An Evaluation of the Responsiveness of the Pre-Kindergarten ASHA NOMS

Évaluation de la sensibilité du Pre-Kindergarten NOMS de l'ASHA

Nancy Thomas-Stonell
Sharon McConney-Ellis
Bruce Oddson
Bernadette Robertson
Peter Rosenbaum

Abstract

The responsiveness of the ASHA Pre-Kindergarten National Outcomes Measure System (Pre-K NOMS) was evaluated with 213 children between the ages of 3 and 5 years. Change scores on the Pre-K NOMS were compared with ratings of change made independently by both the child's parent and treating clinician. Agreement between the parents and clinicians on the question of whether or not communication skills had improved in real life communication situations was used to select a 'consensus' group (N = 169). Parents and clinicians agreed that 96% of the children in this 'consensus' group had made 'real life' changes in their communication skills compared to only 68% identified by the Pre-K NOMS. The Pre-K NOMS failed to detect 29% of the children who made observable 'real life' changes during therapy as judged independently by both treating clinicians and the child's parent. The Guyatt Responsiveness Index (GRI) is a statistical measure of responsiveness. The GRI value suggests that in a treatment efficacy study, the Pre-K NOMS would not be responsive enough to detect meaningful changes or to discriminate between two treatments.

Nancy Thomas-Stonell, B.Sc., D.S.P.
Bloorview Research Institute,
Bloorview Kids Rehab
Graduate Department of Speech-Language Pathology
University of Toronto;
Toronto, Ontario Canada

Sharon McConney-Ellis, M.A.
Bloorview Research Institute
Bloorview Kids Rehab
Toronto, Ontario Canada

Bruce Oddson, PhD
School of Human Kinetics
Laurentian University
Sudbury, Ontario Canada

Bernadette Robertson, LCST
Bloorview Research Institute
Bloorview Kids Rehab
Toronto, Ontario Canada

Peter Rosenbaum, M.D.
CanChild Centre for Childhood Disability Research
McMaster University
Hamilton, Ontario Canada

Abrégé

La sensibilité du Pre-Kindergarten National Outcomes Measure System (Pre-K NOMS – Système national de mesure des résultats pour les enfants de la prématernelle) de l'ASHA a été testée auprès de 213 enfants de 3 à 5 ans. Les scores de changement au test du Pre-K NOMS ont été comparés à l'évaluation de changement faite indépendamment par les parents des enfants et les cliniciens traitants. Pour établir un « groupe de consensus » (N=169), nous avons utilisé les réponses où les parents et les cliniciens s'entendaient à la question qui demandait si les capacités de communication s'amélioraient ou non dans la vie quotidienne. Les parents et les cliniciens convenaient que 96 % des enfants du « groupe de consensus » avaient connu de vrais changements dans leurs capacités de communication. Par contre, le Pre-K NOMS montrait que seulement 68 % de ces enfants avaient connu un changement. Le Pre-K NOMS n'a pas réussi à identifier 29 % des enfants qui ont connu de vrais changements observables durant la thérapie, selon une évaluation indépendante des cliniciens traitants et des parents. L'index de sensibilité Guyatt est une mesure statistique de la sensibilité. Il arrive à des valeurs suggérant que, pour une étude sur l'efficacité du traitement, le Pre-K NOMS ne serait pas assez sensible pour déceler des changements significatifs ou pour établir une différence entres deux traitements.

Key words: outcome measure, evaluation, responsiveness, communication disorders, preschool children

Introduction

Establishing the outcome of speech and language interventions has become an issue of great importance in the profession because speech-language disorders are now recognized as a major health problem in young children (Beitchman, Nair, Clegg, & Patel, 1986; Cleave, 2001; John & Enderby, 2000; Ontario Association of Families of Children with Communication Disorders, 2000). There is an increasing demand from healthcare payers, hospital administrators and clients to document the practical benefits of therapy (Curlee, 1998). Outcome measures are designed to monitor the benefits of treatment in 'real world' conditions (Cleave, 2001). The benefits of outcome measurement include: informed decision-making for clinical decisions, improved program planning, identification of problem areas, and better alignment of goals and objectives with client needs (Forer, 1995).

Outcome measures are simply 'tools'. As such they must have demonstrated clinical utility, reliability, validity and, in the case of change-detecting measures, clinical responsiveness (Law et al., 1999). Outcome measures require responsiveness to client-relevant and clinician-relevant change (Long & Dixon, 1996). Responsiveness is a type of validity that refers to a measure's ability to detect a change over time, whether due to treatment or to natural evolution of function (Norman, Stratford, & Regehr, 1997). An outcome measure should change along with the real changes in health status and be able to discriminate between those clients who improve and those who do not (Deyo & Centor, 1986; Gorelick, Scribano, Stevens, & Schultz, 2003).

The purpose of the present study was to evaluate the responsiveness of the Pre-Kindergarten version (Pre-K NOMS) of the recently developed American Speech-Language-Hearing Association (ASHA) National Outcome Measurement System (NOMS, 2000). Unless the reliability, validity and responsiveness of the NOMS is established, professionals may draw incorrect conclusions from the data collected (Rosenbaum, Russell, Cadman, Gowland, Jarvis, & Hardy, 1990). If a measure has poor responsiveness the researcher and/or clinician may conclude that a treatment or program is ineffective, when in fact it may be the measure that has failed to detect real changes that have occurred (Lipsey, 1983).

