Gestural & Oral Communication and Symbolization –
validating a tool for analysing outcome of early
intervention

Per-Anders Bringfelt

Abstract. Parents to pre-schoolers with severe developmental language
delays usually report elevated communication by their children after
participating in a parental communication course. Valid, reliable and
convenient observational outcome measures are warranted. Video
recordings of three children’s communication were collected at Pre-
and Post-intervention, and at a Follow-up. These were analysed
independently by the author and a blinded clinician focusing on the
child’s Gestural and Oral Communication and Symbolization (GOCS)
ability, an instrument developed for the current study. In addition,
parents were interviewed with the Vineland Adaptive Behaviour
Scales (VABS) and completed the Swedish Communicative
Development Inventories questionnaires (SECDI). Strong significant
correlations were found between symbolic GOCS data and
corresponding VABS and SECDI data. Follow-up interviews and
additional parent information largely supported GOCS data. Inter-
observer intra-class correlations were excellent. Reliability and
methodological issues, as well as further development and validation
of GOCS are discussed.

Keywords: Outcome assessment, child communication, language,
gestures, speech.
Developmental language delay (DLD) is one of the most common child developmental problems with a prevalence of 4-6% (Tomblin, Smith, & Zhang, 1997). Miniscalco (2007) showed that a large portion of children with language problems at 2.5 years of age retain their problems and/or have concomitant developmental problems at school entrance. Most often these children have delays/impaired in more than one area of speech, receptive/expressive language and communication. In the following text, such children are referred to as ‘children with significant/severe DLD’.

Literature repeatedly demonstrates the importance of parental responsiveness for children’s language and communicative development (Hudson, Levickis, Down, Nicholls, & Wake, 2015; Masur, Flynn, & Eichorst, 2005; Tamis-LeMonda, Bornstein, & Baumwell, 2001). When a child has communicative difficulties unproductive parent-child interaction patterns may emerge early and parents meet great challenges in maintaining responsive style (Conti-Ramsden, 1990; Slonims & McConachie, 2006). A wide range of parent-mediated interventions have therefore been developed. Theoretically based on an interactionist perspective on language development (Tomasello & Farrar, 1986; Vygotsky, 1978) these programs assume that parent-child interaction patterns may be enhanced, with positive effects on children’s communicative and linguistic development, especially at early stages (Girolametto, Weitzman, McCauley, & Fey, 2006; Green et al., 2010; McConachie, Randle, Hammal, & Le Couteur, 2005; Roberts & Kaiser, 2012; Siller, Hutman, & Sigman, 2013).

Within the Swedish early intervention project AKKtiv (Ferm, Andersson, Broberg, Liljegren, & Thunberg, 2011) the parent training course ComAlong was developed. It targets parents of pre-school children with severe communication difficulties due to varying conditions, e.g. Autism Spectrum Disorders (ASD) and Intellectual Disability (ID), all of which often benefit from Augmentative or Alternative Communication (AAC).

A ‘DLD version’ of ComAlong was subsequently gradually developed and implemented between 2009 and 2012, and targets parents to toddlers with significant communication difficulties, but no other confirmed diagnoses. The courses are provided in group format with a maximum of eight participating families, and take place in clinical settings. A course typically runs over eight weeks, with six 2.5 hours’ sessions. The general scope is mainly retained and may be summarized in five main areas: (1) Responsive communication, (2) Communication and communicative development, (3) Play, (4) Communication strategies, and, (5) AAC. With parental responsiveness as a red thread, the other areas are covered in lectures, video examples, discussions and hands-on practice. Home assignments are also provided which parents are encouraged to video-record. The assignments are followed up with careful feed-back on the parent’s and child’s communication, as displayed in shared videos or as related verbally. In short, parents are guided to elicit communication from their children and to expand on their children’s existing communication. This is expected to increase the amount of child communication and to extend symbolic communication if present, e.g. beyond a few spoken words to short sentences. For more details on the course content, see Ferm et al. (2011).

According to compilations of course evaluation surveys, a clear majority of parents – in both versions of the ComAlong course – report that their own communication skills have improved during courses. Parents often describe they have become more
responsive and less directive in their communication with their children. A group study of parent-child dyads where children had ASD, ID, and/or multiple disabilities has shown some increase in parental responsiveness in video-recorded parent-child interaction at Post- intervention, compared to Pre-intervention, and compared to a small non-participant group (Broberg, Ferm, & Thunberg, 2012). Collection of such video data on DLD-dyads has just begun.

Also children’s communicative and linguistic gains has shown to be higher compared to measures of other developmental domains in one study (Callenberg & Ganebratt, 2011). These findings however rely solely on parental report by the Vineland Adaptive Behaviour Scales (VABS) interviews (Sparrow, 2011) and the Swedish Early Communicative Development Inventories (SECDI), (Berglund & Eriksson, 2000) and might partly stem from increased parental awareness of communication during intervention. To evaluate treatment effects complementary observational data is warranted.

Video-recordings are commonly used in analysis of communication in young children, for example in established methods like the Communication and Symbolic Behaviour Scales Developmental Profile (CSBS-DP) by Wetherby and Prizant (2008) and Early Social Communication Scales (ESCS) by Mundy et al. (2003). The advantages of video-recordings for research are obvious. Communication is multimodal. Repeated observations facilitate accurate notations and inter/intra-observer reliability checking.

A method for video observation would preferably be time-effective and require limited expertise and training, especially if larger numbers of children are to be observed, and/or if research funding is limited. It should also allow for repeated measuring without ceiling effects. When reviewing established methods, problems were recognized according to these criteria. Conversation Analysis (Psathas, 1995) for example, rely on detailed transcription and complex analysis, which indeed is time-consuming. Other methods, like the abovementioned CSBS-DP and ESCS, use pre-defined codes to capture different aspects of communication. These two instruments and other current methods for observation spring from ASD research and use scores related to autism and/or infants. They also focus on making distinctions between ‘regulative’ and ‘social’ communication, since predominant use of regulative interaction may be typical in ASD. Thus, the risk of ceiling effects with these instruments was considered significant when including DLD populations. Moreover, it was presumed to be troublesome and time consuming to code distinctions between child initiatives and responses, since young children with severe DLD most often have concomitant speech difficulties (Kot & Law, 1995). Examples of only moderate reliability has been seen in a randomized controlled trial on a ASD population (Green et al., 2010).

The most important question is: What change, if any, would be detectable in children with severe DLD, whose parents take part in a relatively brief intervention like ComAlong? In course evaluations, parents – both to children with DLD and to those with ASD and/or ID – have often reported their children having become more social and/or interested/willing with respect to communication. With parents being the main stakeholders, such change in children has strong social validity as an outcome measure. A key question is how such ‘interest’ may be observed and measured.
A straightforward way to operationalize ‘interest’ in intentional communication might be to count easily observed communicative actions. Indeed, such descriptions have commonly been provided by parents in ComAlong courses: “Tries to talk” or “Points more”. Such communicative behaviours may be divided in two main form categories: *Oral* or *Gestural*. Use of gestures in emerging language predicts language growth (Ellis & Thal, 2008; Rowe & Goldin-Meadow, 2009).

