

KEY WORDS

SPEECH-LANGUAGE
PATHOLOGY

OUTCOME RESEARCH

RELIABILITY

PSYCHOMETRIC PROPERTIES

Bruce Oddson, PhD

School of Human Kinetics,
Laurentian University
935 Ramsey Lake Rd.
Sudbury, ON
P3E 2C6
Canada

**Karla Washington, PhD,
CCC-SLP, S-LP(C)**

Department of
Communication Sciences
and Disorders
University of Cincinnati
Cincinnati, OH
USA, 45267

**Bernadette Robertson,
S-LP(C)**

Holland Bloorview Kids
Rehabilitation Hospital
150 Kilgour Rd
Toronto, ON
M4G 1R8
Canada

**Nancy Thomas-Stonell,
BSc, DSP, S-LP(C)**

Holland Bloorview Kids
Rehabilitation Hospital
150 Kilgour Rd,
Toronto, ON
M4G 1R8
Canada

Peter Rosenbaum, MD

CanChild Centre for
Childhood Disability
Research
Institute for Applied
Health Sciences
1400 Main Street West
Hamilton, ON
L8S 1C7
Canada

RESEARCH NOTE:

Inter-rater Reliability of Clinicians' Ratings of
Preschool Children Using the FOCUS®:
Focus on the Outcomes of Communication
Under Six

NOTE DE RECHERCHE :

La Fiabilité Inter-évaluateurs des Évaluations
d'enfants d'âge Préscolaire par des Cliniciens
Utilisant la Méthode FOCUS® (*Focus on the
Outcomes of Communication Under Six*)

Bruce Oddson
Karla Washington
Bernadette Robertson
Nancy Thomas-Stonell
Peter Rosenbaum

Abstract

The present study was conducted to evaluate the reliability of the FOCUS® (*Focus on the Outcomes of Communication Under Six*). The FOCUS® is intended to measure the functional change in real-world communication associated with speech-language therapy. Clinicians administer this test by interviewing parents before and after treatment. The inter-rater reliability of detected change was high when evaluating preschool children (n = 13). Results indicate that this measure is reliable for both practical and research purposes. The necessity of establishing reliability is discussed, as are limitations of the FOCUS®.

Abrégé

La présente étude a été menée pour évaluer la fiabilité de la méthode FOCUS® (*Focus on the Outcomes of Communication Under Six*). FOCUS® veut mesurer le changement fonctionnel dans la communication associée aux thérapies en orthophonie. Les cliniciens administrent ce test en interviewant les parents avant et après traitement. La fiabilité inter-évaluateurs du changement détecté était élevée lors de l'évaluation d'enfants d'âge préscolaire (n = 13). Les résultats indiquent que cette mesure est fiable à la fois en clinique et en recherche. La nécessité d'établir la fiabilité est discutée, ainsi que les limites associées à FOCUS®.

Children need proficiency in communication at school, home, work and play (Byles, 2005; McCormack, 2009) and communication is central to the normal acquisition of many cognitive and academic skills, including literacy (Catts, Bridges, Little, & Tomblin, 2008; Catts, Fey, Tomblin, & Zhang, 2002; Justice, Bowles & Skibbe, 2006; Warr-Leeper, 1993; Warr-Leeper, Wright, & Mack, 1994). Preschool and school-age children with communication disorders are more likely to demonstrate social and behavioural problems that can negatively impact peer acceptance and popularity in school (McCabe, 2005; Warr-Leeper et al., 1994). Deficits in communication have an overall negative impact on children's abilities to participate in every day life situations such as learning in school, taking turns, playing, and most critically, establishing and maintaining friendships with others (Warr-Leeper et al., 1994; Washington, 2007).

Communication disorders are defined as persistent difficulties in the use or understanding of spoken and/or written language (Paul, 2001). These disorders frequently become apparent in early childhood (McLeod & Harrison, 2009), with approximately 6% of the preschool population identified as having a significant communication disorder (Law, Boyle, Harris, Harkness & Nye, 2000). They can be life-long and may result in poor life outcomes, including failure to complete high school, difficulties obtaining and maintaining gainful employment and an increased incidence of psychiatric disorders and arrests (Warr-Leeper, 2001; Johnson, Beitchman, Brownlie, 2010). However, there is some evidence supporting the positive outcomes of intervention for children with communication disorders (Law, Garrett, & Nye, 2003; Washington, Warr-Leeper, & Thomas-Stonell, 2011). Thus, measuring the outcomes of intervention is important in order to establish the effectiveness of speech and language therapy for this population of children (Thomas-Stonell, Oddson, Robertson, & Rosenbaum, 2009; 2010).