There is no clear consensus on the most appropriate way to measure responsiveness (Liang, 1995; May, 1997; Murawski & Miederhoff, 1998). Some researchers argue that responsiveness should be measured using effect size statistics or responsiveness indices (Davidson, 2003; Norman et al., 1997). Others state that two treatments may differ significantly according to statistical analyses without reflecting clinically important change, and recommend that responsiveness be measured using clinically relevant assessments of change (Bain & Dollaghan, 1991; Frytak, 2000; Long & Dixon, 1996). In the absence of a gold standard, May (1997) recommends correlating the scores of the instrument under study with an external measure of change such as client-clinician agreement. Studies examining clinician and parent judgments of preschoolers' speech and language skills have found that parents and clinicians are able to make accurate and reliable judgments of children's skill levels (Glascoe, 1997; Rafaat, Rvachew, & Russell, 1995). This methodology has been successfully used to validate the responsiveness of the Gross Motor Function Measure (GMFM), an internationally accepted evaluative measure for children with cerebral palsy (Russell et al., 1989), and was applied in the current study.

Change can be also be evaluated using statistical measures such as effect size statistics, standardized response means and the Guyatt Responsiveness Index (GRI) (Guyatt, Walter, & Norman, 1987; Murawski & Miederhoff, 1998; Norman et al., 1997). Davidson (2003) compared several responsiveness indices and stated that the GRI was the responsiveness index of choice. The GRI is recommended when samples can be partitioned into two groups: those clients making improvement and those not making improvement according to an external criterion (Corzillius, Fortin, & Stucki, 1999; Tuley, Mulron, & McMahn, 1991). The GRI estimates responsiveness by comparing the average amount of change in an 'improved' group against the variability in the 'stable' group. It accounts for measurement error (i.e. large variability in scores in the stable group), ensuring that instruments with poor reliability are not falsely rated as responsive (Corzillius et al., 1999). Measured change will not be meaningful unless the average amount of change obtained in the improving group is sufficiently larger than the amount of change found in the stable group. The GRI statistic provides an estimate of the most critical aspect of responsiveness – the ability to detect change where change has occurred. It provides a statistical estimate of whether or not a client undergoing treatment is likely to leave the distribution of pre-treatment scores. This definition of treatment efficacy is most applicable in studies such as the one reported here, given that 'normal speech' was not an expected endpoint for most of the children by the end of their treatment block.

Only recently have comprehensive outcome measurement instruments become available in speech-language pathology. Two measures developed for use with a variety of communication disorders are the Therapy Outcome Measures (TOM) developed in England by Enderby (1997) and the American Speech-Language-Hearing Association (ASHA) National Outcome Measurement System (NOMS) (ASHA, 1998-1999). This study focused on the ASHA NOMS as the TOM use terminology and categories that do not transfer well to pediatric clinical practice in Canada. The ASHA NOMS is designed to be an evaluative measure to quantify functional changes in communication skills after speech and language interventions (ASHA, 1998-1999). It is currently being used throughout the United States, and there has been a lot of interest in using the NOMS or adapted versions of the NOMS in Canada (Canadian Association of Speech-Language Pathologists & Audiologists, 2003; Jacoby, Lee, Kummer, Levin, & Creaghead, 2002; Robertson & Devlin, 2002). There are three versions of the NOMS: Adults,

K-6 Schools and Pre-Kindergarten. The Pre-Kindergarten NOMS (Pre-K NOMS) was designed for children between 3 and 5 years of age. It is designed to be used by a treating clinician for any child who has attended a minimum of two speech and language sessions and received some form of intervention (e.g., direct treatment, consultation, home programming). The Pre-K NOMS consists of six Functional Communication Measures (FCMs) which address different clinical areas: Articulation/Intelligibility, Cognitive Orientation, Pragmatics, Spoken Language Comprehension, Spoken Language Production, and Swallowing (NOMS, 2000). The FCMs are 7-point rating scales designed to measure changes in a child's functional communication ability between the beginning and the end of treatment (Mullen, 2003). Functional communication is defined as the 'ability to convey or receive a message, regardless of the mode, to communicate effectively and independently in natural environments' (Goldsmith, 1994).

The FCM scales range from least functional (level 1) to most functional or competent (level 7). A client who achieves an improvement of 1 point on the rating scale is deemed to have made measurable progress or functional communication change (ASHA, 1998-1999); however, there is no published validity information to support this assertion. The FCM levels contain descriptors of communication skills and consider both the intensity of cuing and the environment in which the child is able to communicate (NOMS, 2000). For example, the difference between Level 2 and Level 3 of the Spoken Language Comprehension FCM is the ability of the child to demonstrate understanding of a limited number of common objects, action labels and simple directions in highly structured repetitive daily routines (Level 2) versus novel situations (Level 3). The FCMs are only scored if they specifically relate to the child's treatment program. Clinical judgment is used to rate clients' skills, which are judged according to expected skill levels for the child's chronological age. Clinicians select the FCM level that best reflects the majority of the child's communication skills.

The ASHA NOMS was developed by a committee of content experts, peer reviewed for face validity, then field tested and revised (Dobrzykowski, 1999). Our literature review (November 2004) revealed only one published study using the Pre-K NOMS. This study retrospectively evaluated therapy outcomes using the Pre-K NOMS, but it did not evaluate the responsiveness of the measure (Jacoby et al., 2002). No articles evaluating the reliability, validity or responsiveness of the Pre-K NOMS have been published to date (R. Mullen, personal communication, November 10, 2003).

