Study on pragmatic assessment data reliability in children with typical language development

Resumo

O objetivo desse estudo foi verificar o momento com maior fidedignidade de dados do processo de avaliação da linguagem, para realizar o levantamento do perfil pragmático infantil. Participaram cinco crianças, com desenvolvimento típico de linguagem, e idades entre 7 anos e 1 mês e 8 anos e 11 meses. Foram realizados 150 minutos de gravação, em uma situação de interação da criança com a pesquisadora, divididas em cinco sessões individuais de 30 minutos. Houve análise posterior dos dados, segundo o protocolo de habilidades comunicativas verbais (HCV), sendo delineado o perfil pragmático individual de cada filmagem (30 minutos) e de toda a amostra (150 minutos), para a comparação (sessões 1 a 5 total geral das sessões) dos índices de fidedignidade (IF) e status de confiabilidade (SC). Para o cálculo do IF e do SC, respectivamente, foram realizadas as análises individuais interobservador e intraobservador. Os resultados apresentados pelas crianças 1 e 2 alcançaram maior IF na sessão 2; os da criança 3 apresentaram valores semelhantes de IF nas sessões 3, 4 e 5; os da criança 4 obtiveram o maior IF nas sessões 1 e 3; e os da criança 5 alcançaram o mesmo valor de IF em todas sessões. Com relação ao SC, a sessão 2 apresentou maior porcentagem de altíssima confiabilidade para a maioria das crianças, seguida da sessão 3. Na análise realizada por categoria de HCV, a sessão 3 apresentou maior SC para as habilidades dialógicas, narrativo-discursivas e total geral de HCV. No geral, observa-se que as sessões 2 e 3 foram as que permitiram alcançar maior IF e SC na análise realizada para delineamento do perfil pragmático infantil.
INTRODUCTION

Pragmatics refers to the correspondence between the language abilities and the principles governing the functional use of language, including the social conditions and the rules that govern the use of language in a communicative interaction\(^1\). In Speech-Language Pathology, studies on the development of pragmatic abilities are new in relation to studies of morphosyntactic, semantic and phonological development\(^2\).

In the 70s, the pragmatic approach was included into the studies of language that emphasizes the communicative factors of the language because of the need to relate this to the context. This approach has modified the way of carrying out the child’s language, because the interest has become the communication abilities in general (encompassing speech acts, communicative intentions of the speaker and its communicative functions), demonstrating that words and phrases presented change in meaning depending on the context in which they were produced\(^3\).

A part of the analysis focuses on the pragmatic use of communicative abilities; the description allows defining the pragmatic profile of the subject, contributing to a more effective communication skill they use in different contexts and with different interlocutors\(^3\).

Language samples provide a clear description of communicative abilities that the individual uses and allows for a detailed analysis of its dimensions and processes. The analysis of spontaneous communication is the assessment procedure which provides a more accurate description of the level of language development, and allows a wide variety of analyzes (pragmatic, phonological, syntactic, semantic, etc.), therefore, reducing the risk of subjective interpretations or loss of information\(^3\).

Some studies have shown that 30 minutes of recorded language sample already shows enough to outline the pragmatic profile of children with both typical language development\(^2\), and children with language disorders of various etiologies, such as autism\(^4,5\), Asperger syndrome\(^4,5\), Down syndrome\(^6\) and hearing loss\(^7\). There are studies in the literature\(^1,8,9\) which reported recording times with less than 30 minutes of length, with samples up to five minutes of recording\(^9\), though only for the pragmatic analysis using protocols that only classify abilities as appropriate or inappropriate – differing from the instrument used in this study, or in populations with specific language characteristics\(^1,8,9\).

However, for clinicians and researchers working on the design of pragmatic profile of children, there is the question about the best time to perform this analysis, especially when it comes to the initial sessions (usually takes place when the clinical assessment). Therefore, the objective of this study was to verify the time with more reliability to survey the children’s pragmatic profile, in terms of obtaining the greatest possible reliability in the data.

CLINICAL CASE PRESENTATION

The research presented here is characterized as a multiple-case study. All procedures were submitted and approved by the Research Ethics Committee of the Bauru School of Dentistry, Universidade de São Paulo (FOB-USP), under protocol number 060/2009. This study was conducted at the Speech-Language Pathology and Audiology Outpatient Clinic of the same institution.

