al., 2007; McLeod & Baker, 2014). Methods vary based

84   on practical factors such as each child's age and temperament, parent preferences and

85   concerns, the practice context, availability of different tools, and SLPs' competencies

86   and preferences. Therefore, identification of communication impairment types in clinical

87   cohorts currently relies on assessments conducted in real-world contexts that are often

88   subject to variation in methods across children and SLPs.

89   The lack of tools and approaches for reliably categorizing children's

90   communication impairments in the clinical context is a barrier to the use of data to

91   inform research and clinical decisions. As an example, Cunningham et al. (2018) were

92   unable to stratify outcomes by impairment type using data from 46,872 children in a

93   large health system because data on impairment type were entered without any

94   consistency or agreed-upon definitions, or were missing entirely. An efficient and

95   reliable way for SLPs to categorize children's communication impairments based on

96   their routine clinical assessments is needed. Such a method would allow research

97   teams and programs to collect accurate data about the proportion of children in different

98   communication disorder categories, conduct analyses within and across categories,

99   better understand clinical caseloads, and more. When combined with outcome data, this

100 type of clinical data can be used to stratify outcomes by impairment type, resulting in

101 more targeted evidence to support decisions surrounding wait-list management, triage,

102 therapy services, and resource allocation at both clinical and research levels.

103   Several studies have assessed reliability of SLPs' assessments using different

104 clinical tools. For example, Mumby et al. (2007) explored reliability for SLPs' diagnoses

105 of apraxia of speech by having them watch post-stroke videos and rate the presence

106 and severity of apraxia of speech. SLPs demonstrated high interrater agreement for

107 both. John and Enderby (2000) assessed reliability of SLPs' ratings of impairment,

108 activity, participation, and well-being on the Therapy Outcome Measures using both

109 standardized samples and assessments conducted in the clinical context, with the

110 highest reliability observed for assessments done in the clinical context. Other related

111 studies assessed interrater reliability for SLPs' use of classification systems (Barty et

112 al., 2016; Hidecker et al., 2011). Most of these studies investigated reliability for specific

113 populations. Reliability for the categorization of impairments in children with general

114 communication difficulties and disorders has not yet been reported.

115   In 2017-18, our team began work to develop a clinically feasible and reliable

116 method for SLPs to report data on children's communication impairments. More

117 specifically, we worked with expert SLPs from across the Ontario Preschool Speech

118 and Language Program to establish consensus about the categories and definitions of

119 communication impairments for preschoolers in their large provincial program that

120 serves over 60,000 children each year (Cunningham et al., 2019). SLPs reached

121 consensus in four rounds of a modified Delphi study (Izaryk & Skarakis-Doyle, 2017),

122 and the final categories and definitions were reported in Appendix B in Cunningham et

123 al. (2019). With permission, these categories and definitions are presented here as

124 supplementary material. This work was a first step towards the development of a data

125 collection tool that, in part, could be used by SLPs to quickly collect and report reliable

126 data about children's communication impairments at program and population levels.

127 Once developed, the tool can be used to collect data that are ecologically valid, which

128 can be used to ensure the right children receive the right treatments at the right time,

129 thereby maximizing the cost effectiveness of services (Jobse et al., 2014).

130   The purpose of this preliminary study was to have SLPs use Appendix B from

131 Cunningham et al. (2019) as a *clinical consensus document* to determine whether they

132 could reliably categorize preschoolers' communication impairments in clinical practice,

133 where assessment methods are not standard. Attempting to establish reliability was an

134 important step in the development of a clinical data collection tool.

Method

Institutional Review Board

137   This work was part of a larger quality improvement and program evaluation

138 project being completed with the Ontario Preschool Speech and Language Program.

139 Western University's Research Ethics Board (REB) considered the project not to be

140 research as described in the Canadian Tri-Council Policy Statement V.2 (Research

141 Exempt from REB Review, Article 2.4).

142 **Participants**

143   Six SLPs representing three Ontario Preschool Speech and Language Program

144 community sites (2 SLPs from each site) participated. All SLPs had Masters-level

145 training in speech-language pathology, an average of 14.7 years of clinical experience

146 ($SD = 3.93$), and an average of 12.5 ($SD = 6.1$) years of experience working exclusively

147 with preschoolers. Five SLPs conducted assessments with preschoolers on a weekly

148 basis and one conducted assessments monthly.