Outcome measurement is the backbone of evidence-based practices (Beitchman, Nair, Clegg, & Patel, 1986; Majnemer & Mazer, 2004) and is needed to determine the impact of speech and language therapy. Information about outcomes is critical for informing speech-language pathologists (S-LPs) about which children, with which kinds of problems, will benefit most from specific types or schedules of intervention (Beitchman et al., 1986; Thomas-Stonell et al., 2009). In particular, given the impact of communication impairments on children's lives, it is important to develop tools to measure the real-world outcomes of speech-language therapy. Although it is a common practice to use standardized tests to evaluate change, this is not a correct use because most standardized tests are neither reliable nor sensitive when used this way (Kerr, Guildford & Bird, 2003; Rosenbaum, et al., 1990). For example, when a change in child's age leads to a change of

normative group, scores (at least centiles) can go down even in the midst of therapeutic progress.

A critical component in establishing intervention outcomes is the use of reliable and valid outcome measures (Thomas-Stonell et al., 2009; 2010; Yorkston, Klasner, & Swanson, 2001). Outcome measures are simply 'tools', and as such must have demonstrated utility, reliability, validity, and in the case of change-detecting measures, clinical responsiveness (Law et al., 2000). Most of the measurement tools available to S-LPs do not adequately measure the full range of progress observed in successful interventions (Thomas-Stonell et al., 2009).

Outcome measures need to have established validity and responsiveness before being adopted into widespread use (Hall, 1997; Turk, 1997; Unsworth et al., 2004; van der Putten, Hobart, Freeman, & Thompson 1999). Tests that do not properly reflect the progress a child has made are likely to lead to inappropriate clinical decisions. On a broader scale, a test that is unsound could easily lead researchers to make poor comparisons between treatment models or programs, and hence, poor clinical and programmatic decisions. Measures that are not valid or adequately responsive may be worse than no measurement at all.

Previous work in the measurement of therapy outcomes in speech-language pathology has not produced measurement tools with proven validity and responsiveness for two reasons (Thomas-Stonell et al., 2009; 2010; Washington, 2007). The first is a paucity of outcome measurement research in the field (Washington, 2007). The second is that the primary interest of most clinicians has been to measure progress specific to the treatment plan (Thomas-Stonell et al., 2009), usually at the 'impairment' level. This has driven the creation of a large number of measures, each targeting a limited range of the communication repertoire (e.g., voice disorders), each based on a particular assessment and treatment approach, and none corresponding to the diversity of the challenges in the preschool population. There are two more broadly-based exceptions – the American Speech and Hearing Association National Outcome Measure System (ASHA NOMS) and the Australian Therapy Outcome Measure System (AusTOMs). The ASHA NOMS does not have proven reliability and responsiveness (Thomas-Stonell et al., 2009). The AusTOMs has been shown to be reliable (Perry et al., 2004) and valid (Unsworth et al., 2004; van der Putten et al., 1999). However, it provides only a single rating of participation, thus offering little information about which real world skills have changed. Its simplicity may limit its capacity to evaluate change specifically and responsively.

Achieving good communication outcomes for children is important; therefore measuring the outcomes of intervention appropriately is both relevant and necessary.

The FOCUS© is a new and innovative broad-based outcome measure developed to provide a reliable, valid, and effective measure of progress in speech-language therapy for young children. Unlike most speech and language outcome measures, it evaluates changes in both 'Capacity' – what the child is capable of doing in an ideal environment such as a structured, therapeutic sessions – and 'Performance' – what the child does in various environments such as home, nursery school, and daycare. It is intended to identify real world changes in communication (e.g., intelligibility, sentence grammar, vocabulary, socialization) associated with speech therapy (Thomas-Stonell et al., 2009; 2010). The FOCUS© can be completed either by a parent as a checklist of 50 items, or by a clinician after a 15-minute interview with a parent. The vocabulary was developed out of responses from 210 parents of children enrolled in speech-language therapy in a previous study (Thomas-Stonell et al., 2009; 2010). Since the FOCUS© is based on the comments of the parents themselves, the questions have strong face validity and good correspondence to the vocabulary parents are accustomed to using in describing therapy progress. In previous work it was established that the FOCUS© items have high internal consistency (Thomas-Stonell et al., 2009; 2010). When used with parents, the test-retest reliability was found to be very high (Thomas-Stonell et al., 2010).