It is important that the reliability, validity and responsiveness of the NOMS be established prospectively, because information gathered with poor measurement tools may be flawed, and this has serious implications for the analysis and interpretation of data (Frytak, 2000). This concern formed the rationale for the present study. The following specific hypotheses about the responsiveness of the NOMS were explored in this study:

Hypothesis A: Children rated independently by both clinician and caregiver as having made functional change will show change on the Pre-K NOMS, while those rated by both as not making functional change will show no change on the Pre-K NOMS;

Hypothesis B: children with 'mild' communication disorders will show more change on the Pre-K NOMS than children who have 'moderate' and 'severe' communication disorders; and

Hypothesis C: the Guyatt Responsiveness Index (GRI) will be >2, indicating that the Pre-K NOMS is a responsive measure.

Method

Participants

Six organizations in Ontario that provide speech and language services to preschool children collaborated on this study. These organizations were: Bloorview Kids Rehab (Toronto), Children's Hospital of Eastern Ontario-First Words (Ottawa), Ottawa's Children's Treatment Centre, Porcupine Health Unit (Timmins), Pathways Health Centre for Children (formerly Sarnia and District Children's Treatment Centre) and the York Region Preschool Speech and Language Program (Beyond Words). Following ethical approval from all participating centres, families were invited to participate in the study. All families who had children between the ages of 3 and 5 years receiving speech and language therapy were invited to participate. Families were also included if their child turned 3 or 6 years old within the treatment block. Organizations with small speech and language caseloads invited every family whose child met the age criterion to participate. Organizations that served large numbers of children designated specific clinicians who invited every family on their caseload whose child met the age criterion to participate in the study. Written information was made available in both English and French. Although interpreters were available if families did not speak either English or French, this service was not requested.

Responsiveness studies are strengthened when the data are collected on a large, diverse population. The sample must be varied, and include participants who improve with treatment and those who do not, as responsive outcome measures need to be able to distinguish between these two groups. This study enrolled 221 children and their families. Complete data were obtained on 213 children (96%). Eight children were lost to follow-up (four children moved, two clinician data forms were missing due to a job change, and two parent follow-up forms were missing). The children's ages ranged from 2.5 to 5.7 years with a mean age of 4.1 years (SD =.66). The majority of children (84%) were between 4 and 5 years of age. Sixty-eight percent of the children were boys. Approximately 25% of families reported that more than one language was spoken in the home. Treatment was provided in either English (95%) or French (5%).

Medical information was collected from the medical records of the participating organizations. Thirty-eight

percent of study participants had associated medical diagnoses. The most prevalent diagnosis identified was Neuromotor Disorders (e.g., cerebral palsy, muscular dystrophy) at 14%, followed by Developmental Delay (10%). The prevalence of Autism and Related Disorders and Syndromes (i.e., Down, Fetal Alcohol) was 5% each.

Each child's speech and language skills were evaluated according to assessment procedures developed by each agency using a combination of standardized assessment instruments and clinical observations. All children were identified during this assessment as meeting the criteria for admission to speech and language therapy. To ensure that standard terminology was used across sites, clinicians were asked to identify the primary and secondary communication disorder (if any) of each child using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) disorder codes, which were also used by ASHA when collecting demographic information (National Centre for Health Statistics, n.d.). Clinicians reported that 41% of children had Developmental Speech Disorders, meaning either delayed or disordered phonologic development. The next most common communication disorder was Developmental Language Production at 22%, and Developmental Language Comprehension at 16%. A large number of children (*N*=149) also had a secondary communication disorder identified. The most common secondary communication disorder was Developmental Language Production at 53%, followed by Developmental Language Comprehension in 17%.

The mean duration of treatment among the six organizations was 3.3 months, ranging from 2.1 to 6.0 months. To facilitate data collection, a maximum treatment interval of 6 months was established. Nine children (4.2%) continued to receive speech and language treatment beyond the 6-month interval. Treatment frequency ranged from one session every 6 weeks to five sessions a week; however, consistent with current service delivery models in Ontario, most children (85%) received therapy once a week. Eighty-five percent of therapy sessions were between 31 and 60 minutes in length. The mean cumulative length of treatment was 8.7 hours. Individual treatment was provided 61% of the time and group treatment was provided 32% of the time. The remaining 7% of the time, consultation was provided to community staff and/or parents who were provided with speech and language programs. Typically, two goals were addressed in treatment. The most frequently scored FCMs were Articulation/Intelligibility (34%), Spoken Language Production (33%) and Spoken Language Comprehension (17%). The other three FCMs (i.e., Cognitive Orientation, Pragmatics, Swallowing) were all scored for fewer than 10% of the children.

Procedure

Forty-three speech-language pathologists participated in the study. All speech-language pathologists were registered with the College of Speech-Language Pathology and Audiology of Ontario, had a Master's or Doctoral degree in speech-language pathology and had completed a 6-month mentorship program. Participating speech-language pathologists were trained in the use of the Pre-K NOMS. Dr. Robert Mullen, Director of ASHA's National Centre for Treatment Effectiveness in Communication Disorders, provided a copy of ASHA's training manual (Pre-kindergarten Speech-Language Pathology Users Guide) and evaluation procedures to ensure that the research clinicians met ASHA's training standards (NOMS, 2000). A training video was created so that all clinicians could practice scoring the Pre-K NOMS forms. All participating clinicians scored the written scenarios provided by ASHA and achieved >85% on the scoring of these scenarios before beginning the data collection.