149   SLPs completed assessments of 39 children, with each community site

150 completing between 11 and 15 assessments. Assessments lasted approximately one

151 hour and included a mix of standardized testing (e.g., Clinical Evaluation of Language

152 Fundamentals Preschool; Wiig et al., 2004), informal assessment, and parent interview.

153 They were conducted at the discretion of the SLP as per their clinical judgement,

154 standard of care at their site, and based on the needs and preferences of the child and

155 family. Assessments were scheduled based on SLPs' availability, and children were

156 assigned to participating SLPs based on their position on the waiting list.

157 Complete data were obtained for 38 children ($M = 38.31$ months, $SD = 13.65$, 55%

158 male). Most ($n = 37$) had no medical diagnoses, and three were undergoing testing for

159   suspected neurodevelopmental disability. Other reported concerns included premature

160   birth ($n = 1$), behavioural issues ($n = 1$), English as a second language ($n = 2$), recurrent

161   ear infections ($n = 1$) and low socioeconomic status ($n = 1$). Children's communication

162   was described using the Communication Function Classification System (CFCS;

163   Hidecker et al., 2011), and assessed using Focus on the Outcomes of Communication

164   Under Six (FOCUS; Thomas-Stonell et al., 2010). Both were required as part of a

165   program evaluation project. The FOCUS is a parent-report outcome measure used to

166   assess growth in communicative participation. The CFCS is used to categorize

167   communication abilities into one of five levels of function (from effective sender and

168   receiver with all partners, to seldom effective sender and receiver with familiar partners).

169   **Materials**

170   SLPs used the clinical consensus document to categorize children's

171   communication impairments. The document includes three sections. First, SLPs must

172   describe children's communication broadly using seven categories (age appropriate

173   communication, speech sound delay or disorder, language difficulty or disorder, fluency

174   disorder, voice or resonance disorder, feeding or swallowing disorder, and emergent

175   literacy concern). If a language difficulty or disorder is identified, SLPs must further sub-

176   categorize it, first as either developmental language difficulty/disorder or language

177   difficulty/disorder associated with a biomedical condition (Bishop et al., 2017), and

178   second, as involving receptive language, expressive language, and/or social

179   communication (Cunningham et al., 2019). If a speech sound delay or disorder is

180   identified, SLPs must further sub-categorize it as being rule or motor based. If motor

181   based, SLPs must identify the impairment as articulation, childhood apraxia of speech,

182   or dysarthria (McLeod & Baker, 2017). SLPs entered anonymized data into an online

183 survey that followed the sections of the clinical consensus document in REDCap, a

184 secure online data collection system hosted at Western University (Harris et al., 2009).

185 **Procedure**

186   Pairs of SLPs at three community sites completed assessments between

187 September 2018 and April 2019. Prior to beginning data collection, SLPs were asked to

188 familiarize themselves with the clinical consensus document and study procedures.

189 Procedures differed slightly across sites. At two sites, one SLP conducted and

190 videotaped each assessment session. Videotaped sessions were focused on the child

191 and their interactions with the SLP. Videotaping was stopped before the SLP presented

192 assessment observations and recommendations to the family. Immediately following the

193 assessment appointment, the SLP saved the video recording using a non-identifying

194 code and shared the video and child's chronological age with the second SLP, who

195 watched the video. Both SLPs used the clinical consensus document to independently

196 categorize children's communication impairments, which were subsequently entered

197 online using a secure REDCap link. Each child was assigned an anonymized

198 identification number by the assessing SLP. One SLP completed in-person

199 assessments for the first half of children at their site, and then SLP partners switched so

200 that the assessor became the one who viewed videotaped sessions. At the third

201 community site, one SLP conducted the assessment session, but rather than viewing a

202 videotape, the partner-SLP observed through a one-way mirror, leaving before the

203 assessing SLP discussed observations and recommendations with the family. These

204 SLPs then independently entered their classifications into REDCap.

205 **Analyses**

206   Since this study aimed to determine interrater reliability for SLPs' categorization

207 of children's communication impairments, Cohen's Kappa statistic was used to calculate

208 agreement between SLP pairs (Landis & Koch, 1977). Agreement was calculated

209 separately for the broadly-focused impairment categories, language difficulty/disorder

210 sub-categories, and speech sound delay/disorder sub-categories. For each component,

211 SLPs were able to select any combination of impairment categories, so overall

212 agreement as well as agreement for each individual category was calculated.