The present study investigated the inter-rater reliability of clinicians using the FOCUS© to rate change following speech-language interventions with a very broad range of children. Establishing high reliability is a necessary precondition for the appropriate use of any measurement tool; the majority of measures available in speech evaluation do not have proven reliability when looking at their ability to detect change.

Method

Participants

Ethics approval was granted at Holland Bloorview Kids Rehabilitation Hospital, Laurentian University, and each participating centre. Families and clinicians (speech-language pathologists) were recruited from three sites in the province of Ontario, Canada. Families were invited to participate in the study as they were referred to their respective programs. Every family approached agreed to participate in the study (100% recruitment rate). Informed consent was obtained from each participating family.

A total of 13 preschoolers with speech-language impairments and four S-LPs participated. Preschool participants ranged in age from 3 years, 1 month to 6 years, 4 months ($M = 57$ months). Sixty-two percent ($n=8$) were male. Many of the children (62%) had also been identified as having a specific medical diagnoses, including cerebral palsy ($n=5$) and hypotonia ($n=3$). Participating S-LPs

rated each preschooler's communication level using the Communication Function Classification System (CFCS) (Hidecker et al., 2011). The CFCS classifies communication performance into one of five levels (5 = lowest function, 1 = best function). The CFCS focuses on Activity and Participation levels as described in the World Health Organization's International Classification of Functioning (Hidecker et al., 2011). A parent, caregiver, and/or a professional who is familiar with the individual selects the person's communication level. Five preschoolers (38%) were classified as "effective sender and receiver with familiar partners" (Level 3 communicator). There was no attrition in this sample.

Procedure

Inter-rater reliability for change scores was evaluated by asking pairs of S-LPs ($n = 4$; including one of the authors) to use the FOCUS© to assess each preschooler ($n = 13$) at two different time points (Time 1 and Time 2). Assessments were based on interaction with the child and interview of a parent(s). Each assessment (Time 1 and Time 2) was completed during the evaluation of speech and language skills. Participating S-LPs independently completed each assessment using the FOCUS©. To establish change scores, it was necessary to administer the FOCUS© before and after intervention. Thus, the FOCUS© was administered during an assessment session and then again within a 2½ -month period (mean 89 days, ranging from 34 to 112), which followed speech-language intervention.

Data Analysis

Using IBM SPSS Statistics (Version 20) the scores for each case were screened for outliers. Two cases were identified as outliers at more than 3 standard deviations from trimmed mean of the differences between each judges' scores. These cases were removed. Inter-rater reliability was assessed using the absolute agreement intra-class correlation coefficient (ICC) calculation.

Results

Based on eleven cases, the inter-rater reliability of FOCUS© change scores was acceptable, $ICC = .70$ (CI .24-.91). Change scores ranged from -18 to +116, $M = 57$, $SD = 36$.

Discussion

The FOCUS© is a new and innovative broad-based outcome measure of preschoolers' communication. Unlike most speech and language outcome measures, it has been designed to evaluate changes in both 'Capacity' – what the child is capable of doing in an ideal environment such as a structured, therapeutic sessions – and 'Performance' – what the child does in various environments.

Reliability is a critical attribute of an outcome measure. Measures with poor reliability cannot be used to measure the outcome of therapy. First, to the extent that reliability is poor, sound individual conclusions cannot be made. Second, even as a research tool, poor reliability will lead to distortions and inconsistent assessments of a concept's relationship to other measured variables (Cochran, 1968; Fuller 1987; Gleser, Carroll, & Gallo, 1987). Few measures in the domain of speech language pathology have established reliability.

The present results indicate that the test-retest and inter-rater reliabilities of the FOCUS© are acceptably high. It is important to note that ICC scores are very sensitive to differences in populations. Our sample was highly heterogeneous and reflected the caseloads of several different programs. This may elevate estimates of ICC values relative to ICCs obtained with more homogenous samples. However, the reported values excluded scores for two clients. In both of the cases the disagreement between clinicians was larger (161 and 86) than any treatment effect we have found in any of the validation studies to date. On examination it was found that one of the clinicians had made the ratings of both cases with insufficient discussion with the parents. The clinician relied primarily on direct observation of the child with only limited input from the parent for FOCUS© items that would have benefitted from parental description and elaboration. Since some FOCUS© items (e.g., plays well with others) might not be directly observable by the clinician at the time of the assessment, it is highly recommended that the completion of these items only be completed after a discussion with parent(s). For the above-mentioned examination, such a discussion was not sufficiently completed, resulting in a discrepancy in FOCUS© ratings between the clinician who completed a more lengthy parent discussion and the one who did not. The manual that comes with the FOCUS© will reflect this experience and be clear about the administration guidelines.