According to the ASHA NOMS training manual, the treating clinician completes the Pre-K NOMS scales at the beginning and end of treatment for the FCM areas that correspond to the treatment goals. Skills are rated from least functional (level 1) to most functional or competent (level 7) compared to expected skill levels for their chronological age. In this study, the same clinician was then asked to complete a structured questionnaire rating the child's skills compared to the child's peers. This questionnaire was designed to be comparable to the Pre-K NOMS by rating change in the six communication areas addressed by the Pre-K NOMS (i.e., Articulation/Intelligibility, Cognitive Orientation, Pragmatics, Spoken Language Comprehension, Spoken Language Production, and Swallowing) as well as providing an overall rating of change. At the same time, one or both of the child's parents independently completed the questionnaire to provide an additional judgment of communication changes. The parents were asked to complete the questionnaire as they have the most insight into whether any communication changes resulted in improved 'functionality' for the child in their environments.

At the completion of treatment, the treating clinicians completed the Pre-K NOMS a second time. The child's parents and the clinicians re-rated the child's communication skills compared to their peers using the same structured questionnaire. After completing these ratings, parents and clinicians were provided with the questionnaires they had completed at the start of treatment to remind them how they had initially rated the child's communication skills in each area. This methodology is recommended to improve the reliability of judgments of change and has been used effectively by Russell et al. (1989) to evaluate the responsiveness of the Gross Motor Function Measure (following Guyatt, Berman, Townsend, & Taylor, 1985 and as recommended by Norman et al., 1997). This information was used to guide parent and clinician judgments for Part 2 of the questionnaire, which was only completed at the end of treatment. Parents and clinicians judged (i) how much change had occurred in the six communication areas addressed by the Pre-K NOMS, (ii) whether communication changes had resulted in 'real life' communication changes, and (iii) how much change they had noted in 'overall' communication skills.

Hypothesis A

Clinicians and parents independently judged whether changes in communication skills had resulted in 'a difference in the child's ability to communicate independently in real life situations'. This 'yes' or 'no' question was based on ASHA's definition of functional communication change (Goldsmith, 1994). When parents and clinicians judged that 'real life' change had occurred, they were asked to complete open-ended questions describing the changes that had been observed and the impact of these changes on their child at home and in the community. This procedure helped ensure that parents and clinicians were reporting treatment effects that had a broad impact on the client's ability to function in their world (Bain & Dollaghan, 1991). Parent and clinician agreement on this question was used to divide the sample into 'change' or 'no-change' groups, establishing a 'consensus' group against which the Pre-K NOMS was compared. Kappa was used to examine the agreement between this classification and change detected by the Pre-K NOMS (> 1 FCM level). For the Pre-K NOMS, the criterion for measuring functional change was set as broadly as possible. The Pre-K NOMS were classified as having measured functional change if there was a change of > 1 FCM level in any of the FCM scales scored.

Hypothesis B

At the start of treatment, clinicians gave an overall rating of the child's communication skills compared to their peers. Responses to this question were used to group the children into three equal groups – 'mild', 'moderate' and 'severe' communication impairments – for purposes of this analysis only. At the end of treatment, clinicians rated how much change overall had occurred using an 8-point rating scale (i.e., from 0 = "got worse" to 7 = "excellent change"). A chi-square analysis was used to determine if any of these groups had made more change than expected, defined by a rating of 4 or higher. If one of the mild, moderate or severe groups showed more change according to the chi-square analysis, it was hypothesized that the Pre-K NOMS would also show significantly more change for this group.

Hypothesis C

The Guyatt Response Index (GRI) was used to measure responsiveness statistically (Frytak, 2000). The GRI is the mean improvement in the improving group divided by the standard deviation in the non-improving group. The sample was divided into 3 groups: (i) those that improved according to both parent and clinician consensus judgments, (ii) those that did not improve according to both parent and clinician consensus judgments, and (iii) those children where parents and clinicians disagreed about change. As the non-improving 'consensus' sample was only 4% (N = 7), the GRI denominator also included data from the 44 children whose parents and clinicians disagreed about improvement. These data were required to provide a reliable estimate of variance for calculating the GRI. Ratings of 'real life' change on the 8-point rating scale were used to calculate the GRI values for parents and clinicians. Ratings on the FCM scales were used to calculate the GRI for the Pre-K NOMS. While there is no generally accepted benchmark of good responsiveness in the literature, a GRI value near 2 indicates that the improving group has scores noticeably different from those in the non-improving group. This value indicates that a measure has the ability to detect differences between an 'improved' and 'not improved' group.

Results

Clinicians indicated that 82% of the 213 children had made 'real life' improvements following treatment. Parents rated 90% of the children as having demonstrated 'real life' improvements. The Pre-K NOMS indicated that only 58% of the children made 'functional communication changes' (i.e. 'real life change') on any of the FCMs (i.e. > 1 FCM level). Agreement between parents' ratings of 'real life' change and the Pre-K NOMS was 59% (see Table 1). Agreement between clinicians' own Pre-K NOMS ratings and their independent ratings of 'real life' change was 69% (see Table 2). Agreement between parents and clinicians about the presence of 'real life' change was 79% (see Table 3).