213   An important consideration for the analyses was the minimum sample size

214 required to detect a meaningful kappa coefficient. Main factors affecting effect size for

215 our analyses included the number of categories being compared and the kappa

216 coefficients used for hypothesis testing (Bujang & Baharum, 2017). When categorizing

217 children's communication impairments broadly, agreement across seven categories was

218 compared, and when sub-categorizing speech sound delay/disorder and language

219 difficulty/disorder, agreement across two and three categories was compared. The null

220 hypothesis in all cases was no agreement between raters ($K_1 = 0$), and we expected

221 substantial agreement between raters across all categories ($K_2 = 0.9$). With 80% power

222 and alpha set at 0.05, a minimum sample size of 7 for the 2x2 categorizations, 5 for the

223 3x3 categorizations, and 2 for the 7x7 categorizations was required to detect the

224 minimum kappa coefficient of 0.9. Where the minimum sample size requirement was

225 not met, percent agreement is reported, but kappa coefficients were not calculated.

226   **Results**

227 **Categorization of broad impairment categories**

228   Pairs of SLPs at three community sites provided complete data for 38 children.

229 Percent agreement and kappa coefficients were calculated for each broadly-focused

230   impairment category when seven or more children were identified by at least one SLP

231   as belonging in that category (see Table 1). Based on Landis and Koch (1977),

232   benchmarks for the strength of agreement between raters using Kappa are poor (k

233   <0.00), slight (k = 0.00-0.21), fair (k = 0.21-0.40), moderate (k = 0.41-0.60), substantial

234   (k = 0.61-0.80), and almost perfect (k = 0.81-1.00).

235   SLPs had substantial agreement for children with age appropriate

236   communication and language difficulty/disorder, and almost perfect agreement for

237   children with speech sound delay/disorder. Percent agreement was high for fluency

238   disorder, voice and resonance disorder, and emergent literacy concern, but too few

239   children were identified to calculate a kappa coefficient for these categories.

240   When asked to classify children's communication broadly, SLPs could select any

241   combination of impairment categories (e.g., speech sound delay/disorder + language

242   difficulty/disorder), so overall interrater agreement was also calculated. The feeding and

243   swallowing disorders category was not included in the calculation as no children were

244   identified in this category. Across all possible combinations of impairment categories,

245   percent agreement was moderate (63.16%, k = 0.54, z = 7.19, p < .01, *95%CI* [.46-.56]).

246   Disagreements were most often related to the presence of an emergent literacy concern

247   or fluency disorder in addition to a language difficulty or disorder. With these categories

248   removed, overall agreement improved to 79.41% (k = 0.71, z = 7.32, p < .01, *95%CI*

249   [.68-.83]).

250   **Sub-categorization of language difficulty or disorder**

251   Sub-categorizations were explored for children identified as having language

252   difficulty or disorder broadly. SLPs had almost perfect agreement that 3 children had

253   language difficulty/disorder associated with a biomedical condition, and disagreed for

254   one child (94.74% agreement, k = 0.83, z = 3.66, p < .01, *95%CI* [.50-1.00]). Almost

255   perfect agreement was also observed for whether language impairments were

256   developmental. SLPs agreed that 15 children had developmental language

257   difficulty/disorder, and disagreed for one child (94.74% agreement, k = 0.83, z = 3.66, p

258   < .01, *95%CI* = [.50-1.00]).

259   Percent agreement and kappa scores for SLPs' categorization of language

260   difficulties is presented in Table 2. Moderate agreement was observed for the

261   categorization of receptive and expressive language difficulty (see Table 2). When

262   asked to sub-categorize language difficulties, SLPs could select any combination of

263   receptive, expressive, and social communication difficulties. Overall agreement was

264   substantial at 78.98% (k = 0.63, z = 4.75, p < .01, *95%CI* [.47-0.85]), with

265   disagreements most often related to the presence of receptive language and social

266   communication difficulties in addition to an expressive difficulty.

267   **Sub-categorization of speech sound delay and disorder**

268   Pairs of SLPs agreed that 9 children had speech sound delay or disorder and

269   sub-categorizations for those children were explored. Overall agreement for whether

270   impairments were rule-based, motor-based or mixed was fair (66.67%, k = 0.36, z =

271   1.29, p = .10, *95%CI* = [-0.62-1.00]). SLPs had fair agreement that speech difficulties

272   were rule-based for 4 children. Agreement for each individual category is presented in

273   Table 2. There was low agreement for whether impairments were motor based, but

274   SLPs agreed that two of the four identified children had a speech sound delay/disorder

275   that was mixed. Agreement for type of motor-based impairment was not calculated as

276   only five children were identified and agreement for the presence/absence of this

277   impairment was low.