Also measures of linguistic growth are desirable, since most of the targeted children have ability or potential to produce at least some orally mediated words, or other symbol communication (e.g. manual signs). It appeared feasible to count oral contributions from children containing at least one comprehensible word, and gestural contributions containing at least one comprehensible symbolic token (e.g. a manual sign, or a word/symbol mediated by another AAC method). Lexical diversity is another measure of language growth, and ‘Number of different words’ (NDW) have been used as such with similar populations (Roberts & Kaiser, 2012). NDW is however confined to orally mediated language. For children with extremely limited speech skills, their diversity of gesturally mediated symbols was considered useful as a complement or as a first-choice of expressive language (i.e. iconic or emblematic gestures, nodding or shaking head, manual signs, words/symbols mediated by pointing at communication boards or manipulating other augmentative or alternative communication devices). Thus, the following two complementary measures were adopted: ‘NDW Gestural’, and ‘NDW Oral’.

‘Eye Gaze’ was considered as having face validity as a potential measure of interest in communication, being a major element in most interpersonal interactions. In research on child development Joint attention (coordinated eye gaze-pointing/showing) is widely studied, since it is enhancing the development of intentional communication (Bretherton & Bates, 1979). Joint attention is however a more complex behaviour. For example, only moderate reliability regarding shared attention was found in the previously mentioned study by Green et al (2010). A major rationale for coding Eye Gaze only, was that it may be more easily observed in clinical and natural settings.

The aim of the present work was to develop a feasible method for observation of children’s interest in communication and expressive linguistic skills. This method will hereafter be referred to as GOCS (acronym for Gestural and Oral Communication and Symbolization). Parent reported data was used as benchmarks for preliminary validation. Specific research questions were:

1) May GOCS provide feasible and reliable outcome measures of communication and expressive language, in two to four-year-olds with severe DLD and/or concomitant difficulties, when parents take part in a ComAlong course?

2) How does GOCS video data relate to parent reported data of child development in conjunction with the course, as measured by VABS, SECDI and parents’ own descriptions?

**Methods**

The study was approved by the regional ethical review board (Dnr: 608-06). Participants gave written consent, after having been thoroughly informed about the purpose of the study and about their right to withdraw from participation at any time.
Decisions about study participation did not influence the access to services. All identities were kept confidential.

Descriptive mixed methods case studies of three children with severe DLD were carried out. Data of child communication was collected within a Pre-test Post-test design, in conjunction with a two months’ parent mediated intervention (the ComAlong course). Follow-up assessments were made four months later. Comparisons of video data and parent-reported data were then carried out regarding children’s communication.

Participants
Participants were consecutively recruited during three months, at regular visits at the SLP clinic of a major Swedish university hospital. Three families gave informed consent, with one parent from each participating in the course. The children were all boys. Child characteristics are summarized below (table 1).

Table 1

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>“Ben”</th>
<th>“Gus”</th>
<th>“Nic”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, months</td>
<td>42</td>
<td>33</td>
<td>42</td>
</tr>
<tr>
<td>Heredity</td>
<td>-</td>
<td>Stuttering; late talking; persisting spelling difficulties</td>
<td>-</td>
</tr>
<tr>
<td>Language delay</td>
<td>Mixed receptive-expressive</td>
<td>Mixed receptive-expressive</td>
<td>Mixed receptive-expressive</td>
</tr>
<tr>
<td>Hearing</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal; History of recurrent OME¹</td>
</tr>
<tr>
<td>Speech</td>
<td>Deviant²</td>
<td>Delayed³</td>
<td>Deviant⁴</td>
</tr>
<tr>
<td>Parent concerns</td>
<td>Unintelligible speech; Poor social skills</td>
<td>No speech; No interest in verbal imitation; Selective hearing. Playing alone at pre-school</td>
<td>Unintelligible speech; Syndrome, not otherwise specified; Cleft soft palate, closed at 8 months</td>
</tr>
</tbody>
</table>

¹ Otitis media with effusion, ² Frequent use of word-initial glottal stops, ³ Mainly yes/no, deictic and feedback morphemes, ⁴ Plosives and fricatives absent. Suspected developmental apraxia of speech.

The children were raised in monolingual middle class families, where the participating parent had more than 12 years of education. By checking case histories, and through an extra clinical visit conducted by the author, all children were confirmed to be eligible. They had markedly poor expressive skills, with no clinical or parent reported evidence of utterances containing more than one morpheme for the 2-year-old “Gus”, or two morphemes by the 3-year-olds “Ben” and “Nic”. All children had mixed language delays, delineated by a percentile score < 5 on the receptive sub-scale of Reynell Developmental Language Scales (Edwards et al., 1997). All had at least one additional
predictor for long term communication or language problems, e.g. few spontaneous
imitations, poor social skills and family history of persisting speech/language/literacy
difficulties, entailing high priority for intervention (Paul, 2007). Two had been assessed
by psychologists with recommendations for follow up or referral for
neurodevelopmental assessment. One of the two had a suspected apraxia of speech and
motor problems.

Intervention
Three parents (one father, two mothers) took part, along with non-study-parents to four
other children, in a ComAlong course (DLD version). The course content has been
outlined in the introduction.

Data collection
Data were collected at three time points; pre- and post-intervention and at a follow up
four months later, with a few exceptions. For an overview of data collection, see table 2.

Video recordings of play session
At all time-points, approximately 10 minutes of semi-structured play between child-
clinician (the author) was video-recorded at the SLP clinic. The rationale for sampling
communication and language from clinician-child interaction, rather than from parent-
child dyads, was that an experienced clinician was supposed to more likely elicit
optimal communication and linguistic performance during a brief encounter. If
promising, the video observation method would then be transferred to parent-child-
interaction recordings in naturalistic settings. The parent was present throughout the
visit to make the child feel comfortable. At the end of each session, the parent was
asked if the child’s communication was representative for his/her current abilities.
Further video-data procedures are described in the Analyses section.

Parent reported data
Parent reported data was collected at three time points; pre- and post-intervention and at
follow-up. The Swedish Early Communicative Development Inventories (SECDI-WS,
i.e. Words and Sentences version) were handed over to the parents at the end of the
clinician-child play sessions, to be filled out at home and capture child vocabularies
(maximum score: 711). Within a week the author completed VABS parent interviews,
with the main purpose to tap the children’s current level of adaptive skills and
communicative abilities. Specifically, in VABS ‘Expressive Communication’ subscale,
the parents were asked about a range of developmentally related expressive
communicative behaviours and how often and/or how independently these were carried
out. Each item was scored with either 0, 1, 2 rendering a maximum raw score of 108.
Additionally, short written statements were collected about the currently most difficult
aspects of parent-child communication.

At post-intervention only, perceived change regarding children’s communication were
collected from course evaluations.

At follow-up only, an additional semi-structured short interview was undertaken, in
which parents were asked if ‘any important change’ in their children’s communication
had taken place ‘during the last 12 months’. This wording was used in order not to
impose any change due to the intervention. If any change was reported, the parent was asked to describe it and when it occurred.