Table 1

Percent Agreement between Parent Ratings of Change and Pre-K NOMS

	Parent Ratings		
Pre-K NOMS Ratings	No Change	Change	Totals
No Change	5.6%	36.7%	42.3%
Change	4.2%	53.5%	57.7%
Totals	9.8%	90.2%	100%

Note: Overall agreement between parent rating and the Pre-K NOMS ratings is 59.1%

Table 2

Percent Agreement between Clinician Ratings of Change and Pre-K NOMS

	Clinician Ratings		
Pre-K NOMS Ratings	No Change	Change	Totals
No Change	14.9%	27.4%	42.3%
Change	3.3%	54.4%	57.7%
Totals	18.1%	81.9%	100%

Note: Overall agreement between clinician ratings and the Pre-K NOMS ratings is 69.4%

Table 3

Percent Agreement between Parent and Clinician Ratings of Change

	Clinician Ratings		
Parent Ratings	No Change	Change	Totals
No Change	5.5%	6.3%	11.8%
Change	14.9%	73.3%	88.2%
Totals	20.4%	79.6%	100%

Note: Overall agreement between parent and clinician ratings of change is 78.7%

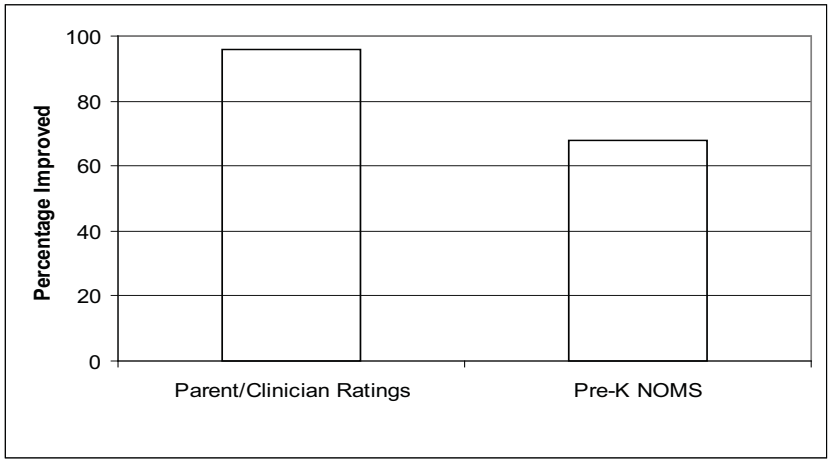

Figure 1. Percent improvement for the consensus group (N = 169) according to parent/clinician ratings and the Pre-K NOMS

Hypothesis A

In 169 of the 213 children, parents and clinicians agreed that change had or had not occurred. This group (N = 169) is referred to as the 'consensus' group. Ninety-six percent of the children in this group were judged to have made 'real life' changes, while both parents and clinicians agreed that 4% of these children had made no change. The Pre-K NOMS showed functional improvements for 64% of the children (see Figure 1). There was significant but very modest agreement between the 'consensus' group and the Pre-K NOMS regarding which children were and were not improving (K = .142, p = .001); however, the Pre-K NOMS missed 29% of the children that demonstrated 'real life' communication changes as judged by both parents and clinicians. In all cases where there was a disagreement between the parent/clinician judgments and the Pre-K NOMS, parents and clinicians agreed that the child had made 'real life' improvement, but the Pre-K NOMS did not measure any change.

For the three most frequently scored FCMs (i.e., Articulation/Intelligibility, Spoken Language Comprehension and Spoken Language Production), parents and clinicians noted 'real life' improvements in 96% of the children while the Pre-K NOMS showed improvements in only 50% of these children. There was no significant difference in responsiveness for the three FCMs examined. In addition, neither chronological age nor medical diagnosis affected the responsiveness of the Pre-K NOMS.

Hypothesis B

To determine whether the Pre-K NOMS simply represents a stricter criterion for determining functional change, agreement was examined for a subset of children rated as having made more than expected change (overall ratings of change > 4) according to the judgments made by parents (n = 96) and clinicians (n = 51). There was a significant relationship between clinicians' severity ratings at the start of treatment and clinicians' ratings of change (X^2 (12) = 32.1, p < .01) (see Figure 2). Children who were classified as having 'moderate' communication disorders showed significantly more treatment change than children with 'mild' or 'severe' communication disorders in two treatment areas: Spoken Language Production (Odds Ratio (OR) = 3.31, p = .019) and Spoken Language Comprehension (OR = 3.05, p= .003) according to the parent and clinician ratings. The Pre-K NOMS detected no difference in the amount of change for this moderate group (OR = .79, p = .06) compared to the mild and severe groups on either FCM. Agreement between ratings and Pre-K NOMS declined for this subgroup of children. This does not support the view that the primary difference between the Pre-K NOMS and consensus judgments is a difference in criterion level for measuring change.

Hypothesis C

The GRI was calculated to evaluate statistically the responsiveness of the Pre-K NOMS. Both the clinicians' (GRI = 3.1) and parents' ratings of change (GRI = 3.3) were greater than 2, indicating that they are highly responsive. The GRI for the Pre-K NOMS was 0.9, indicating poor responsiveness. This GRI value indicates that the Pre-K NOMS would not be able to detect differences between an improved and non-improved group even when improvement falls into the statistically moderate range (i.e., d = .5). When compared to the variability in test scores, the amount of change measured by the Pre-K NOMS is not sufficiently large to reliably detect treatment change.

