

# Validity of Using Short Segments for Analyzing Speech Pathology Supervision Conferences with the MOSAICS

P. Hagler and R. Fahey

## Abstract

Short-segment samples of entire supervisor/student conferences have been shown to be valid when analyzed with the Underwood Category System (Underwood, 1973) and McCrea's Adapted System (McCrea, 1980). Neither interaction analysis system recognized the potential for equal contribution by both supervisor and supervisee in a conferencing situation. This study sampled conference interaction during the beginning five minutes, the last five minutes, and five minutes randomly selected from the middle of speech-language pathology supervision conferences using the MOSAICS system of interaction analysis (Smith, 1978). The MOSAICS system allows coding of the same behaviors for both conference participants. Frequency data from each sampling procedure were compared to one another and to data from the entire conference. Results suggested that five-minute segments were, to a large extent, valid samples of supervision conferences analyzed with the MOSAICS.

## Introduction

Use of short-interval segments for analysis of clinician/client interaction has been a common practice for many years in speech pathology. Their use has been based primarily on the findings of Schubert and Laird (1975), Boone and Prescott (1972) and Brookshire, Nicholas and Krueger (1978). Establishment of the validity of short-segment samples of clinician/client interaction has facilitated the self-monitoring practices of students as well as speech and hearing professionals. It should also contribute to the supervision and research processes where they pertain to treatment activities. The short-segment sampling strategy suggested by Boone and Prescott (1972) for clinician/client interaction analysis was used by Culatta and Seltzer (1976) in the application of their 12-category, supervisor/clinician interaction analysis system. Randomly selected 5-minute segments of supervision conferences were used, but comparisons of the 5-minute segments with whole conference data were not made.

Two validity studies for short-segment sampling have been done with data from speech pathology supervision conferences. Casey (1980) used McCrea's Adapted System (McCrea, 1980) to investigate the validity of five-minute samples of speech and language super-

vision conferences, as representative of events of the entire conference. Casey (1980) collected samples using the following procedures: 1) the beginning five minutes of the conference; 2) the last five minutes of the conference; 3) a random five-minute segment selected from the middle of the conference; and 4) two random 2 1/2 minute segments from the middle of the conference. Results indicated that there were no significant differences between the mean scores obtained from the four sampling procedures and those obtained from analysis of the entire conference. It was also found that none of the sampling procedures could be considered more representative than any other. However, Casey (1980) recommended caution in use of five-minute segments when analyzing two subcategories of conference behavior, self-exploration and empathic understanding, because of their low frequency of occurrence. It was also recommended that, if two 2 1/2-minute samples are selected from the middle of a conference, they should be non-overlapping.

Underwood (1973) also reported validity of short-segment sampling from supervisor/clinician conferences, using a 15-category system which described ten supervisor behaviors and four clinician behaviors. A randomly selected five-minute segment was found to be representative of the entire conference. The rationale for categorizing more than twice as many supervisor behaviors than clinician behaviors was that supervisor behavior was believed to set the emotional tone and communicative atmosphere for the conference. While this belief may have been logical, there were no data to suggest that a supervisee contributed less than a supervisor to the nature of the relationship. The Underwood (1973) system, by not listing each category for both supervisor and supervisee, precluded the coding of some behaviors that could be exhibited by either participant. Similarly, McCrea's Adapted System (McCrea, 1980) was based on scales which originated in the mental health professions. They categorized clinician (supervisor) behaviors that were thought to relate to client (supervisee) self-exploration in psychotherapy. While both studies provided validity data for short-segment sampling procedures, neither interaction analysis system recognized the potential for equal contribution by both supervisor and supervisee during conferencing. Coding of different behaviors on the part of supervisors and supervisees has been a common denominator for short-segment validity studies of interaction analysis to date. This study attempted to establish validity of short-segment samples using an interaction analysis system that recognized the potential for equal contributions by both supervisor and supervisee. The analysis system enabled examination of

---

Paul Hagler,  
University of Alberta, Edmonton

Ronald Fahey,  
Phonic Ear Ltd., Mississauga.

---

---

several more interactive parameters than the Culatta and Seltzer (1976) system and included content and teaching cycle data, as well.