Table 2

Methods, measures and time points for data collection

<table>
<thead>
<tr>
<th>Parent reported data of child’s communication</th>
<th>Video data of child’s communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-intervention, Post-intervention and Follow-up:</td>
<td>Pre-intervention, Post-intervention and Follow-up:</td>
</tr>
<tr>
<td>Written parental statement about current child-parent communication problems.</td>
<td>Sampling and recording of approx. 10 minutes of semi-structured play session with child – clinician interaction, tapping Gestural communication, Oral communication, Eye Gaze.</td>
</tr>
<tr>
<td>VABS II interview: Expressive communication</td>
<td></td>
</tr>
<tr>
<td>SECDI Words and Sentences: Vocabulary production checklist</td>
<td></td>
</tr>
<tr>
<td>Post-intervention only:</td>
<td></td>
</tr>
<tr>
<td>Course evaluations: Parent descriptions of change</td>
<td></td>
</tr>
<tr>
<td>Follow-up only:</td>
<td></td>
</tr>
<tr>
<td>Retrospective semi-structured interview about change during the last 12 months</td>
<td></td>
</tr>
</tbody>
</table>

Analyses

Video recordings

The author developed a manual for GOCS by coding a range of recordings from home and clinical child-adult interactions by participants in previous ComAlong courses (DLD-version). This resulted in five different coding variables: Contributions that were ‘Oral Symbolic’, ‘Oral Other’, ‘Gestural Symbolic’, ‘Gestural Other’, and ‘Eye Gaze’. These will be returned to in following paragraphs.

GOCS is based on the concept ‘Communicative contribution’ which needs definition. Developed within ‘Activity based Communication Analysis’ (Allwood, 2007) it was there defined as "...what a communicator contributes to an interaction, through spoken (sometimes written) words and gestures (used in broad sense for all visible communicative body movements) at a given point in time, before being replaced by a contribution from another communicator... " It has previously been used in research on communication disorders (Ahlsén, 2008; Thunberg, 2007). Importantly, a ‘Communicative contribution’ is not equivalent to a ‘turn’, as defined in Conversation Analysis, since also feedback utterances or short contributions while someone else is “holding the turn” are considered. In the present study - aiming for easiness and reliability of observation – oral and gestural contributions were observed separately. ‘Contribution’ was thus re-defined as: A single or coherent set of communicative actions within a mode.

The video recordings of the three children, nine recordings in total, were analysed after the Follow-up assessments, in random order. Coding was performed according to the
GOCS manual by the author and by an independent observer who had received four hours of training from the author in the coding procedure, on non-study child recordings. The independent observer, who had five years of experience working as a child speech and language pathologist, was not involved in the project and blind to the conditions of the children, and to the time points of the recordings. This observer was thoroughly informed about the play materials and procedures in the play-sessions to compensate for the author’s first-hand knowledge. To mask the author to the time points of the recordings, these had immediately after recording been re-labelled with random digit codes, and been edited to contain approximately 10 minutes of play and conversation, with a minimum of contextual clues regarding the time points of recordings. For playback, VLC media player™ on a PC with high-end headphones was used. Specially designed Microsoft Excel™ sheets were used for coding, as described below. An overview is also provided (see figure 1).

Firstly, every recording was looked through once focusing oral contributions. If an oral contribution contained at least one readily intelligible word, it was transcribed orthographically directly in the Excel sheet for later analysis, and counted as ‘Oral Symbolic’. If containing no intelligible words, the contribution was counted as ‘Oral Other’. Importantly, an explicit interpretation from the adult in interaction with the child could render symbolic status to a contribution, unless contradicted later on. Possibly intelligible word-like sound-strings were also transcribed, to be able to give them word status further on. Other unintelligible passages of a contribution were transcribed as ‘X’. Onomatopoeia was never rendering symbolic status. This stand was taken due to the vast and fine grained variation in the use of onomatopoeia with respect to symbolic status, observed in clinical recordings, which entailed large time consumption and coder training to ensure reliability. Examples of sound and word productions that would be coded as oral contribution are provided (figure 1, right side, top panels). ‘Oral Symbolic’ and ‘Oral Other’ were summed up into the variable ‘Oral Total’.

Secondly, any gestural contribution was coded. If containing any or several pre-defined forms of symbolic communication (as derived from the context e.g. expanded on by the parent) one count for ‘Gestural Symbolic’ was made. Form-meaning relations were transcribed for later analysis (e.g. [sign-cow]). Accepted forms are displayed (figure 1, top left panel). To make quantification of additional body communication manageable, only ‘Giving’, ‘Showing’ and ‘Pointing’ were accepted. If child used such gestures in a communicative contribution (but no symbolic gestures), that rendered one count for ‘Gestural Other’ to be summed up with symbolic gestures into a ‘Gestural Total’ variable.

Finally, ‘Eye Gaze’ was coded as a separate mode of communicative contribution, if directed at the interacting SLP or at the caregiver. Any gaze at the facial region was considered valid (bottom neck – top head).

Further video analysis
The author carried out descriptive statistics using the data from each observer, child and time point, according to the following paragraphs. Figure 1 provides a flowchart overview.
‘Oral Total’, ‘Gestural Total’ and ‘Eye Gaze’ were summed into the composite variable ‘Overall Rate of Communication’, aiming to reflect the overall rate of communicative attempts and/or interest. ‘Oral Symbolic’ and ‘Gestural Symbolic’ were added together, resulting in the composite ‘Rate of Symbolization’ aiming to reflect linguistic level.

<table>
<thead>
<tr>
<th>Gestural contributions</th>
<th>Oral contributions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gestural Symbolic</strong></td>
<td><strong>Oral Symbolic</strong></td>
</tr>
<tr>
<td>Iconic gestures</td>
<td>Intelligible word forms:</td>
</tr>
<tr>
<td>Emblematic gestures, including Nodding or Shaking head Manual signs</td>
<td>Ambiguous word forms due to homonymy, e.g. [Ɛː] and [Ɛː'] supposedly meaning &quot;här&quot; or &quot;där&quot; (in english here/there);</td>
</tr>
<tr>
<td></td>
<td>Conventional feedback-morphemes (e.g. ‘M-hm’, “Eh?”)</td>
</tr>
<tr>
<td>Symbols mediated by pointing at Communication Boards or manipulating other AAC devices</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gestural Other</th>
<th>Oral Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pointing, Showing, Giving</td>
<td>Oral communicative sounds; Unintelligible speech attempts; Onomatopoeia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gestural Total</th>
<th>Oral Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye-gaze directed at adult’s face</td>
<td></td>
</tr>
</tbody>
</table>

**Overall Rate of Communication**

**Rate of Symbolization**

<table>
<thead>
<tr>
<th>NDW-Gestural</th>
<th>NDW-Oral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of different words - Gestural</td>
<td>Number of different words - Oral</td>
</tr>
</tbody>
</table>

**Symbolic Diversity**

*Figure 1. Taxonomy, coding and analysis of video-data. Communicative interest/efforts/activity operationalized as frequency of Gestural and Oral communicative contributions (Symbolic or Other), summed up together with ‘Eye Gaze’ into Overall Rate of Communication. Grey panels: Language skills operationalized as contributions with symbolic status, summed up in terms of Rate of Symbolization and Symbolic Diversity.*

Additionally, the Number of Different Words, Oral ‘NDW-Oral’ was counted for each transcript, by listing unique items uttered by the child. Ambiguous word forms, e.g. [Ɛː] and [Ɛː’] ‘här’/’där’ (here/there), not possible to distinguish between, not even with leads from the context, were counted as one entry only. In the same way, the Number of
Different Words, gesturally mediated, ‘NDW-Gestural’ were listed and counted. These two measures aimed to capture the children’s expressive ‘Symbolic Diversity’ as spoken orally or by means of gestures or AAC devices.