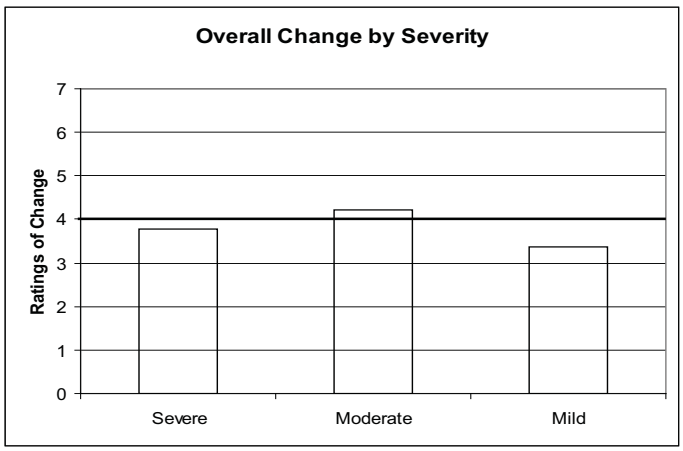

Figure 2. Clinicians' ratings of change for children rated as having mild, moderate and severe communication impairments. A rating of 4 or higher indicates greater than expected change.

Discussion

It is important that an outcome measure be able to detect changes that are relevant and important to both clinicians and families. To help ensure that observed changes reflected 'real life' improvements, parents and clinicians were asked to describe the impact of observed changes on the children. Treatment effects that have a broad impact on the client's ability to function are one criterion of 'clinically significant' change as defined by Bain & Dollaghan (1991). Parents and clinicians described changes such as improved intelligibility for familiar and/or unfamiliar listeners, improved ability to express wants and needs, less frustration when communicating and better socialization with peers. The finding that the Pre-K NOMS missed 29% of the children that demonstrated such 'real life' changes according to parents and clinicians suggests that the Pre-K NOMS is not yet a responsive tool.

Additional information was gathered from clinicians to provide insight into why clinicians rated 'real life' change on the questionnaires but did not reflect this change on the FCM scores. The primary reason for the discrepancies was that changes had occurred in areas not measured by the Pre-K NOMS! Clinicians' commented on improvements in attention, confidence, socialization with peers and increased participation in school/daycare. They noted improvements in the child's ability to repair communication breakdowns, resulting in improved intelligibility and independence. These comments indicate that the children had made observable changes in their ability to use communication skills within their community. These changes reflect a positive 'real life' change at the level of participation according to the World Health Organization's International Classification of Functioning, Health and Disability (WHO, 2003).

The Pre-K NOMS FCM scales focus on changes in communication skills at the ICIDH-2 levels of Body Functions and Activity. They do not measure changes at the Participation level. In addition, only the specific FCMs that correspond to treatment goals are scored. Although the Pre-K NOMS manual states "the FCMs are not intended to reflect the entire evaluation or to describe all aspects of a child's communication abilities", the finding that the Pre-K NOMS missed changes in 29% of the children suggests that they may be too focused on impairment level changes. The Pre-K NOMS need to focus more on changes at the levels of activity and participation.

At times, parents and clinicians also observed changes in areas that directly related to the FCMs (e.g., improved speech intelligibility; using longer, more complex sentences; improved ability to follow directions), that they were not able to reflect on the relevant FCM scales. Some of the FCM descriptors require changes in two or three skill areas before moving to the next FCM level. For some children, changes in one skill area were large enough to be judged by the clinician and parent as important 'real life' change but this change did not result in an increased FCM score. For example, one child learned to produce consonant-vowel-consonant combinations which increased intelligibility with unfamiliar listeners. Both the parents and the clinician felt that this child had made 'real life', clinically important gains. The clinician initially scored this child as a Level 4 on the Articulation/Intelligibility scale but did not change the score to Level 5 at discharge, despite improved intelligibility with unfamiliar listeners, because intelligibility for familiar listeners remained the same, at 70-75%, below the consistently intelligible (i.e. 80-100%) level as defined by the Pre-K NOMS. Therefore, the clinician rated the child as making 'real life' change on the questionnaire, but these was no corresponding change on the Pre-K NOMS.

The Pre-K NOMS may also miss functional changes because ratings are based on a comparison between the child's current skills and the norm-based skills expected for a child's chronological age (NOMS, 2000). A child may make 'real life' changes in communication skills but remain at the same level on the Pre-K NOMS because the improvement is less than the improvements achieved by their typically developing peers over the treatment period. Clinicians noted that this was especially true for children with severe communication impairments. The problem of making norm-referenced comparisons for children with functional limitations has been addressed elsewhere (Rosenbaum et al., 1990).

A few of the discrepancies were also caused by variations in the way clinicians scored the FCMs when the child demonstrated changes in skill levels within treatment sessions (i.e., changes in activity/capacity), but these skills had not transferred to other environments (i.e., participation/performance). There can be differences between 'capacity' and 'performance' in specific environments. Some clinicians increased the FCM score but stated that there was no 'real life' improvement on the questionnaire because the child was not able to apply these new skills outside of the

treatment session. Other clinicians did the opposite. To ensure that the measure is scored reliably, ASHA needs to clarify for clinicians how they should score the FCMs in this circumstance. As Goldsmith (1994) has emphasized that "functional communication is the ability …to communicate effectively and independently in natural environments", the current authors feel that clinicians' ratings should reflect children's communication performance in their home and community. Thus, the FCM would not be scored as improved if the skill had not yet generalized.