The interaction analysis system which allowed recording of the same behaviors by both participants was the Multi-dimensional Observational System for Analysis of Interactions in Clinical Supervision (MOSAICS) by Richard Weller (1971). From Weller's original analysis system, four ratios derived from coded pedagogical moves and substantive-logical utterances were identified by Smith (1978) as useful in analysis of supervision conferences in speech pathology. Twenty-one teaching cycles were specified. These represented the various sequences of discourse moves made by both participants in response to a structuring or soliciting move by the supervisor. Smith's adapted version was reported by Smith and Anderson (1982). Tradition reliability and validity analyses were done by Weller (1971), during the development of the multidimensional observational system, and again by Smith (1978), when the system was adapted for use in speech language pathology. In both cases, validity measures were based on the coding of whole conferences. Pedagogical moves by each participant were coded for the MOSAICS system as: structuring (STR), soliciting (SOL), responding (RES), reacting (REA), or summary reacting (RSM). Each move was also coded according to its substantive-logical characteristics. Moves related to the use of language were coded as: defining (DEF) or interpreting (INT). Those related to the diagnostic process were coded as: fact stating (FAC), explaining (XPL), evaluation (EVL), or justification (JUS). Prescriptive process moves were coded as: suggestions (SUG), explanations of suggestions (SGX), opinions (OPN), or justification for opinions (OPJ). Ratios and teaching cycles were calculated from frequency data on these coded events. Neither Weller (1971) nor Smith (1978) established the validity of short-segment sampling. Reliable coding of supervision conferences required that each conference tape be played back a minimum of three times, with different, specified coding decisions made on each occasion. With conferences often one-half hour or longer in length, it became apparent that a valid short-segment sampling procedure would be of value, both as a self-improvement device for busy supervisors and as a research tool when large numbers of conferences must be analyzed.

This study was designed to determine whether short-segment samples of speech language pathology supervision conferences were representative of entire conferences, when coded with the MOSAICS, an interaction analysis system accounting for the potential for equal contribution by both participants.

## **METHOD**

### **Participants**

Four experienced supervisors and seven undergraduate student clinicians made up seven subject pairs used in this study. Three supervisors were members of the Canadian Speech and Hearing Association. The

fourth supervisor was eligible for membership. All had previous supervision experience with two or more students. Student participants were undergraduates enrolled full-time in the Department of Speech Pathology and Audiology at the University of Alberta. They participated as part of a regular clinical practicum assignment. Five students had obtained fewer than 150 clinical practicum hours. The remaining students were nearing the end of their clinical practicum training. Subject pairs were selected upon their affirmative response to a request to video tape a conference. A few supervisors were unable or unwilling to have conferences taped. Subjects were told that the study pertained to validity of sampling procedures. Mostly they were told what the study was not; that is, it was not an evaluation of their supervisory behavior or style. Subjects' limited awareness of the purpose of this study was required by an ethical review committee, but was not thought to constitute a threat to internal validity. It seemed unlikely that such general information could have been the basis for subjects effecting measurable changes in their conferencing behavior when the measuring tool, MOSAICS, was unfamiliar to them.

### **Procedure**

One entire conference with each subject pair was video-taped. Conferences were at least 20 minutes in length with some conferences lasting forty minutes. Mean conference length was approximately thirty-five minutes. Complete written transcripts were made following practised conventions adapted from those used for transcription of child language recordings (Bloom & Lahey, 1978). In all cases, conferences were part of a regular conferencing schedule.

Each conference was coded beginning to end from the written transcript using primary content areas identified by Smith and Anderson (1982). Interjudge and intrajudge reliability were established. Point-to-point agreement was measured in all cases. The principal researcher, experienced in the use of MOSAICS, trained one research assistant for coding. Interjudge reliability for delineation and correct coding of pedagogical moves was 82 percent. This was established before data collection began. The research assistant collected all data. Intrajudge reliability for the research assistant was also determined. Random five-minute segments from 3 randomly selected written transcripts were coded a second time after a two-week interval. Percentages of point-to-point agreement were determined for speaker, moves, and substantive-logical categories, and the mean percentage for each transcription was calculated. The mean percent intrajudge agreement for the three coded samples was 94%; with 100% agreement on speaker, 96% agreement on moves, and 86% agreement on substantive-logical categories.

Segments of complete conferences were also coded when sampled as follows: a) five minutes from the beginning of each conference; b) five minutes from the end of each conference; and c) five minutes randomly selected from the middle of each conference.