Discussion

278

279   One pair of SLPs at each of three community sites reliably categorized

280 preschoolers' communication impairments into broad categories. Percent agreement

281 and kappa coefficients were substantial to almost perfect when SLPs judged children to

282 have age appropriate communication, speech sound delay/disorder, or language

283 difficulty/disorder.

284   Agreement was lower (moderate) when overall reliability across all possible

285 combinations of children's broadly-focused communication impairments was assessed.

286 We believe disagreements were due in part to the prevalence of those disorder types

287 and SLPs' exposure to them (Feinstein & Cicchetti, 1989), but they may also be due to

288 a lack of standardized measures being used. Informally, we know it was common for

289 SLPs to forego standardized assessments for young children, relying instead on

290 informal and criterion-referenced assessment to identify an impairment. With these

291 categories removed, percent agreement increased to substantial (85%, $k = 0.79$).

292 Moving forward, it may be prudent to analyze data for children's broadly-focused

293 impairments by individual impairment categories rather than overall.

294   Almost perfect agreement was observed for SLPs' categorizations of children's

295 language difficulty/disorder as being developmental or associated with a biomedical

296 condition, indicating SLPs could reliably differentiate between these categories. This

297 terminology for children's language impairments (Bishop et al., 2017) was new to many

298 of the SLPs participating in this study, and we believe the clinical consensus document

299 served as a good knowledge translation tool to support differentiation between these

300 categories in the preschool population. By developing the document together with SLPs,

301 we were able to create written materials with language use and formatting that were

302 easily assimilated by SLPs, and unforeseen implementation barriers may have been

303 addressed through the development process (Grimshaw et al., 2012). More targeted

304 knowledge translation work will be needed as the data collection tool is finalized and

305 launched to ensure all SLPs have the knowledge to differentiate between these

306 categories. This will likely include a multifaceted approach involving a combination of

307 educational outreach and recruitment of local opinion leaders (Grimshaw et al., 2012).

308   Agreement was lowest when SLPs were asked to sub-categorize children's

309 speech sound delay/disorder. SLPs had fair agreement for when an impairment was

310 rule-based, but disagreed about the presence of a motor component. This finding may

311 be explained in part by a new theory on articulatory phonology (Namasivayam et al.,

312 2020), which argues that phonetic and phonological skills are interconnected and affect

313 one another. It may therefore be that SLPs struggled because children in fact had mixed

314 speech impairments that could not be differentiated.

315   Findings are consistent with studies reporting SLPs can reliably categorize

316 children's impairments (John & Enderby, 2000; Mumby et al., 2007) and abilities (Barty

317 et al., 2016; Hidecker et al., 2011) following clinical assessments. Findings support

318 including the impairment categories outlined in the clinical consensus document in the

319 developing data collection tool with some modifications and cautions. The broadly-

320 focused impairment categories can be included, but caution should be used when

321 interpreting data for the categories where Kappa could not be calculated. Data should

322 be interpreted for individual categories rather than for combinations of categories, as

323 additional work is needed to support reliable categorization for multiple categories. The

324   language-disorder sub-types and sub-categories of expressive and receptive difficulties

325   can also be included.

326   **Limitations**

327   Within the language difficulties/disorders category, caution should be used when

328   interpreting categorizations for social communication impairment due to our inability to

329   calculate a Kappa coefficient for this category. Currently, there is not sufficient evidence

330   to suggest SLPs can reliably differentiate children's speech sound delays/disorders and

331   additional work is needed before these sub-categories are used. Other limitations

332   include the small number of children identified in some impairment categories, the

333   number of SLPs who completed assessments, and that the study was conducted in

334   relatively uncontrolled community settings. Our analysis of SLPs' agreement provides

335   some evidence of reliability, but it is possible that there were data entry errors or that

336   SLPs described children's communication inaccurately.

337   Next steps are to revise the categories based on SLPs' reliability, develop

338   additional sections for the data collection tool that will amass information about predictors

339   of outcome, then pilot the tool in its entirety to collect clinical data and perceptions of

340   clinical utility and usability.