Due to time variation in recorded interaction, video data were converted to minute rates (e.g. 6.82 Eye gazes per minute). Since no norm referenced data would be at hand for any of the observational measures, it was decided a 20% difference would be considered noteworthy for comparison with parental notions about ‘any important change’.

**VABS interviews and SECDI-WS vocabulary checklist data**

Data from the VABS interview subscale ‘Expressive Communication’ and the SECDI-WS Vocabulary production checklist were compiled as raw scores.

**Semi-structured interviews, Course evaluations and statement data**

The semi-structured interviews were transcribed verbatim and summaries were made according to the following procedure. For each question the main content of the parents’ notions was condensed into one or two sentences. Pertinent and illustrative citations were used, if available. These summaries were checked with parents through a telephone call, as soon as possible, by reading them out loud. Minor completions and corrections were made by the parents. Parents’ statements from course evaluations and written statements about main problems in everyday child-parent communication were scrutinized. Altogether, any data that might either support or contradict recent change indicated by GOCS (at time points one and two) was brought into tables for comparison, case-wise. For an example, see table 4 in the results section.

**Statistics**

**Reliability of observational video-data**

To obtain estimates of reliability regarding the video observation method, agreement between the author and the independent observer was calculated in SPSS v. 22, using two-way mixed effects single measures intra-class correlation coefficients, ICC (Shrout & Fleiss, 1979) for each pair of coded variables. A high ICC value would mean small noise variability (within and between observers) relative to the variability in the children’s performance. The following benchmarks were used: >.80 Excellent, .80-.60 Good, .60-.40 Fair, <.40 Poor. An ICC value of 80, for example, suggests 80% of the variability in the data can be explained by child performance.

To investigate the magnitude and nature of disagreements in every variable, Bland-Altman diagrams were utilized (Bland & Altman, 1986). The between-observer difference (bias) was plotted against the between-observer mean for each recording (N=9). For each variable, the mean between-observer difference was computed to investigate if one observer scored generally higher than the other. The placement of the corresponding 95% confidence interval was used to indicate if bias appeared significant, or random in nature. By visual inspection of the plots, presence of relations between bias and the magnitude of the variables was considered. Raw data was scrutinized regarding explanations to major disagreement.
Validation by comparisons of GOCS data with VABS and SECDI-WS data

Two-tailed Pearson correlation comparisons were carried out regarding sub-scale raw scores from VABS and SECDI-WS on one hand, and presumably related video data on the other:

1. VABS Communication Expressive subscale vs. GOCS ‘Symbolic Diversity’ - a composite of the number of different orally and gesturally mediated words (NDW-Oral + NDW-Gestural)
2. SECDI-WS Vocabulary production vs. Number of different words, orally mediated (NDW-Oral)

Since GOCS data from the two observers could not readily be used interchangeably, between-observers mean values were utilized.

Validation by case data

Parents’ descriptions of their children’s development were compared with video data, case by case. Each case was summarized in a table.

Results

Feasibility of GOCS

GOCS took four hours of training for an SLP working with pre-school children to master. When analysing the data, 10 minutes of video took roughly 90 minutes to code. Compilation and analysis of data were largely automatized by scripts in Excel, except for the symbolic diversity measures, which required a few extra manoeuvres.

Reliability of GOCS

The ICC results indicated excellent inter-observer agreement for all but one variables in GOCS (Table 3).

Table 3

<table>
<thead>
<tr>
<th>Variable</th>
<th>Observer 1 M (SD)</th>
<th>Observer 2 M (SD)</th>
<th>ICC [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Symbolic</td>
<td>5.4 (3.5)</td>
<td>4.9 (3.5)</td>
<td>.94 [.78, .99]</td>
</tr>
<tr>
<td>Gestural Symbolic</td>
<td>0.9 (0.3)</td>
<td>1.1 (0.4)</td>
<td>.80 [.35, .95]</td>
</tr>
<tr>
<td>Rate of Symbolization¹</td>
<td>6.3 (3.8)</td>
<td>6.0 (3.7)</td>
<td>.95 [.89, .99]</td>
</tr>
<tr>
<td>Oral Total</td>
<td>8.8 (4.2)</td>
<td>8.1 (3.3)</td>
<td>.95 [.79, .99]</td>
</tr>
<tr>
<td>Gestural Total</td>
<td>3.2 (1.1)</td>
<td>3.6 (1.2)</td>
<td>.89 [.60, .98]</td>
</tr>
<tr>
<td>Eye gaze</td>
<td>2.8 (1.5)</td>
<td>2.9 (1.2)</td>
<td>.90 [.63, .98]</td>
</tr>
<tr>
<td>Overall Rate of Communication¹</td>
<td>14.8 (6.2)</td>
<td>14.5 (4.9)</td>
<td>.96 [.85, .99]</td>
</tr>
<tr>
<td>NDW*-Oral</td>
<td>2.1 (1.6)</td>
<td>1.8 (1.6)</td>
<td>.97 [.88, .99]</td>
</tr>
<tr>
<td>NDW*-Gestural</td>
<td>0.6 (0.2)</td>
<td>0.6 (0.2)</td>
<td>.91 [.65, .98]</td>
</tr>
</tbody>
</table>

Note: *NDW=Number of Different Words. ¹Composites of the preceding variables.
ICC also indicated excellent agreement regarding the composites ‘Overall Rate of Communication’ (Oral and Gestural contributions regardless of symbolic status, added up with Eye gaze) and ‘Rate of Symbolization’ (Symbolic Oral and Symbolic Gestural contributions). For ‘Gestural Symbolic’ the ICC was ‘good’. Confidence intervals were generally broader for gestural measures including Eye gaze, than in oral measures.

The Bland-Altman analyses are presented below. Between-observer differences are plotted against the means between observers, for each recording, e.g. “Ben 1” referring to Ben at time point 1 (see figures 2a-6). Mean biases, i.e. means of inter-observer differences regarding variables, and 95% confidence intervals are provided.

For oral contributions and NDW-Oral, the biases were positive, confirming that Observer 1 generally had coded oral communication more frequently than Observer 2 (see figures 2a-3), in line with the ICC data table.

![Figure 2a. Oral Total.](image)

![Figure 2b. Oral Symbolic.](image)

Bland-Altman diagrams for Oral contributions. Differences between observers are plotted against the means between observers. Mean observer differences (biases) for each variable and their 95% confidence intervals are represented by thick and thin lines, respectively.

‘Oral Total’ (figure 2a): Mean bias was 0.7 [-0.2, 1.6]. Bias appeared proportionally related to the mean, i.e. to the amount of speech attempts. Raw data inspection revealed Observer 1 had generally coded repeated oral communicative attempts as separate oral contributions while Observer 2 tended to merge them.

‘Oral Symbolic’ (figure 2b): Mean bias was 0.5 [-0.4, 1.4]. No obvious relation between bias and mean was present.

‘NDW-Oral’ (figure 3): Mean bias was 0.3 [0.0, 0.6]. The placement of the confidence interval indicated that bias was systematic. No obvious relation between bias and mean was present.
Figure 3. NDW-Oral (Bland-Altman diagrams for oral variables continued). Differences between observers are plotted against the means between observers. Mean observer bias and its 95% confidence interval are represented by thick and thin lines.

For gestural contributions and NDW-Gestural the mean biases were negative, indicating Observer 2 had generally coded gestural contributions more frequently than Observer 1, including Eye Gaze (see figures 4a-6 below). The diagrams reflect the smaller frequencies of gestural contributions, compared to oral contributions. Relations between bias and the mean were suggested by visual inspection of most gestural Bland-Altman graphs (figures 4b, 5 and 6).