One of the most important attributes of an outcome measure is the ability to discriminate between those clients who improve and those who do not (Deyo & Centor, 1986). A GRI of only 0.9 for Pre-K NOMS suggests that they would not be able to distinguish between an improved and non-improved group in studies that compare two treatment modalities or those where treatment is of limited duration. In such studies, using a measure that is not responsive will almost always lead to the conclusion that no change has occurred, whatever the true answer might be.

The ideal way to evaluate the responsiveness of an outcome measure is to compare it against a treatment of known efficacy or another 'gold standard' measure (May, 1997). Unfortunately, there are no speech and language outcome measures for preschool children with proven reliability, validity and responsiveness that can be used as a gold standard. In other fields, independent agreement between parents and clinicians about the presence or absence of change has been used as an external measure of change. In this study, the treating clinician completed both the ASHA Pre-K NOMS form and the questionnaire. Neither clinicians nor parents were blind to treatment and may, therefore, be biased towards observing change. It is important to note, however, that many times clinicians and parents were able to describe relevant 'real world' changes outside of the domains measured by the Pre-K NOMS. This does not mean that the Pre-K NOMS is an invalid or inaccurate representation of the therapeutic process within the FCM scored, but suggests that it may not be suitable for measuring the 'real life' progress that clinicians and parents agree to be important.

Conclusion

Outcome measures must evaluate the entire spectrum of skills from body structures/body functions to activity to life participation in order to evaluate the full impact of intervention on a child's life. Responsive outcome measures are needed to identify reliably those children who benefit from speech and language therapy. The Pre-K NOMS was not a responsive outcome measure for this population. It underestimated the number of children who made important 'real life' changes observed by both clinicians and parents during therapy. Similarly, the GRI value suggests that the Pre-K NOMS would not be responsive enough to discriminate between an 'improved' and a 'non-improved' group or between two treatment groups experiencing different amounts of change. It is possible that the responsiveness of this measure could be improved by clarifying the scoring of FCMs when new skills are consistently observed within treatment sessions, but have not transferred to other settings, and by scoring all of the FCM scales for each child rather than only the scales corresponding to specific treatment goals. This study suggests that the Pre-K NOMS may focus too much on changes at the level of impairment and not enough on changes at the levels of activity and participation. Changes at the levels of activity and participation, both within and outside of therapy sessions, need to be considered when measuring progress. The responsibility of finding reliable ways of measuring 'real world' change lies with all clinicians and researchers.

References

American Speech-Language-Hearing Association National Data Report-Rehabilitation Settings (1998-1999). Rockville, MD: American Speech-Language-Hearing Association.

Bain, B., & Dollaghan, C. (1991). Treatment efficacy: The notion of clinically significant change. *Language, Speech and Hearing Services in the Schools*, *22*, 264-270.

Beitchman, J. H., Nair, R., Clegg, M., & Patel, P. G. (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.

Canadian Association of Speech-Language Pathologists and Audiologists (2003). *Manitoba Speech-Language Pathology Outcome Measure Report – CASLPA Committee Report*. Retrieved November 10, 2004, from http://www.caslpa.ca/english/resources/manitoba_outcomemeasures_report.asp

Cleave, P. L. (2001). Design issues in treatment efficacy research for child language intervention: A review of the literature. *Journal of Speech-Language Pathology and Audiology, 25*(1), 24-34.

Corzillius, M., Fortin, P., & Stucki, G. (1999). Responsiveness and sensitivity to change of SLE disease activity measures. *Lupus*, *8*, 655-659.

Curlee, R. F. (1998). Foreword. In C. A. Frattali (Ed.), Assessing functional outcomes in neurogenic populations. *Seminars in Speech and Language*, *19*(3), 205.

Davidson, M. (2003). *Measuring activity limitation in low back pain: A comparison of five questionnaires*. La Trobe University Web site: http://www.lib.latrobe.edu.au/thesis/public/adt-LTU20051220.175347/index.html

Deyo, R. A., & Centor, R. M. (1986). Assessing the responsiveness of functional scales to clinical change: An analogy to diagnostic test performance. *Journal of Chronic Disease*, *39*, 897-906.

Dobrzykowski, E. A. (1999). Functional communication measures. *Journal of Rehabilitation Outcomes Measurement. 3*(3), 57-58.

Enderby, P. (1997). Therapy outcome measures: Speech-language pathology user's manual. London, UK: Singular Publishing Group.

Forer, S. (1995, June/July). Outcomes evaluation in subacute care. Rehabilitation Management, 138-140.

Frytak, J. (2000). Measurement. *Journal of Rehabilitation Outcomes Measurement*, *4*, 15-31.

Glascoe, F. (1997). Parents' concerns about children's development: Prescreening technique or screening test? *Pediatrics*, *99*(4), 522-528.

Goldsmith, T. (1994). Clinical documentation in managed health care environment. American Speech-Language-Hearing Association, *Managing Managed Care*, *8*, 49-56.