341   **Conclusions**

342   SLPs' categorizations of children's communication impairments using a clinical

343   consensus were generally reliable. Findings support use of the clinical consensus

344   document to collect data for most impairment categories, but additional work is needed

345   to confirm reliability for some less commonly occurring broadly-focused impairment

346   categories and for sub-categorization of speech sound delay and disorder.

347 **References**

348 Antoniadis, A., & Lubker, B. B. (1997). Epidemiology as an essential tool for establishing

349   prevention programs and evaluating their impact. *Journal of Communication*

350   *Disorders, 30*(4), 269-284.

351 Barty, E., Caynes, K., & Johnston, L. M. (2016). Development and reliability of the

352   Functional Communication Classification System for children with cerebral palsy.

353   *Developmental Medicine and Child Neurology, 58*(10),1036-41.

354 Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2017). Phase 2 of

355   CATALISE: a multinational and multidisciplinary Delphi consensus study of problems

356   with language development: Terminology. *Journal of Child Psychology and*

357   *Psychiatry, 58*(10): 1068-1080.

358 Broomfield, J., & Dodd, B. (2001). Mainstream paediatric speech and language therapy

359   service population: epidemiology. *International Journal of Language &*

360   *Communication Disorders, 36,* 447-52.

361 Bujang, A., & Baharum, N. (2017). Guidelines of the minimum sample size requirements

362   for Cohen's Kappa. *Epidemiology Biostatistics and Public Health, 14*(2), e12267-1-

363   10.

364 Caesar, L. G., & Kohler, P. D. (2009). Tools Clinicians Use: A Survey of Language

365   Assessment Procedures Used by School-Based Speech-Language Pathologists.

366   Communication Disorders Quarterly, 30(4).

367 Cunningham, B. J., Hanna, S. E., Rosenbaum, P., Thomas-Stonell, N. & Oddson, B.

368   (2018). Factors contributing to preschoolers' communicative participation outcomes:

369   Findings from a population-based longitudinal cohort study in Ontario Canada.

370   *American Journal of Speech-Language Pathology, 27*(2), 737-750.

371 Cunningham, B. J., Kwok, E. L., Turkstra, L., & Oram Cardy, J. (2019). Establishing

372 consensus about preschoolers' speech-language impairments in a large-scale

373 community program. *Journal of Communication Disorders*, Nov-Dec, 82: 105925.

374 Dockrell, J., George, R., Lindsay, G., & Roux, J. (1997) Problems in the Identification

375 and Assessment of Children with Specific Speech and Language Difficulties.

376 *Educational Psychology in Practice, 13*(1), 29-38.

377 Eadie, P., Morgan, A., Ukoumunne, O. C., Ttofari, E. K., Wake, M., & Reilly, S. (2015).

378 Speech sound disorder at 4 years: prevalence, comorbidities, and predictors in a

379 community cohort of children. *Developmental Medicine & Child Neurology, 57*(6):

380 578-584.

381 Feinstein, A. R., & Cicchetti, D. V. (1989). High agreement but low kappa: I. The

382 problems of two paradoxes. *Journal of Clinical Epidemiology, 43*(6), 543-549.

383 Grimshaw, J.M., Eccles, M.P., Lavis, J.N., Hill, S. J., & Squires, J. E. (2012). Knowledge

384 translation of research findings. *Implementation Science,* 7**,** 50.

385 Harris, P.A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009).

386 Research electronic data capture (REDCap): A metadata-driven methodology and

387 workflow process for providing translational research informatics support. *Journal of*

388 *Biomedical Informatics* 42(2), 377–81.

389 Hidecker, M. J., Paneth, N., Rosenbaum, P., Kent, R. D., Lillie, J., Eulenberg, J. B.,

390 Chester, K., Johnson, B. Michalsen, L., Evatt, M., & Taylor, K. (2011). Developing

391 and validating the Communication Function Classification for individuals with cerebral

392 palsy. *Developmental Medicine and Child Neurology, 53*(8), 704-710.

393 Izaryk, K., & Skarakis-Doyle, E. (2017). Using the Delphi technique to explore complex

394 concepts in speech-language pathology: An illustrative example from children's

395   social communication. American Journal of Speech-language Pathology, 26,

396   1225–1235.

397   Jobse, B. N., Adams, O., & Levy, I. (2014). Population health management: A Canadian

398   perspective on the future of health systems. *Canadian Journal of Physician*

399   *Leadership, 4*(1), 8-15.