The study included five children, three boys and two girls, selected according to the following criteria: (a) children with typical language development, excluding those who were suspected with language disorders in the speech-language and hearing screening performed by the researcher; and (b) children aged from 7 years and 1 month to 8 years and 11 months.

We opted for the age group from 7 to 8 years old, in an attempt to minimize interference in the process of language acquisition during the assessment of the pragmatic profile proposed, since children with typical language development who are older than 7 years have already developed pragmatic abilities by then. Moreover, it is expected that children in this age group make use, predominantly, of verbal language over other forms of communication.

For the assessment of typical language development, we performed a speech-language screening. This consisted of a questionnaire about the child’s development, which was applied with parents present, and a checklist developed specifically for this study, with data about the child’s oral language, which was completed by the researcher after the interaction with the child, which lasted approximately 30 minutes. In this screening, if there was any complaint/suspected risk factors for speech disorders, language (expressive and/or receptive) or hearing, the child would participate in the research and would be redirected to the necessary clinical procedures available at the Speech-Language Pathology and Audiology Outpatient Clinic of FOB-USP (clinical and/or complete audiological assessment, aiming to establish a speech-language diagnosis).

We collected 150 minutes of recording time with each child in five 30-minute samples (sessions with the child’s interaction with the researcher), totaling 750 minutes of recording. The recording took place in a situation of spontaneous interaction between the child and the researcher, with the use of previously selected playing materials. The children in the study had no previous contact with the researcher, to avoid the influence of the variable of familiarity with the interlocutor. The playing materials were selected according to age and gender of children, and each session was planned in order to make available, in interactive space, materials and toys suitable not only for age and gender, but that should facilitate verbal interaction, so that all communication abilities proposed by protocol\(^5\), could be used by children in the sample.

Subsequently, each recording was transcribed and transferred to the Protocol of Verbal Communication Abilities (VCA)\(^5\) for recording and analysis of pragmatic verbal abilities. As described in the protocol, are considered verbal communication abilities (VCA): dialogical abilities (DA) (which are the basic abilities to initiate and maintain a dialogue), regulation abilities (RA) (which correspond to emissions, which aims to regulate some behavior), narrative-discursive abilities (NDA) (which reports what is real or imaginary, and narrative-based reasoning) and noninteractive verbal abilities (NIA) (such as...
the use of meta-language). A more detailed description of each category is found at Appendix 1.

The pragmatic profile was reached after the transcribed recordings and after the categorized VCA used in each turn of dialogue from both persons (adult and child). We performed quantitative analysis of each child’s ability to calculate the percentage of use of each ability in relation to the total abilities used. For example, from the 100 verbal communication abilities of the sample, 10 were of initiate shift, i.e., 10% of the total.

We analyzed separately the VCA used by each child, to create the pragmatic profile of each subject and in each session, we verified the reliability level of data, by the analysis of inter-observer and intra-observer

**Inter-observer analysis**

The inter-observer agreement regarding all recordings was analyzed, as a way to control the reliability of data by two independent observers, being the researcher, the observer 1 and a speech pathologist trained in the analysis of the categories considered in the protocol used, the observer 2.

The observer 1, having already prior knowledge of the categories used by the protocol, conducted training with observer 2 in order to minimize any possible doubts in the analysis.

All VCA protocol were categorized, for both observers (1 and 2) and then the comparison was made between the analysis of the observers, each of the recordings, turn by turn, taking as basis the analysis of the researcher (taken as a basis for comparison) in order to calculate the correlation.

The agreement was analyzed in each session for each child, calculated by the point-to-point technique. It was considered reliable data with at least 75% agreement\(^{(10)}\) according to the formula:

\[
\text{Agreement} = \frac{\text{Agreements}}{(\text{agreements + disagreements})} \times 100
\]

For each of the participants it was calculated a reliability index (RI), by comparing the values generated by the formula, presented between the overall average of the sessions and the five sessions. Thus, it was possible to comparatively analyze the data of the five children, individually and by average, allowing the definition of which session(s) provided the biggest RI data, in the design of the pragmatic profile of the child.

As the result of the RI, there was a variation of 75.2% in session 1 of the child 2, and 100% in sections 3, 4 and 5 of the child 3, as well as in sessions 1 and 3 of the child 4, and in all sessions of the child 5.

Regarding the result of the average RI of each child, there was a variation of 86.7%, submitted by two children, 100% obtained by the child 5. It should be noted that for children 3, 4 and 5 the results from the RI sessions were above 99%, while for the first child, the results were above 85% and, for the second child, these values were above 75%.

