SLPI PAPER #19:
MONITORING THE CONSISTENSY OF YOUR SLPI TEAM MEMBERS' RATINGS

Dr. Frank Caccamise
SLPI Co-Developer/Consultant
September 1997 (1st ed.), February 2008 (11th edition)

Introduction

The purposes of this SLPI PAPER are: (1) to provide you with information about the importance of SLPI Coordinators monitoring the SLPI ratings of their SLPI Team Members, (2) to encourage programs using the SLPI to conduct SLPI reliability and validity studies, and (3) to report on SLPI reliability and validity studies.

Why Monitor SLPI Ratings?

It is important for programs using the Sign Language Proficiency Interview (SLPI) to monitor the consistency of SLPI independent ratings prior to team discussion both across SLPI Team Members (inter-rater reliability) and to monitor how well Team Members' independent ratings prior to team discussion compare to official ratings (validity). The two primary reasons for the above monitoring are: (1) to help ensure fairness to candidates/interviewees taking the SLPI, and (2) to assist in planning and providing in-service training as appropriate for SLPI Team Members.

In regards to the first reason, fairness to interviewees taking the SLPI, it is important that interviewees are given the opportunity to show their best sign language communication skills. This requires that interviewees: (a) are provided information that is intended to help prepare them for taking the SLPI, (b) receive SLPI interviews that are conducted in a manner that allows them the opportunity to show their best sign language communication skills, and (c) receive fair and accurate SLPI ratings. Relative to the latter the use of more than one rater with each rater providing   independent ratings prior to rater team discussion helps to provide an on-going check on the consistency of raters within normal SLPI rating procedures. In addition, SLPI interviewees, raters, and coordinators may request re-ratings and re-interviews if they believe rating results or interviews do not accurately reflect interviewees' true skills. It is the SLPI process as a whole, therefore, that helps to ensure interviewees have the opportunity both to show their best sign language communication skills and to have these skills rated in a consistent manner.

The second reason for monitoring consistency of raters' ratings, to assist in planning and providing SLPI Team Members with appropriate in-service training, is important to ensure SLPI Team Members are interpreting the SLPI Rating Scale in a correct and consistent manner. For example, if rater A provides ratings that are consistently different from the ratings of other SLPI Team Members, her/his SLPI Coordinators should consider conducting one-on-one in-service training with Rater A. Also, if two raters rate an interview as Advanced and another rater rates this same interview as Intermediate, their SLPI Coordinator should consider meeting one-on-one with this latter rater or bringing all three raters together to discuss the appropriate rating for the interviewee. In addition to one-to-one and small group in-service training, it is important to periodically bring all SLPI Team Members together for rating practice and discussion. The literature for the Language/Oral Proficiency Interview (L/OPI), the spoken language assessment tool upon which the SLPI is based, recommends this type of "re-calibration" occur every two-to-three years for L/OPI raters.

Encouragement and Support for Monitoring the Consistency of Your SLPI Team
in Conducting SLPI Ratings

You do not need to plan and conduct a complex research study to monitor the consistency of SLPI ratings conducted by your SLPI Team; that is, you may answer the following questions by recording results for your normal SLPI rating procedure in a Sign Language Program Database:

1. How often are your raters' independent first ratings the same or within one level of one another?

2. When your raters' first independent ratings are not the same or within one level of one another, how often do second independent ratings following second viewings of SLPI interviews result in ratings that are the same or within one level of one another? [a]

3. If re-ratings are requested by persons taking the SLPI or by the SLPI Coordinator,

    A. How often are ratings of second rating teams in agreement with the ratings of first rating teams?

    B. When ratings of first and second rating teams not in agreement, how do ratings of your third rating teams compare to the with ratings of your first and second ratings teams?

If your resources allow, as part of a study to monitor consistency of raters, you may wish to have two or more rating teams rate the same SLPI interview. Prior to doing this your normal rating procedure should be followed until an official rating has been determined.

SLPI **Final Ratings** and **Official Ratings**

Given the above discussion, when recording results for monitoring consistency of SLPI ratings by your SLPI Team Members, it is important to be clear on the difference between SLPI **final ratings** and SLPI **official ratings**:

1. **Final ratings** are ratings made by individual raters. Each rater always provides an exact or single level rating (for example, Superior Plus, Superior, Advanced Plus, and so forth).