400   John, A., & Enderby, P. (2000). Reliability of speech and language therapists using

401   therapy outcome measures. *International Journal of Language and Communication*

402   *Disorders, 35*(2), 287-302.

403   Korpilahti, P. Kaljonen, A., & Jansson-Verkasalo, E. (2016) Population-Based Screening

404   for Language Delay: Let's Talk STEPS Study. *Psychology*, 7, 205-214.

405   Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for

406   categorical data*. Biometrics, 33*(1), 159–174.

407   McLeod, S., & Baker, E. (2014). Speech-language pathologists' practices regarding

408   assessment, analysis, target selection, intervention, and service delivery for children

409   with speech sound disorders. *Clinical Linguistics & Phonetics, 28*(7-8), 508-531.

410   McLeod, S. & Baker, E. (2017). Classification, causes, and co-occurrence. In S. McLeod

411   & E. Baker (Eds), *Children's speech: An evidence-based approach to assessment*

412   *and intervention* (pp. 37-59). Boston, MA: Pearson.

413   Mumby, K., Bowen, A., & Hesketh, A. (2007). Apraxia of speech: how reliable are

414   speech and language therapists' diagnoses? *Clinical Rehabilitation, 21*(8), 760-767.

415   Namasivayam, A. K., Colemn, D., O'Dwyer, A., & van Lieshout, P. (2020). Speech

416   Sound Disorders in Children: An Articulatory Phonology Perspective. *Frontiers in*

417   *Psychology, 10,* 2998.

418 Nelson, H. D., Nygren, P., Walker, M., & Panoscha, R. (2006). Screening for speech and

419   language delay in preschool children: systematic evidence review for the US

420   Preventive Services Task Force. *Pediatrics, 117*(6), 2336.

421 Skahan, S. M., Watson, M., & Lof, G. L. (2007). Speech-language pathologists'

422   assessment practices for children with suspected speech sound disorders: Results of

423   a national survey. *American Journal of Speech Language Pathology, 16*(3), 246-259**.**

424 Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M.

425   (1997). Prevalence of specific language impairment in kindergarten children. *Journal*

426   *of Speech Language & Hearing Research, 40*(6), 1245-60.

427 Wiig, E. H., Secord, W. A., & Semel, E. (2004). *Clinical Evaluation of Language*

428   *Fundamentals Preschool, Second edition*. San Antonio, TX: Harcourt Assessment

429   Inc.

Table 1.

Percent agreement and kappa scores for classification using broadly-focused impairment categories.

Impairment category (# *identified by at least one SLP*)	% agreement	kappa	Landis & Koch (1977) Strength of Agreement
Age appropriate communication skills (*n=13*)	84.21%	*k* = 0.62, *z* = 3.84, *p* < 0.01, *95%CI* [.34-.89]	Substantial
Speech sound delay or disorder (*n*=11)	94.74%	*k* = 0.86, *z* = 5.33, *p* < 0.01, *95%CI* [.68-1.00]	Almost Perfect
Language difficulty or disorder (*n*=23)	89.47%	*k* = 0.79, *z* = 4.88, *p* < 0.01, *95%CI* [.59-0.98]	Substantial
Fluency disorder (*n*=4)	92.11%	n/a	
Voice or resonance disorder (*n*=3)	92.11%	n/a	
Feeding or swallowing disorder (*n*=0)	n/a	n/a	
Emergent literacy concern (*n*=4)	97.37%	n/a	

430

Table 2.			
Percent agreement and kappa scores for sub-classification of speech and language difficulty and disorder sub-types.			
Sub-difficulty (# identified by at least one SLP)	% agreement	kappa	Landis & Koch (1977) Strength of Agreement
Receptive language difficulty (*n*=7)	78.95%	*k* = 0.46, *z* = 1.99, *p* = 0.02, *95%CI* [.005-.91]	Moderate
Expressive language difficulty (*n*=18)	89.47%	*k* = 0.46, *z* = 2.37, *p* < 0.01, *95%CI* = [-.14-1.00]	Moderate
Social communication difficulty (*n*=5)	89.47%	n/a	
Rule based speech difficulty (*n*=7)	66.67%	*k* = 0.31, *z* = 0.95, *p* = 0.17, *95%CI* = [-.31-.92]	Fair
Motor based speech difficulty (*n*=1)	88.89%	n/a	
Mixed speech difficulty (*n*=4)	77.78%	n/a	

431