With regard to the higher value of RI for each session, the results presented by children 1 and 2 reached the highest value in session 2, the third child had similar values in sections 3, 4 and 5, the results showed that the child 4 had RI results higher in sessions 1 and 3, and the child 5 reached the same value of RI at all sessions. Note that the sessions 2 and 3 had larger RI to 60% of children, corresponding to three of a total of five children.

In general, for all children, the values of RI became higher right from the second session and after, approaching 100% (with indices ranging from 99.5% to 100%).

**Intra-observer analysis**

It was also conducted an individual intra-observer analysis (the base model was the observer 1), by verifying the percentage values calculated for the abilities assessed, session by session (sessions 1-5) and the total sum of sessions (sum of the five sessions) of each of the children. For each of these percentages found, it was given a degree and a level of reliability. The degree ranged from A to F, and the level ranged from A1 to F2 (Chart 1).

**Chart 1. Reliability parameter for calculating the reliability status (RS)**

<table>
<thead>
<tr>
<th>Scale</th>
<th>Degree</th>
<th>Level</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 – 90.0</td>
<td>A</td>
<td>A1</td>
<td>100 – 99.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A2</td>
<td>94.9 – 90.0</td>
</tr>
<tr>
<td>89.9 – 70.0</td>
<td>B</td>
<td>B1</td>
<td>89.9 – 80.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B2</td>
<td>79.9 – 70.0</td>
</tr>
<tr>
<td>69.9 – 50.0</td>
<td>C</td>
<td>C1</td>
<td>69.9 – 60.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C2</td>
<td>59.9 – 50.0</td>
</tr>
<tr>
<td>49.9 – 30.0</td>
<td>D</td>
<td>D1</td>
<td>49.9 – 40.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>39.9 – 30.0</td>
</tr>
<tr>
<td>29.9 – 10.0</td>
<td>E</td>
<td>E1</td>
<td>29.9 – 20.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E2</td>
<td>19.9 – 10.0</td>
</tr>
<tr>
<td>9.9 – 0.0</td>
<td>F</td>
<td>F1</td>
<td>9.9 – 5.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F2</td>
<td>4.9 – 0.0</td>
</tr>
</tbody>
</table>

Then, we performed a comparison of the degrees and levels of percentage usage of each verbal communication abilities in each session, with the grand total of the sessions for each child, to check the reliability parameter (reliability status – RS). The closer these degrees and levels were, the greater were the trust status of the comparison. The standardization of this type of reliability analysis was based on and adapted from a manual\(^{(11)}\).

The SC (Chart 2) varied in different gradations, and the comparison of data between sessions was considered of very high reliability when the letters were the same grade and the same level (e.g. C1 and C1); of high reliability when letters were in the same grade, but different levels (e.g. C1 and C2), considered medium reliability when there was both different degree and level (e.g. C2, D1); low reliability when there was a difference of two degrees and two levels of letters (e.g. C2 and D2); very low reliability when there was a difference of two degrees between letters (e.g. C2 and E2), and was considered unreliable when there were more than two degrees difference between letters (e.g. C2 and F2).
These percentage ranges for calculating a proposal was already used in a work done in the area of public health, based on already standardized data\(^{(11)}\) and does not require the application of statistical tests to validate it.

To check which session allowed us to achieve a higher percentage of high degrees of reliability analysis by the proposal, we will describe the data regarding the percentage of RS in each of the possible levels and degrees (ranging from very high degree of reliability to unreliable data).

For all children, the highest RS had a higher percentage, ranging from 57.1% in session 5 for the Child 4, 82.1% in sessions 3 and 5 for the child 5.

The values of the percentage of the highest reliability status of each session, of each child can be seen in Table 1.

Table 1. Percentage of highest reliability status of each session for each child

<table>
<thead>
<tr>
<th>Child</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st session</td>
<td>60.7</td>
<td>67.9</td>
<td>64.3</td>
<td>67.9</td>
<td>67.9</td>
</tr>
<tr>
<td>2nd session</td>
<td>67.9</td>
<td>71.4</td>
<td>75.0</td>
<td>67.9</td>
<td>67.9</td>
</tr>
<tr>
<td>3rd session</td>
<td>71.4</td>
<td>64.3</td>
<td>60.7</td>
<td>67.9</td>
<td>82.1</td>
</tr>
<tr>
<td>4th session</td>
<td>64.3</td>
<td>64.3</td>
<td>67.9</td>
<td>71.4</td>
<td>78.6</td>
</tr>
<tr>
<td>5th session</td>
<td>56.3</td>
<td>67.9</td>
<td>64.3</td>
<td>57.1</td>
<td>82.1</td>
</tr>
</tbody>
</table>

To check which session allowed achieving the highest percentage of RS to the overall VCA total and for each VCA category analyzed using this protocol, there was a comparison of this value for each of the sessions, with the overall total of the sessions by category of VCA and the overall total of these abilities (Table 2).