2. **Official ratings** are ratings reported to persons taking the SLPI. Official ratings may be reported as **ranges** (for example, Novice Range) or as **exact single level** ratings (for example, Novice and Novice Plus).

_

____

[a]Second independent ratings prior to discussion are made when all three raters first independent are not within one level of one another.

SLPI Reliability and Validity: Three Rater Team Procedure

In a 1996 NTID research proposal F. Caccamise, W. Newell, H. Lang, and D. Metz outlined studies designed to investigate the following three questions:

1.   How consistent are ratings within SLPI Rating Teams (three raters per team)
2.   How consistent are raters in interpreting the SLPI Rating Scale? (three raters per team)
3.   How consistent are ratings across SLPI Rating Teams of three raters?

Rater Consistency <u>Within</u> SLPI Rating Teams and in Interpreting the SLPI Rating Scale

In a 1999 report for a pilot study designed to investigate the first two questions posed above, F. Caccamise and W. Newell provided analyses of 205 SLPIs conducted at NTID from 1993-to-1996. Relative to the first question, consistency of ratings within SLPI Rating Teams, analysis showed that for the 191 SLPIs receiving official ratings all three NTID SLPI Rating Teams members (prior to any rater discussion) were within one level of one another 136 times (72.2%) for first independent ratings and were within one level of one another 181 times (94.8%) by second independent ratings. Relative to the second question (consistency of raters in interpreting the SLPI Rating Scale), analysis showed that for the 191 SLPIs receiving official ratings: (1) of 573 first independent ratings, 546 (95.3%) were within one level of official ratings; and (2) of 168 second independent ratings, 164 (97.6%) were within one level of official ratings. These pilot study results support that ratings within NTID SLPI Rating Teams of three are consistent and that NTID SLPI Raters are able to consistently interpret the SLPI Rating Scale. In addition, when all three raters' first independent ratings are not within one level of one another, second independent ratings significantly improve consistency among raters prior to discussion among raters.

During the academic 2001-02 and 2002-03 academic years (AY01-02 and AY02-03), NTID used the SCPI Individual Rater Procedure and then in AY03-04 NTID re-implemented the SLPI Three Rater Team Procedure. With this re-implementation of the SLPI Three Rater Team Procedure a study was initiated to investigate the reliability and validity of the SLPI Three Rater Team Procedure as conducted at NTID. Based on an analysis of 160 SLPIs conducted at NTID from the AY03-04 Fall Quarter through the AY05-06 Fall Quarter the following was reported:

The results of the reliability analyses show that for first independent ratings greater than 86% of raters produced ratings that were within one rating level of the other members of their rating teams. When all three rating team members' first independent ratings were not within one rating level of one another, adding a second independent rating prior to discussion boosted the reliability of ratings within one rating level to greater than 96%. These results confirm that the reliability for NTID SLPI raters' independent ratings prior to negotiated discussion, review, and determination of interviewees' official ratings, was very high.

The results of the validity analysis for this study indicate relatively strong validity for raters' pre-negotiated independent ratings. Only 2.5% of first independent ratings fell outside one level of the official ratings. Adding a second independent rating prior to rater team discussion when the first independent ratings of the three raters were not all within one level of one another reduced this percentage to 0.6%. These results suggest that the

influence of idiosyncratic construct-irrelevant variance on construct validity prior to negotiation is minor, and that the negotiation process is effective in reducing the impact of this variance on SLPI validity in the majority of rater-disagreement cases……

The SLPI evaluation protocol was designed to converge toward increasingly reliable and valid official ratings through the use of multiple stages of independent ratings and negotiated discussion and review of interviewees' interviews in relation to a standardized scaled linguistic performance profile grounded in a theory of language use and learning. The NTID reliability and validity data reported in this paper provide clear evidence that the SLPI evaluation protocol successfully accomplishes that goal….