Figure 4a. Gestural Total. Figure 4b. Gestural Symbolic.

Bland-Altman diagrams for gestural variables. Differences between observers are plotted against the means between observers. Mean observer differences (biases) for each variable and their 95% confidence intervals are represented by thick and thin lines, respectively.
Gestural Total’ (figure 4a): Mean bias was ~ -0.3 [-.8 - -.1], and most disagreements were only fractions of one unit.

‘Gestural Symbolic’ (figure 4b): Mean bias was ~ -0.2 [-.4 - -.04]. The placement of the confidence interval indicated this negative bias was systematic.

‘NDW-Gestural’ (figure 5): Mean bias was close to zero: ~ 0 [.05 - .1].

Figure 5. NDW-Gestural (Bland-Altman diagrams for gestural variables continued). Differences between observers are plotted against the means between observers. Mean observer differences (bias) and its 95% confidence interval are represented by thick and thin lines, respectively.

Figure 6. Bland-Altman diagram for Eye Gaze. Differences between observers are plotted against the means between observers. Mean observer difference (bias) and 95% confidence interval are represented by thick and thin lines, respectively.

Regarding Eye Gaze (figure 6), the mean bias was also close to zero: ~ -0.1 [-.5 - .4].

Notably, Bland-Altman diagrams showed that disagreement was often seemingly small. However, small differences appeared potentially important, especially in measurement of small changes for individual children. Therefore, data from both observers have been utilized in the final parts of the Results section.

Validity: GOCS case data on change compared with parental report
In the following pages, each case is summarized in a table (tables 4, 5 and 6), followed by comments. GOCS data from both observers are reported since it would not readily be used interchangeably. If data from either or both observers indicated a noteworthy change at Post-intervention or Follow-up, it has been marked with a footnote. If also consistent with parent report, the footnote is marked by an asterisk. Corresponding
parental report has been similarly highlighted: (e.g. ‘More words, language development fully ongoing’). Apparent contradictions between GOCS and parent report has also been marked.

Table 4

Ben’s case data from at the three time-points of the study: Parent report followed by counts per minute for each GOCS variable, from both observers

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Pre-intervention</th>
<th>Post-intervention</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age 42 months</td>
<td>Age 44 months</td>
<td>Age 48 months</td>
</tr>
<tr>
<td>PARENT REPORT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concerns</td>
<td>Unintelligible speech</td>
<td>Lacking skills to circumvent lack of words</td>
<td>-</td>
</tr>
<tr>
<td>Reported change</td>
<td>‘More words, language development fully ongoing’</td>
<td>More conscious regarding speech: ‘I can – baby cannot’</td>
<td></td>
</tr>
<tr>
<td>GOCS Observer1 (Observer 2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral Symbolic</td>
<td>5.0 (6.3)</td>
<td>10.2&lt;sup&gt;OC&lt;/sup&gt; (8.1&lt;sup&gt;OC&lt;/sup&gt;)</td>
<td>11.1 (10.8&lt;sup&gt;OC&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Gestural Symbolic</td>
<td>0.8 (1.2)</td>
<td>1.4&lt;sup&gt;OC&lt;/sup&gt; (2.0&lt;sup&gt;OC&lt;/sup&gt;)</td>
<td>1.4 (1.4&lt;sup&gt;OC&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Rate of Symbolization&lt;sup&gt;2&lt;/sup&gt;</td>
<td>5.8 (7.5)</td>
<td>11.6&lt;sup&gt;OC&lt;/sup&gt; (10.1&lt;sup&gt;OC&lt;/sup&gt;)</td>
<td>12.5, (12.2&lt;sup&gt;OC&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Oral Total</td>
<td>10.0 (9.1)</td>
<td>14.5&lt;sup&gt;OC&lt;/sup&gt; (11.8&lt;sup&gt;OC&lt;/sup&gt;)</td>
<td>15.3 (11.8)</td>
</tr>
<tr>
<td>Gestural Total</td>
<td>2.4 (3.5)</td>
<td>4.9&lt;sup&gt;OC&lt;/sup&gt; (5.6&lt;sup&gt;OC&lt;/sup&gt;)</td>
<td>3.4&lt;sup&gt;OC&lt;/sup&gt; (3.1&lt;sup&gt;OC&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Eye Gaze</td>
<td>1.2 (2.1)</td>
<td>4.0&lt;sup&gt;OC&lt;/sup&gt; (3.9&lt;sup&gt;OC&lt;/sup&gt;)</td>
<td>4.8 (4.8&lt;sup&gt;OC&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Overall rate of Communication&lt;sup&gt;2&lt;/sup&gt;</td>
<td>13.6 (14.7)</td>
<td>23.4&lt;sup&gt;OC&lt;/sup&gt; (21.3&lt;sup&gt;OC&lt;/sup&gt;)</td>
<td>23.5 (19.7)</td>
</tr>
<tr>
<td>NDW-Oral</td>
<td>1.7 (1.9)</td>
<td>3.0&lt;sup&gt;OC&lt;/sup&gt; (2.1&lt;sup&gt;OC&lt;/sup&gt;)</td>
<td>5.5&lt;sup&gt;OC&lt;/sup&gt; (5.0&lt;sup&gt;OC&lt;/sup&gt;)</td>
</tr>
<tr>
<td>NDW-Gestural</td>
<td>0.8 (0.8)</td>
<td>0.7 (0.6&lt;sup&gt;OC&lt;/sup&gt;)</td>
<td>0.9&lt;sup&gt;OC&lt;/sup&gt; (1.0&lt;sup&gt;OC&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Symbolic Diversity&lt;sup&gt;2&lt;/sup&gt;</td>
<td>2.5 (2.7)</td>
<td>3.7&lt;sup&gt;OC&lt;/sup&gt; (2.7&lt;sup&gt;OC&lt;/sup&gt;)</td>
<td>6.4&lt;sup&gt;OC&lt;/sup&gt; (6.0&lt;sup&gt;OC&lt;/sup&gt;)</td>
</tr>
</tbody>
</table>

<sup>OC</sup> ‘Observed Change’ according to the current observer, defined as ≥ 20% difference compared to previous time point. <sup>OC</sup> ‘Observed Change’ also deemed consistent with PARENT REPORT about change since previous time point. <sup>1</sup> PARENT REPORT notion deemed consistent with GOCS data. <sup>2</sup> Apparent contradictions between GOCS and parent report. <sup>2</sup> Composite of the preceding variables.