With respect to the overall total of VCA used (all categories added: DA, RA, NDA and NIA) session 3 showed higher RS for three children, accounting for 60% of the sample. Session 5 showed greater RS for two children (40%) and sessions 1 and 4 showed higher RS for one child each, corresponding to 20% of the sample.

Overall, in the design of children’s pragmatic profile of the sample analyzed, the sessions 2 and 3 achieved the highest index values of reliability, indicating that these sessions were similarly analyzed by the two observers, whereas the pragmatic assessment is subjective. The highest values of reliability status for the general analysis of verbal communicative abilities were also obtained in sessions 2 and 3, and the third session was the one that obtained the highest values of reliability status, in the analysis of abilities by category. This demonstrates that in sessions 2 and 3, the number of VCA was nearest to the total of the sessions, indicating a situation closer to reality.

DISCUSSION

This study aimed to identify the moment of greatest reliability to perform a reliable and punctual pragmatic analysis, based on an analysis of 750 minutes of recordings, with proven reliability and situations by the data, i.e. no inferences. A high level of reliability shows that the analyzed data is reliable, even if it is a situation where there may be subjective interpretations, as it is the case of pragmatic analysis.

In the study presented here, we observed a high rate of agreement between the two observers, and such a fact was evidenced by the high percentage of reliability indices in all sessions for all children and the fact that no session submitted this index below the minimum correlation value that was adopted by the formula, above 75%. One hypothesis for the emergence of a high level of reliability in the assessments can be the training of observers. The previous knowledge and experience in relation to the abilities assessed, regardless of which protocol is adopted, the training is essential to obtain more reliability.

However, the training of observers could only be done because there were prior planning sessions. In the design of a pragmatic profile, there needs to be planning for each session, which is proportional to the appearance of all verbal communication abilities to be evaluated. The literature makes clear the importance of activities to be planned and the structured situations, noting further that the success of rehabilitation depends on a correct planning of the evaluation process\(^{4,12}\). The pragmatic profile found in this study was varied and can be attributed to the diversity of planning sessions with playing materials (assorted toys) to allow the use of the analyzed abilities.

The highest index values of reliability status found in sessions 2 and 3 in the study presented here indicates that, in the design of a pragmatic profile of a child, regardless of the initial session,
other sessions can still obtain reliable data because the data are consistent in the sessions observed, even with no familiarity with the interlocutor. In a work done with children with specific language, it was also noticed that pragmatic analysis should be performed on medial or late samples, as the initial time values were different regarding to the communication means used.

As the literature indicates when family situations are provided more communicative initiatives and greater responsiveness to the caller are more likely to occur, providing a performance closer to the actual abilities of each individual. It suggests the hypothesis that if – in the study presented here – the caller was familiar (parent or caregiver), high levels of trust and reliability data could have been obtained from the first session.

It is emphasized that the data presented here are grounded on an extensive data sample – totaling 750 minutes of analyzed verbal communication abilities – even if it is a case study, in which there are a limited number of participants. Every knowledge area requires data showing empirical evidence and observed in daily practice. Thus, an extensive work based on samples contributes to build the area.

**FINAL COMMENTS**

The study presented here indicates that the sessions 2 and 3 were the ones that provided in relation to the other sessions (sessions 1, 4 and 5), the highest values of reliability.

The moment with more reliability then to outline the pragmatic profile of children with typical language development is near to the beginning of the contact - not the initial session, but the consecutive ones.

Given the rigor of the methodology presented, we suggest to replicate this study in a larger sample and also with various interlocutors, so that clinicians and researchers working with pragmatic aspects will be able to base their analysis on reliable data.

**REFERENCES**

Appendix 1. Verbal communication abilities

**Dialogical abilities (DA)**

Start of turn (ST). Ability to initiate a dialogue, even when no subject was discussed, with the exception of conventional social greetings. Ex: one party says to the other “Let’s play ball?”