Gorelick, M. H., Scribano, P. V., Stevens, M. W., & Schultz, T. R. (2003). Construct validity and responsiveness of the Child Health Questionnaire in children with acute asthma. *Annals of Allergy, Asthma & Immunology*, *90*, 622-628.

Guyatt, G. H., Berman, L. B., Townsend M., & Taylor, D. W. (1985). Should study subjects see their previous responses. *Journal of Chronic Disability, 38*(12), 1003-1007.

Guyatt, G.H., Walter, S., & Norman, G. (1987). Measuring change over time: Assessing the usefulness of evaluative instruments. *Journal of Chronic Disability. 40*(2), 171-178.

Jacoby, G. P., Lee, L., Kummer, A. W., Levin, L., & Creaghead, N. (2002). The number of individual treatment units necessary to facilitate functional communication improvements in speech and language of young children. *American Journal of Speech-Language Pathology, 11*, 370-380.

John, A., & Enderby, P. (2000). Reliability of speech and language therapist using outcome measures. *International Journal of Language & Communication Disorders*, *35*(2), 287-302.

Law, M., King, G., Russell, D., MacKinnon, E., Hurley, P., & Murphy, C. (1999). Measuring outcomes in children's rehabilitation: A decision protocol. *Archives of Physical Medicine and Rehabilitation*, *80*, 626-636.

Liang, M. H. (1995). Evaluating measurement responsiveness. *Journal of Rheumatology, 22*, 1191-1192.

Lipsey, M. (1983). A scheme for assessing measurement sensitivity in program evaluation and other applied research. *Psychological Bulletin, 94*(1), 152-165.

Long, A. F., & Dixon, P. (1996). Monitoring outcomes in routine clinical practice: Defining appropriate measurement criteria. *Journal of Evaluation in Clinical Practice, 2*, 71-78.

May, L. (1997). The challenge of measuring change: Responsiveness of outcome measurements. *Canadian Journal of Rehabilitation, 10*, 15-24.

Mullen, R. (2003, November). National Outcomes Measurement System (NOMS). Paper presented at 13th Annual Research Symposium Outcomes Research and Evidence-Based Practice, Chicago, IL

Murawski, M. M., & Miederhoff, P. A. (1998). On the generalizability of statistical expressions of health related quality of life instrument responsiveness: A data synthesis. *Quality of Life Research, 7*, 11-22.

National Centre for Health Statistics (n.d.). International classification of disease, ninth revision, clinical modification. Retrieved June 20, 2006 from http://www.cdc.gov/nchs/about/otheract/icd9/abticd9.htm.

National Outcomes Measurement System (2000). Pre-Kindergarten speech-language pathology users guide. Rockville, MD: American Speech-Language-Hearing Association.

Norman, G. R., Stratford, P., & Regehr, G. (1997). Methodological problems in the retrospective computation of responsiveness to change: The lesson of Cronbach. *Journal of Clinical Epidemiology, 50*, 869-879

Ontario Association of Families of Children with Communication Disorders (2000). General information on communication: Did you know that. Retrieved April 27, 2004 from http://www.oafccd.com/factshee/fact36.htm

Rafaat, S.K., Rvachew, S., & Russell, R.S.C. (1995). Reliability of clinician judgements of severity of phonological impairment. *American Journal of Speech-Language Pathology 4*, 39-46.

Robertson, M., & Devlin, D. (2002, April). Manitoba Speech-Language Outcomes Measure. Paper presented at the Canadian Association of Speech-Language Pathologists and Audiologists 2002 Annual Conference, Victoria, BC.

Rosenbaum, P.L., Russell, D. J., Cadman, D. T., Gowland, C., Jarvis, S., & Hardy, S. (1990). Issues in measuring change in motor function in children with cerebral palsy: A special communication. *Physical Therapy, 70*, 125-131.

Russell, D. J., Rosenbaum, P., Cadman, D. T., Gowland, C., Hardy, S., & Jarvis, S. (1989). The Gross Motor Function Measure: A means to evaluate the effects of physical therapy. *Developmental Medicine and Child Neurology, 31*, 341-352.

Tuley, M. R., Mulron, C. D., & McMahn, C. A. (1991). Estimating and testing an index of responsiveness and the relationship of the index to power. *Journal of Clinical Epidemiology, 44*(4-5), 417-421.

World Health Organization (n.d.). International Classification of Functioning Disability and Health. Retrieved December 16, 2003, from http://www3.who.int/icf/icftemplate.cfm.

Author Note

The authors wish to acknowledge the financial support of The Hospital for Sick Children Foundation and the Bloorview Children's Hospital Foundation. P. Rosenbaum holds a Canada Research Chair in Childhood Disability from the Canadian Institutes of Health Research. The authors express sincere appreciation to the families who participated in this research project. We especially wish to acknowledge the Research Coordinators and Speech-Language Pathology clinical staff at the: Children's Hospital of Eastern Ontario – First Words Program, Ottawa Children's Treatment Centre, Pathways Health Centre for Children (formerly Sarnia and District Children's Treatment Centre), Porcupine Health Unit, York Region Preschool Speech and Language Program (Beyond Words), and Bloorview Kids Rehab for their clinical expertise and assistance with data collection.

All correspondence should be addressed to Professor N. Thomas-Stonell at the Bloorview Kids Rehab, 150 Kilgour Rd., Toronto, Ontario, Canada, M4G 1R8.

Received: May 24, 2005

Accepted: July 20, 2006