In Ben’s case, observers were mainly consistent regarding the absence/presence of noteworthy change (marked with ‘OC’ in table 4). Parent report and GOCS data was partly consistent: Reports of Ben recently had gotten ‘more words’ and about language development ‘fully ongoing’ corresponded to elevated Oral Total, Oral Symbolization and NDW-Oral. Few apparent contradictions between GOCS and parent report were present. On the other hand: At Post-intervention, there were noteworthy changes in video data with respect to Gestural Total, Eye Gaze and Overall Rate of Communication. Ben’s father had expressed notions of such change (e.g. regarding Eye Gaze), but his report lacked sufficient precision in time to validate GOCS data.
Table 5

Gus’ case data from at the three time-points of the study: Parent report followed by counts per minute for each GOCS variable, from both observers

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Pre-intervention</th>
<th>Post-intervention</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age 33 months</td>
<td>Age 35 months</td>
<td>Age 39 months</td>
</tr>
<tr>
<td>PARENT REPORT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concerns</td>
<td>Absent speech*, No interest in verbal imitation, Selective hearing/Receptive difficulties.</td>
<td>Intelligibility when speaking¹</td>
<td>Compliance when needed to hurry</td>
</tr>
<tr>
<td>Reported change</td>
<td></td>
<td>Started to talk¹</td>
<td>Interest in verbal repetition to learn¹</td>
</tr>
<tr>
<td>GOCS Observer1 (Observer 2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral Symbolic</td>
<td>4.1 (4.6*)</td>
<td>4.6 (5.0)</td>
<td>7.2OC* (6.6OC*)</td>
</tr>
<tr>
<td>Gestural Symbolic</td>
<td>0.7 (0.7)</td>
<td>0.8 (1.1OC)</td>
<td>0.7 (0.6OC)</td>
</tr>
<tr>
<td>*Rate of Symbolization²</td>
<td>4.8 (5.3)</td>
<td>5.4 (6.1)</td>
<td>7.8 OC* (7.2)</td>
</tr>
<tr>
<td>Oral Total</td>
<td>6.8 (6.5)</td>
<td>10.1 OC* (10.1 OC*)</td>
<td>9.4 (8.4)</td>
</tr>
<tr>
<td>Gestural Total</td>
<td>2.9 (3.0)</td>
<td>4.9 OC* (5.6 OC)</td>
<td>2.8 OC* (3.0 OC)</td>
</tr>
<tr>
<td>Eye Gaze</td>
<td>1.7 (1.6)</td>
<td>4.9 OC* (3.9 OC)</td>
<td>3.3 OC* (3.1)</td>
</tr>
<tr>
<td>Overall rate of Communication²</td>
<td>11.4 (11.1)</td>
<td>19.9 OC* (19.6 OC)</td>
<td>15.5 OC* (14.5 OC)</td>
</tr>
<tr>
<td>NDW-Oral</td>
<td>0.7* (0.7*)</td>
<td>1.9 OC* (1.4 OC*)</td>
<td>3.5 OC* (3.5 OC*)</td>
</tr>
<tr>
<td>NDW-Gestural</td>
<td>0.4 (0.3)</td>
<td>0.7 OC* (0.8 OC)</td>
<td>0.6 (0.4 OC)</td>
</tr>
<tr>
<td>Symbolic Diversity²</td>
<td>1.1 (1.0)</td>
<td>2.6 OC* (2.2 OC*)</td>
<td>4.1 OC* (3.9 OC*)</td>
</tr>
</tbody>
</table>

OC* ‘Observed Change’ according to the current observer, defined as ≥ 20% difference compared to previous time point. OC* ‘Observed Change’ also deemed consistent with PARENT REPORT about change since previous time point.¹ PARENT REPORT notion deemed consistent with GOCS data. Apparent contradictions between GOCS and parent report. ² Composite of the preceding variables.

As seen in table 5, data from the observers regarding Gus were mainly consistent about presence/absence of noteworthy change. Video and parent data were partly consistent, namely regarding Oral variables: When parent reported of emerging speech and interest in verbal repetition, Oral Total, NDW-Oral and Symbolic Diversity showed parallel noteworthy alleviation. However, Gus was reported to be non-verbal at Pre-intervention, which was contradicted by GOCS data. Moreover, GOCS gains in Gestural Total and Eye Gaze were never mentioned in parental report.
Table 6

Nic’s case data from at the three time-points of the study: Parent report followed by counts per minute for each GOCS variable, from both observers

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Pre-intervention Age 42 months</th>
<th>Post-intervention Age 44 months</th>
<th>Follow-up Age 48 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>PARENT REPORT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concerns</td>
<td>Intelligibility</td>
<td>Intelligibility</td>
<td>Intelligibility</td>
</tr>
<tr>
<td>Reported change</td>
<td>‘Multiword utterances’;</td>
<td>Increased interest in communication(^1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>‘More manual signs’; Varied purposes of communication.</td>
<td>More manual signs(^*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Puts effort in speaking(^1) Begun to relate events beyond here-now(^1)</td>
<td></td>
</tr>
<tr>
<td>GOCS Observer1 (Observer 2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral Symbolic</td>
<td>1.8 (1.1)</td>
<td>1.0(^{+}) (0.5(^{+}))</td>
<td>3.3(^{OC*}) (1.1)</td>
</tr>
<tr>
<td>Gestural Symbolic</td>
<td>0.8 (1.2)</td>
<td>0.5(^{+}) (0.9(^{+}))</td>
<td>1.1 (^{OC*}) (1.2(^{e}))</td>
</tr>
<tr>
<td>Rate of Symbolization(^2)</td>
<td>2.6 (2.3)</td>
<td>1.5(^{+}) (1.4(^{+}))</td>
<td>4.4 (^{OC*}) (2.3)</td>
</tr>
<tr>
<td>Oral Total</td>
<td>4.8 (4.8)</td>
<td>1.6(^{+}) (1.8(^{+}))</td>
<td>7.6 (^{OC*}) (8.3 (^{OC*}))</td>
</tr>
<tr>
<td>Gestural Total</td>
<td>2.1 (2.1)</td>
<td>1.7(^{+}) (2.7(^{+}))</td>
<td>3.7 (^{OC*}) (3.3 (^{OC*}))</td>
</tr>
<tr>
<td>Eye Gaze</td>
<td>1.9 (2.0)</td>
<td>1.8(^{+}) (2.7(^{+}))</td>
<td>1.4 (^{OC}) (1.6 (^{OC}))</td>
</tr>
<tr>
<td>Overall rate of Communication(^2)</td>
<td>8.9 (9.0)</td>
<td>5.2(^{+}) (7.3(^{+}))</td>
<td>12.7 (^{OC*}) (13.2 (^{OC*}))</td>
</tr>
<tr>
<td>NDW-Oral</td>
<td>0.8 (0.8)</td>
<td>0.6(^{+}) (0.4(^{+}))</td>
<td>1.4 (^{OC*}) (0.5(^{OC}))</td>
</tr>
<tr>
<td>NDW-Gestural</td>
<td>0.8 (0.6)</td>
<td>0.2(^{+}) (0.2(^{+}))</td>
<td>0.6(^{OC}) (0.6(^{e}))</td>
</tr>
<tr>
<td>Symbolic Diversity(^2)</td>
<td>1.6 (1.4)</td>
<td>1.2(^{+}) (0.6(^{+}))</td>
<td>2.0 (^{OC}) (1.1 (^{OC}))</td>
</tr>
</tbody>
</table>

\(^{OC}\) ‘Observed Change’ according to the current observer, defined as \(\geq 20\%\) difference compared to previous time point. \(^{OC*}\) ‘Observed Change’ also deemed consistent with PARENT REPORT about change since previous time point. \(^1\) PARENT REPORT notion deemed consistent with GOCS data. \(^{e}\) Apparent contradictions between GOCS and parent report. \(^2\) Composite of the preceding variables. \(^{+}\) Data considered unrepresentative for child’s current ability according to parent.