It should also be noted that we report values derived from clinicians. Clinicians score the FOCUS© using a combination of clinical observations and parent report. They should not be considered any more reliable than responses recorded directly by parents. Parent test-retest reliability has already been established as being very high (Thomas Stonell et al., 2010), although we have not been able to evaluate test-retest reliability for two parents separately.

Establishing the inter-rater reliability of the FOCUS© lays the groundwork for its eventual use in clinical settings. Reliability is one necessary attribute for useful measurement. Clinical use requires validity to be established using well-judged criteria. Effectively, an unreliable test cannot be used for individual judgments even if its validity is established – because one could never be sure if a specific measured rating was meaningful or not.

The present work represents a separate and independent step for the larger validation and refinement of this measure.

Limitations. Although we have a diversity of speech and language therapy needs represented, these results are from a small sample. Because of the nature of the recruitment process, we do not know if biases have been introduced by the selection of particular children and parents, or therapist assessors. Although each centre was asked to recruit consecutive families, there is no way to assess the impact of selection that may have occurred. The FOCUS© is intended to be applicable for a very broad range of children with a variety of speech and language therapy needs. The present sample includes children with varied severities of communication problems. This diversity means that we may detect higher reliability than would be found in any given treatment program. Similarly, although a wide range of ages was included, we cannot comment on the reliability for any specific age. Finally, we have not yet assessed the use of this instrument by related professionals, such as teachers or communication disorder assistants.

Conclusion

Reliability is a critical precondition for useful and effective outcome measurement. The test-retest reliability of the FOCUS© is high, and the inter-rater reliability is acceptable. These results suggest that the FOCUS© will be reliable for its many proposed uses across a broad range of communication disorders, diagnoses, severities, and ages. The difficulty found in two cases out of thirteen suggest that some standards for training in the use of the FOCUS© and for a minimal amount of time in contact with the client and parents must be established.

References

- Beitchman, J. H., Nair, R., Clegg, M., & Patel, R. (1986). Prevalence of speech and language disorders in 5-year-old kindergarten children in the Ottawa-Carleton region. *Journal of Speech and Hearing Disorders, 51*, 98–110.
- Byles, J. (2005). The epidemiology of communication and swallowing disorders. *Advances in Speech-Language Pathology, 7*, 1–7.
- Catts, H. W., Bridges, M. S., Little, T. D., & Tomblin, J. B. (2008). Reading achievement growth in children with language impairments. *Journal of Speech, Language, and Hearing Research, 51*, 1569–1579.
- Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45*, 1142–1157.
- Cochran, W. G. (1968). Errors of measurement in statistics. *Technometrics, 10*(1), 637–666.
- Fuller, W. A. (1987). *Measurement Error Models*. New York: John Wiley.
- Gleser, L.J., Carroll, R.J., & Gallo, P.P. (1987). The limiting distribution of least squares in an errors in variables linear regression model. *The Annals of Statistics, 15*, 220–233.
- Hall, K. (1997, June). *Choosing and Using Outcome Assessment Tools*. Paper presented at the 21st. Annual Williamsburg Conference: Practical Strategies for Measuring Rehabilitation Outcomes. Williamsburg, PA.