Dialog maintenance (DM). Ability to maintain a topic of conversation proposed by the interlocutor (keeping with the context) or try to focus a person’s attention on a topic already started (this includes features like repeating part of a statement for later continuity, avoiding breaking the dialog). Ex: when a caller says “Let's play ball?” And the other responds “ball? Okay, but only if it's football.”

Insertion of new topics in the dialogue (NT). Ability to suggest, in a dialogue, new topics of conversation. Ex: when they are playing football, one says to the other “I wonder who will win the Premier League?”

Sequential dialogical organization (SO). Ability to respect the conventions of sequential organization of talks, to fill dialogic turn, through features such as:

1. Comments (CM) - emissions used to identify or describe objects, people or actions no other function than to share the information with the caller. Such statements may constitute a complete or verbal vocalizations (including onomatopoeia or songs). Ex: one person says “This car is a VW” and mimics the sound of the car.

2. Direct answers (DAN) - when, after an inquiry made directly or indirectly by the other party, there is the presence of a contextual verbal response or motor acts (accompanied by vocalizations). Ex: a caller asks “Can you pick up the pen to me?” And the other says “Okay here's your pen!”, While the handle and returns another.

3. Imitation (I) - when, to complete a round of dialogue, there is only a repetition of the speech of the speaker or some other issue related to the subject and evoked through dialogue. Ex: one person says “What's your favorite novel?” And the other responds “Novel ... The Cattle King, the Globe and you - all about.”

4. Feedback to the interlocutor (FI) - composed of statements or expressions that indicate only attention to the speech of the other, with the aim of strengthening or repair. Ex: when a speaker is talking and the other exclaims “Uh, huh” or “Right, right...” or “Speak up.”

Failure repair (FR). When there is a full or partial repetition of an issue, to correct any errors in pronunciation or formulation itself or the other. Ex: a caller is speaking “Yesterday, I went to paque, meaning the park.” Change of roles (CR). When there is recreational use of verb forms to correct any errors in pronunciation or formulation itself or the other. Ex: a caller asks “Why is it that the witch wanted to kill Snow White?”

Insertion of new topics in the dialogue (NT). Ability to suggest, in a dialogue, new topics of conversation. Ex: when they are playing football, one says to the other “I wonder who will win the Premier League?”


**Narrative-discursive abilities (NDA)**

Storytelling or telling of an event (ST). Ability to report a fact or story consistently through spontaneous emission, with or without the aid of the party. Ex: a caller starts to tell a story, from pictures he sees “Once there was a girl who lived ... sad.”

Reproduction of stories (RPS). Ability to reproduce all or part of a fact or story told or read by others, with or without assistance from the interlocutor. Ex: when you finish telling a story such as Snow White, the other immediately or later reproduces correctly “Once there was a pretty girl, bright white, bright white as snow ...”.

Interpretation of stories (IS). Ability to draw conclusions and issue opinions on facts or stories and understand them. Ex: after telling a story, ask yourself “Why is it that the witch wanted to kill Snow White?” And the other responds “Because the witch was ugly and bad and had very envious of pretty girl and nice.”

Arguments (ARG). Ability to use own issues to convince the other, using verbal arguments and convincentes. Ex: A caller says “Now, we have to store toys and go home.” And the other responds “But it's still early and my bus will take to pass moreover you promised to let me see the book again.”

**Verbal noninteractive abilities (NIA)**

Use of language to establish one’s own identity (LOI). When a person refers to himself in his verbal emissions. Ex: “Hence, I became very angry and said - do not hit me anymore, I’m strong.”

Symbolic play (SP). Use of language to establish relationships of representation direct or indirect objects, actions, or people with certain verbal expressions. Ex: a caller says, to tell the story of Little Red Riding Hood - “My eyes are big, but it's better to see you!”. Making the intonation of speech lobe.

Metalanguage (ML). When the individual uses speech to refer to their speech or language. Ex: “I think talk is for people to be just so, moving his mouth. I’ve even thought about before just talk - a response to another caller, when asked why people spoke.

Note: It is important to emphasize that a statement may have more than one function, and therefore that all functions used are noted. Ex: when after telling a story, it is a question about it and the caller responds, he is using narratives/discourse abilities (NDA) interpretation of stories (IS) and dialogical abilities (DA) in dialog maintenance (DM) and direct response (DR).