In Nic’s case change between Pre-intervention and Follow-up was analysed. As seen in table 6, observers were only partly consistent about presence/absence and direction of noteworthy change. Disagreement was however only at hand in the symbolic variables. Parental and GOCS data was also partly consistent: Increased interest in communication and efforts in getting things out were reported in parallel to noteworthy change in GOCS Oral Total and Gestural Total, and hence Overall Rate of Communication, despite at odds with declined rate of Eye Gaze. Report of ‘multiword utterances’ and ‘various purposes of communication’ could not be captured/counted with GOCS in its present state.
**Validity: Correlations with VABS and SECDI-WS data**

Data from each child and time point showed a significant and strong positive correlation between ‘Symbolic Diversity’ (NDW-Oral + NDW-Gestural) and VABS Expressive Communication, $r=0.86$, $n=9$, $p=0.003$ (see figure 7). Similarly, a significant and strong positive correlation was found between ‘NDW-Oral’ data and the Vocabulary production scores from SECDI-WS, $r=0.92$, $n=9$, $p<.001$ (see figure 8).

![Figure 7](image1.png) **Figure 7.** Vineland Expressive Communication raw scores (Maximum score: 108) from three children plotted against Symbolic Diversity (the number of different symbols of any mode produced per minute) at the three time points of the study.

![Figure 8](image2.png) **Figure 8.** SECDI-WS Vocabulary Production raw scores (Maximum score: 711) from three children plotted against the Number of Different Words (oral) produced per minute, at the three time points of the study.

**Discussion**

This study was part of an evaluation project regarding a parent based intervention, ComAlong, targeting children with significant DLD. Within this context, observational study of children’s communicative and linguistic change is still warranted, but suitable methods of observation for children with receptive-expressive language disorders are still lacking. Therefore, a preliminary validation study of GOCS – a method for coding Gestural and Oral Communication and Symbolization in pre-school children – was carried out.

The first main finding was that GOCS appeared to be a feasible tool to code early oral and gestural communication in children with DLD, since it took merely four hours of training for an independent second observer to achieve good to excellent ICC values with the first observer. Moreover, time spent to analyse 10 minutes of video recording was approximately 1.5 hours, which appears promising for both research and clinical settings.

The second main finding was the strong significant correlations between symbolic GOCS data and corresponding VABS and SECDI data, suggesting concurrent validity of GOCS’ linguistics measures. Follow-up interviews and additional parent information also largely supported GOCS data, providing further support for validity.
Bland-Altman analyses suggested caution using data from observers interchangeably, especially for outcome measurement on individual basis and regarding minor changes. Scrutinized raw data revealed validity and reliability issues that will be discussed in the following paragraphs, for each GOCS sub-variable:

Regarding ‘Oral Total’ a small positive mean bias was at hand. Moreover, the magnitude of disagreement appeared proportional to the number of communicative attempts. Probably, first-hand acquaintance by Observer 1 (author) with the children, materials and play-session procedure might have inclined him to code repeated oral communicative attempts as separate intentional and meaningful contributions, while Observer 2, lacking such knowledge, might have been inclined to merge them into larger and fewer units (otherwise meaningless). However, besides the three most extreme disagreements, stemming from recordings with comparatively large amounts of speech and poor intelligibility (Ben at time points 2 and 3, and Nic at time point 3), disagreements were overall less than one oral contribution per minute.

Observer 1’s better chances to identify oral communicative attempts as words would be a main explanation to positive mean biases in ‘Oral Symbolic’ and ‘NDW-Oral’. Albeit showing only small positive mean biases, the bias appeared significant regarding the latter. Poor acuity by Observer 2 was also revealed regarding coding deictic oral contributions as words, and in reconsidering symbolic status for initially unintelligible oral contributions. Observer 1 had problems to detect words from Ben at time point 1, concurrent to the adult’s speech (observer’s own!). Emphatic and emotionally loaded protests from the child appeared to add to this problem. Extreme disagreements regarding the two measures discussed referred to recordings with large amounts poorly intelligible speech (Ben 2, Nic 3). Another extreme (Ben 1) stood out in terms of frustration and strong emotions, and concurrent adult and child attempts to speak. Otherwise, disagreements were less than one contribution per minute for ‘Oral Symbolic’ and only fractions of one unit for ‘NDW-Oral’.

Regarding gestural variables, the Bland-Altman diagrams superficially reflected very small mean biases. However, given the overall low rate of gestural communication in recordings, also small differences between observers were considered as potentially problematic. Starting with ‘Gestural Total’, due to Nic’s low rate of communication in the Nic 2 recording, even singular coding mistakes was revealed having had great impact on agreement of whether change had taken place or not. Obviously, reliability problems would be less regarding larger changes, e.g. between Ben’s or Gus’ performances at time points 1 and 2. It may be remembered that ICC confidence intervals were generally broader in gestural than in oral measures, suggesting larger unexplained variances. Besides low rate of gestural communication, raw data inspection suggested two other sources of variance, namely high rate of gestural communication with concurrent oral contributions, and lack of adherence to the coding manual with loss of observation acuity. In Ben 1 for example, Observer 1 had problems to capture all gestural communication by Ben, who was extremely physically active, directional in interaction (with the author, i.e. Observer 1), using deictic gestures simultaneously to deictic words or vocalizations. Similar issues were at hand regarding the ‘Gestural Symbolic’ variable. The Ben 3 recording revealed disagreement on how to deal with combined symbolic gestures (e.g. when a child emphatically was shaking his head twice, or when shaking head in combination with a hand/arm gesture, suggesting
something was gone/missing). This variable, and ‘NDW-Gestural’, had low mean biases in absolute numbers, but vulnerability to singular coding mistakes due to their low frequency (just like Gestural Total). For example, either of the observers did not notice a head shaking for ‘no’ or an iconic gesture, alternatively disagreed on symbolic status of a gesture. ‘Eye Gaze’ mean bias was close to zero, but the measure still had some reliability issues, related to lack of adherence to the manual, e.g. one observer, while coding Eye Gaze, trying to code gestural behaviours related to eye gaze simultaneously, e.g. ‘Showing’. Due to lack of child cooperation Eye Gaze was also occasionally hard to observe reliably. Linear relations between bias and the mean were suggested by visual inspection of most gestural Bland-Altman plots, but might well have been more complex.

In summary, interactions between measure characteristics, and differences within and between children regarding performance appeared critical for inter-observer agreement in GOCS. Additionally, Observer 2 more often coded Gestural communication than did Observer 1, while the opposite was at hand for Oral communication, suggesting observer attention biases depending on mode. Problems of coding children interacting with oneself appeared to add to poor coding acuity and random error. Capturing small changes reliably appears problematic, also with this method. This has led to the following main findings on how to improve reliability of GOCS.

More training or clearer guidelines are needed regarding: 1) Coding of emergent and deviant oral speech; 2) Revising decisions about symbolic status of Oral sound-strings; 3) Oral speech strongly affected by emotions; 4) Boundaries between repeated contributions, in any mode (when the partner does not reply/take an alternating turn); 5) Parallel or intertwined oral and/or gestural contributions; 6) Gestural communication (methods to enhance alertness might be critical and also consideration that here might be observation biases towards either gestural or oral communication); 7) Symbolic status of gestural communication, and the procedure for revising such decisions while coding. Moreover: The importance of focussing gestures separate from eye gaze may be critical and so may continuity in coding to prevent forgetting about materials and rules. Similar between-observer levels of knowledge about children and materials would be mandatory. Reliability problems with ‘Gestural contribution’ suggested observation of passive children may be deceitful and require special training. Reliability problems with combinations of symbolic gestures should be considered to capture grammatical progress. Low frequency behaviours may be given extra attention while coding. Measuring outcome with GOCS after only two months may be too early.