[Caccamise, F., & Samar, V. (in publication). Sign Language Proficiency Interview (SLPI) pre-negotiation inter-rater reliability and rater validity. *Contemporary Issues in Communication Science and Disorders (CICSD)*; projected publication date is Spring 2009]

Consistency <u>Across</u> SLPI Rating Teams of Three

In 1998 F. Caccamise and W. Newell provided a progress report for a study that was designed to investigate the third question posed by Caccamise et al. (1996); that is, reliability of SLPI results across SLPI Rating Teams of three raters. Results showed that of five SLPI interviews rated by three NTID SLPI Rating Teams, all three teams had the same ratings for four interviews, with two teams having the same rating and one team a rating one level lower for the fifth interview. In addition, of four interviews that were rated by two teams, both teams had the same ratings for three interviews, with two teams differing by one level for the fourth interview. These tentative (due to small number) results supported that SLPI results across NTID SLPI Rating Teams of three are consistent. From AY98-99 through AY00-01 the demand for NTID SLPI services resulted in no additional data being collected for this study. As stated earlier, NTID used the SLPI Individual Rater Procedure during AY01-02 and AY02-03. Given the return by NTID to the SLPI Three Rater Team Procedure in AY03-04, as resources allow further study of consistency across SLPI Rating Teams of Three Raters will be conducted.

Reliability Research – SLPI Individual Rater Procedure

In addition to the traditional SLPI Three Rater Team Procedure, in 1998 we began to offer training for an SLPI Individual Rater Procedure and in 2003 we began offering training for an SLPI Two Rater Team Procedure. A brief description of these procedures is provided in SLPI PAPER #9, *Options for Conducting and Sharing Results of SLPI Ratings*

Several programs use or have used the Individual Rater Procedure, and three of these programs, Florida, South Carolina, and Utah Schools for the Deaf and the Blind (FSDB, SCSDB, and USDB), in consultation with the author of this paper have monitored consistency of SLPI Team Members' ratings during the past several years. Also, as stated earlier, use of the SLPI Individual Rater Procedures was initiated at NTID in the Fall of 2001.

Barbara Curtis, FSDB SLPI Coordinator, and Claire Bailey, SCSDB SCPI Coordinator, reported satisfaction with the consistency of ratings obtained via the SLPI Individual Rater Procedures. Both Barbara and Claire, however, stated that they had become dissatisfied with the support for

ratings provided on SLPI Individual Rater Worksheets. Therefore, during AY05-06 both FSDB and SCSDB changed to the SLPI Two Rater Team Procedure and both have indicated this has resulted in improved support for interview ratings on rater worksheets.

NTID, however, had a different experience. As reported in the *NTID Faculty/Staff Communication Research Group (F/SCRG) Final Report* (F. Caccamise, 2005) by S. Schley and F. Caccamise, during AY02-03 an analysis of SLPI research ratings for 11 interviewees with official ratings of Intermediate, Intermediate Plus, and Advanced supported that ratings obtained via the SLPI Individual Rater Procedure were inconsistent across NTID SLPI raters. Considering this analysis of results for the SLPI Individual Rater Procedure, the importance of the Intermediate-to-Advanced SLPI rating range for NTID faculty/staff, and the results of the SLPI Three Rater Team Procedure reliability study for 1993-through-1996 NTID SLPI ratings, it was recommended that NTID return to use of the SLPI Three Rater Team Procedures. As stated earlier, during AY03-04 the SLPI Three Rater Team Procedure was re-implemented at NTID.

Jean Thomas, the USDB SLPI Coordinator, like Barbara and Claire, has reported satisfaction with the consistency of ratings obtained via the SLPI Individual Rater Procedure. In addition, Jean reported satisfaction with the support for ratings provided on SLPI Individual Rater Worksheets. In discussing this with Jean, it was learned that USDB schedules raters to come to a single location at the same time to conduct their ratings. Following their independent ratings this allows raters (1) to discuss both ratings they agree on and ratings they disagree on, and (2) to have immediate discussions to assist in resolving rating differences. Also, as appropriate, the SLPI Coordinator discusses ratings and support for ratings with raters. According to Jean, this procedure serves to maintain a sense of team among USDB SLPI Team Members and serves as an on-going form of SLPI Team member in-service training. The SLPI Individual Rater Procedure as implemented at USDB may well explain why USDB has experienced success with this procedure.

In addition, the SLPI Individual Rater Procedure may be expected to work most successfully when all or most ratings are reported as ranges rather than specific ratings (see SLPI PAPER #16. *Options for Reporting SLPI Ratings*).

<center>Conclusion</center>

Monitoring the consistency of SLPI Team Members' ratings is important to helping ensure the fairness of the SLPI process for persons taking the SLPI and for planning and conducting appropriate in-service training for SLPI Team Members. Hopefully the information in this document has provided you with information that is helpful to you in ensuring the fairness of ratings conducted by your SLPI Team and to help you in planning in-service training on an individual and group basis for your SLPI Team Members.