- Hidecker, M. J. C., Paneth N., Rosenbaum, P. L., Kent R. D., Lillie J., Eulenberg, J.B.,...Taylor K. (2011). Developing and validating the Communication Function Classification System (CFCFS) for individuals with Cerebral Palsy. *Developmental Medicine and Child Neurology*, 53(8), 704-710.
- Johnson, C. J., Beitchman, J. H., & Brownlie, E. B. (2010). Twenty-year follow-up of children with and without speech-language impairments: Family, educational, occupational, and quality of life outcomes. *American Journal of Speech-Language Pathology*, 19, 51-65.
- Justice, L. M., Bowles, R. P., & Skibbe, L. E. (2006). Measuring preschool attainment of print-concept knowledge: A study of typical and at-risk 3- to 5-year-old children using item response theory. *Language, Speech, and Hearing Services in Schools*, 37, 224-235.
- Kerr, M.A., Guildford, S. & Bird, K. R. (2003). Standardized language test use: A Canadian survey. *Journal of Speech-Language Pathology & Audiology*, 27, 10-28.
- Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: Findings from a systematic review of the literature. *International Journal of Language and Communication Disorders*, 35, 165-188.
- Law, J., Garrett, Z., & Nye, C. (2003). *Speech and language therapy interventions for children with primary speech and language delay or disorder*. Cochrane Database of Systematic Reviews, 3.
- Majnemer, A., & Mazer, B. (2004). New directions in the outcome evaluation of children with cerebral palsy. *Seminars in Pediatric Neurology*, 11, 11-17.
- McCabe, P. C. (2005). Social and behavioral correlates of preschoolers with specific language impairment. *Psychology in Schools*, 42(4), 373-387.
- McCormack, J. (2009). Communication impairment in Australian children: A review of recent prevalence studies. *Acquiring Knowledge in Speech, Language, and Hearing*, 9(3), 95-98.
- McLeod, S., & Harrison, L.J. (2009). Epidemiology of speech and language impairment in a nationally representative sample of 4- to 5-year-old children. *Journal of Speech, Language, and Hearing Research*, 52, 1213-1229.
- Paul, R. (2001). *Language disorders from infancy through adolescence*. 2nd Edition. St. Louis, MI: Mosby, Inc.
- Perry A, Morris M, Unsworth C, Duckett S, Skeat J, Dodd K, ... Riley K. (2004). Therapy outcome measures for allied health practitioners in Australia: The AusTOMs. *International Journal of Quality Health Care*, 16, 1-7.
- Rosenbaum, P., Cadman, D., Russell, D., Gowland, C., Hardy, S., & Jarvis, S. (1990). Issues in measuring change in motor function in children with cerebral palsy. A special communication. *Physical Therapy*, 70(2), 125-131.
- Thomas-Stonell, N., Oddson, B., Robertson, B., & Rosenbaum, P. (2009). Predicted and observed outcomes in preschool children following speech and language treatment: Parent and clinician perspectives. *Journal of Communication Disorders*, 42, 29-42.
- Thomas-Stonell, N., Oddson, B., Robertson, B., & Rosenbaum, P. (2010). Development of the FOCUS© (Focus on the Outcomes of Communication Under Six): A communication outcome measure for preschool children. *Developmental Medicine and Child Neurology*, 52, 47-53.
- Turk, M. (1997). Early development-related conditions. In M.J. Fuhrer (Ed.) *Assessing medical rehabilitation practices: The promise of outcomes research* (pp.367-392). Baltimore, MD: Paul Brooks.
- Unsworth, C., Duckett, S., Duncombe, D., Perry, A., Skeat, J., & Taylor, N. (2004). Validity of the AusTOM scales: A comparison of the AusTOMs and EuroQOL-5D. *Health Quality of Life Outcomes*, 2, 64-75.
- van der Putten, J., Hobart, J.C., Freeman, J.A., & Thompson, A. J. (1999). Measuring change in disability after inpatient rehabilitation: Comparison of the responsiveness of the Barthel Index and the Functional Independence Measure. *Journal of Neurology, Neurosurgery & Psychiatry*, 66(4), 480-484.
- Warr-Leeper, G. (1993) The case for management of language disorders in the schools: Indispensable or incidental? *Communication Exchange*, 4(4), 1-5.
- Warr-Leeper, G. (2001). A review of early intervention programs and effectiveness research for environmentally disadvantaged children. *Journal of Speech-Language Pathology and Audiology*, 25, 89-102.
- Warr-Leeper, G., Wright, N., & Mack, A. (1994). Language disabilities of antisocial boys in residential treatment. *Behaviour Disorders*, 19(3), 159-170.
- Washington, K. (2007). Using the ICF within speech-language pathology: Application to developmental language impairment. *Advances in Speech-Language Pathology*, 9(3), 242-255.
- Washington, K.N., Warr-Leeper G., & Thomas-Stonell, N. (2011). Exploring the outcomes of a novel computer-assisted treatment program targeting expressive grammar deficits in preschoolers with SLI. *Journal of Communication Disorders*, 44, 315-330.
- Yorkston, K., Klasner, E., & Swanson, K. (2001). Communication in context: A qualitative study of the experiences of individuals with multiple sclerosis. *American Journal of Speech-Language Pathology*, 10, 126-137.

Authors' Note

Correspondence concerning this article should be addressed to Bruce Oddson, School of Human Kinetics, Laurentian University, 935 Ramsey Lake Rd, Sudbury, ON, P3E 2C6 Canada. Email: boddson@laurentian.ca

Received date: May 15, 2012

Accepted date: April 3, 2013