GOCS was presumed to have strong social validity: According to parents, main positive outcomes of ComAlong courses has been that children’s interest in communication has increased, and/or increased child use of different modes of communication. The study thus aimed to operationalize such interest as frequency of Communicative contributions, Eye Gaze, as well as Symbolic communication skills (Oral or Gestural). The following main findings provide some initial construct validity.

When GOCS data was compared to parent reported data of recent change, large grained consistency was at hand in all cases. Regarding Ben and Gus consistency mainly appeared regarding Oral variables. In the case of Nic both Oral and Gestural GOCS data was in line with parent report of increased interest and ability in communication.
Inconsistencies between GOC-S and parental report were also at hand. First, Ben’s and Gus’ observational Pre-/Post-intervention gains regarding Gestural contributions and Eye Gaze could not be verified by parental report. Parents may have paid more attention to gains in oral than to gestural communication. Moreover, noticing changes within different ‘modes’ in parallel may be difficult. Parents may not have noted gains in eye gaze if problems were subtle, while children’s gains in the GOC-S variable Eye Gaze may be related to increased acquaintance with the clinician and the setting. Second, parent reported about Gus as being mainly non-verbal at Pre-intervention, while observation at that time point revealed Oral Symbolic contributions at least every 15th second. However, Gus’ speech at that time mainly consisted of deictic words, yeses, no’s, and feedback morphemes, which parents might not have considered as speech. Third, parent report of Nic’s increase in manual signing at Follow-up was not consistently captured by GOC-S. Although Gestural Total was notably higher, NDW-Gestural remained very low. Parent may have over-estimated gains in Nic’s manual sign lexicon. On the other hand, a brief video sample may not always capture a child’s actual symbolic skills (regardless of mode).

The results of ‘VABS Expressive Communication’ and ‘Symbolic Diversity’ correlated significantly, suggesting that number of different observed words and symbolic gestures may be a valid way of capturing expressive communication and language status. In addition, the oral symbolic diversity measure ‘NDW-Oral’ correlated strongly with the MacArthur-Bates SECDI-WS Vocabulary production scores, suggesting that observed number of different orally mediated words can be trusted to provide a valid measure of a child’s expressive vocabulary. Since the ‘NDW-Oral’ measure requires extra analysis manoeuvres, a post hoc comparison was carried out between SECDI-WS and the more convenient ‘Oral Symbolic’ (the rate of oral contributions containing at least one intelligible word). Here also, a statistically significant and strong correlation appeared. Thus, for a beginning speaker, also ‘Oral Symbolic’ might be a valid outcome measure. The composite ‘Rate of Symbolization’ might also be useful, especially for children dependent on manual signing and other AAC, but validation through comparison with parent report would require questionnaires capturing gestural symbolic skills. Undesirable ceiling effects in GOC-S may be at hand. One child (Ben) appeared to be close to a ceiling regarding ‘Overall rate of Communication’ at time points 2 and 3, with an assertive and directive communication style. Further development of GOC-S should probably incorporate communicative behaviours reflecting receptive/responsive skills, such as compliance to suggestions and instructions, or looking at objects mentioned or otherwise referred to by the partner. Counting feedback contributions separately and incorporating them into a receptive/responsive sub-measure might also be fruitful. Imitative play and other play behaviours (e.g. actions following the joint activity plot) had indeed initially been considered for coding, but were left out because of delineation problems, perhaps prematurely.

Limitations and further research
The small sample of children and parents was a major limitation in this study. Even though the participant children all had significant mixed receptive-expressive language delays with or without concomitant neurodevelopmental problems, and thus were typical for the target population of ComAlong courses, further validation would require larger numbers of children. Moreover, they would preferably be followed over a longer
period since change may be slow. An inquiry into GOCS’ sensitivity to development in typically developing children would also be informative.

Further considerations regarding GOCS different variables may be warranted. For example, GOCS in its present state dismissed child initiatives, which is widely used as an outcome variable. In contrast, GOCS include imitations/repetitions. Child repetitions may indeed be a highly valid outcome measure in severe communication delays and disorders. Support for this view can be found in the literature, e.g. in the treatment study by Kot and Law (1995), but may be further investigated.

Eye Gaze directed at the facial region of the adult was used as a convenient measure of communicative interest (instead of the more complex Joint attention behaviours, and with no demands of verifiable eye-to-eye gazes). It is well established that Joint attention is closely related to language growth in young children, see e.g. Bon, Daley & Sigman (2004). Whether frequency of Eye Gaze alone would measure up as a ‘proxy measure’ for Joint Attention remains to be proved. However, it has been verified that Joint Attention frequently occur simultaneously to gesture behaviours such as those used in GOCS in children interacting with their caregivers (Tomasello & Farrar, 1986), so the validity of combining separate measures of Eye Gaze and gestures does have some support. Deviant eye gaze behaviour of persistent staring would probably not inadvertently render high GOCS scores, since it is the number of occurrences per time unit that is counted. Still, Eye-gaze might be difficult to capture from video recordings and might be best judged by the person interacting directly with the child. This variable might therefore be questioned.

Child communication elicited by a semi-standardized clinician-child play-session (as in the present study) may be different from parent-child interaction at home, which calls for further investigation. Singular short video samples may be unrepresentative, which was dealt with by asking parents about how typical the children’s performances were. Such caution must be taken also in future research.

Observer bias might well have been present as the author was not fully blinded as video observer. Further validation of GOCS calls for independent, blinded observers. Measurement of intra-observer reliability was not completed within the present pilot study due to time and funding limitations, but will be mandatory in further validation.

Although there were strong significant correlations between validated instruments (VABS and SECDI-WS) and selected sub-measures of GOCS, these results must be interpreted with caution: Data collection by parent report was not totally independent from observational data. Parents might have been influenced by their impressions during clinical visits when they filled out the SECDI questionnaires. Also, the author might have been affected by own impressions of the child from a recent clinical encounter when interviewing parents with VABS, and in turn affected the interviewees. Independent measures would be warranted.

‘20 % change’ in GOCS data was an arbitrary criterion for consistency with a parental notion of change, due to lack of norm referenced data. Any numeric level would be problematic, e.g. when tapping change from a zero level to a level of some occurrence, rendering heavy impact in numbers, while parent notion of progress might be different, e.g. ‘hardly noticeable’. The relationship between parent notions of progress and numerically encoded change may be very complex.
Retrospect interviews made it difficult for parents to recall the exact timing for change. All parents were sure that significant change had occurred during holidays, when they spent more time with their children. Uncontrolled variation in parents’ opportunities to observe their children may thus be a problem in this and similar studies. If interview data would be used in further validation of GOCS (or as a qualitative measure in clinical studies) more precise and frequent probes would be preferred.

Conclusions and implications
GOCS may provide feasible, valid and reliable components in outcome measurement for pre-schoolers with severe communication problems both in research and in clinical settings. GOCS appears time efficient, requires limited expertise and technical equipment, making it promising for clinical research.

Further development and validation of GOCS is warranted, with larger samples and diverse groups of children, use of video-data from parent-child interaction and with consideration of present main findings regarding validity and reliability issues.
REFERENCES