## Analysis

Frequency data from all sampling procedures as well as entire conference data were converted into percentages and compared for significant differences using two-factor mixed ANOVAs with repeated measures on one-factor (Bruning & Kintz, 1977). Four ratios found useful by Smith (1978) in analyzing participants' use of conference time were calculated for both participants under each sampling condition. These became the dependent variables. Participants, supervisor or clinician, became the between-groups factor. Sampling method or segments, that is whole conference, first five minutes, middle five minutes, and last five minutes, became the within-subjects or repeated measures factor. The following ratios were tested: a) analytical/evaluative (A/E); b) initiatory/reflective (I/R); c) diagnostic/prescriptive (D/P); d) complex/simple (C/S). The ratios were computed with the following formulae<sup>1</sup>:

- 1)  $A/E = (FAC+XPL+SUG+SGX)/(EVL+JUS+OPN+OPJ)$
- 2)  $I/R = (STR+SOL)/(RES+REA+RSM)$
- 3)  $D/P = (FAC+XPL+EVL+JUS)/(SUG+SGX+OPN+OPJ)$
- 4)  $C/S = (XPL+JUS+SGX+OPJ)/(FAC+EVL+SUG+OPN)$

The frequency of occurrence of each teaching cycle was determined for complete conferences and for the three five-minute sampling procedures described above. These data, expressed as percentages, were compared for significant differences using a one-factor ANOVA for a repeated measures design (Glass and Stanley, 1970). One analysis was run on data representing each of the six teaching cycles.

## Results

Analyses of variance on three of the four ratios, Tables 1 - 4, revealed no significant main effect for sampling method (segments). There were no interaction effects for participants by sampling method (segments). The only significant effect was for sampling method for the C/S ratio ( $F=9.667$ ,  $p<.001$  with 3 and 36 df). A Tukey-T for multiple comparisons post hoc, shown in Table 5, suggested that the C/S ratios determined from samples taken in the middle five minutes of supervisory conferences were significantly different from those taken in the last five minutes. None of the short-segment samples were significantly different from the whole conference. There was a significant main effect for the participants factor in the analysis of variance for the I/R ratio ( $F=13.533$ ,  $p<.005$  with 1 and 12 df).

Of the 21 teaching cycles found useful by Smith (1978), only six were present in numbers sufficient to justify analysis. Those were: (1) STR, (3) STR REA, (4) STR REA REA..., (14) SOL RES, (18) SOL RES REA, and (19) SOL RES REA REA... Three dots indicated that a teaching cycle continued by repeating the last two moves. Three cycles not listed among those used by Smith (1978) occurred in samples for this study. These, along with the remaining 15 teaching cycles, were not

present in two or more of the samples from all subjects. This low frequency of occurrence was judged to preclude statistical analysis of those particular teaching cycles.

Analyses of variance on the six teaching cycles shown in Table 6 revealed no significant differences in frequency of occurrence for any sampling method.

**Table 1:** Two factor mixed analysis of variance on analytical/evaluative (A/E) ratio data

Source	SS	df	ms	F	p
Total	1.402	55			
Between S's	.721	13			
Participants	.036	1	.036	.632	n.s.
Error <sub>b</sub>	.685	12	.057		
Within S's	.681	42			
Segments	.077	3	.026	1.733	n.s.
Segments x Participants	.054	3	.018	1.200	n.s.
Error <sub>w</sub>	.550	36	.015		

**Table 2:** Two-factor mixed analysis of variance on initiatory/reflexive (I/R) ratio data

Source	SS	df	ms	F	p
Total	1.112	55			
Between S's	.765	13			
Participants	.406	1	.406	13.533	<.005
Error <sub>b</sub>	.359	12	.030		
Within S's	.347	42			
Segments	.017	3	.006	.750	n.s.
Segments x Participants	.036	3	.012	1.500	n.s.
Error <sub>w</sub>	.294	36	.008		

**Table 3:** Two-factor mixed analysis of variance on diagnostic/prescriptive (D/P) ratio data

Source	SS	df	ms	F	p
Total	1.335	55			
Between S's	.809	13			
Participants	.040	1	.040	.625	n.s.
Error <sub>b</sub>	.769	12	.064		
Within S's	.526	42			
Segments	.051	3	.017	1.308	n.s.
Segments x Participants	.019	3	.006	.462	n.s.
Error <sub>w</sub>	.456	36	.013		

**Table 4:** Two-factor mixed analysis of variance on complex/simple (C/S) ration data

Source	SS	df	ms	F	p
Total	.442	55			
Between S's	.232	13			
Participants	.002	1	.002	.105	n.s.
Error <sub>b</sub>	.230	12	.019		
Within S's	.190	42			
Segments	.086	3	.029	9.667	<.001
Segments x Participants	.002	3	.001	.333	n.s.
Error <sub>w</sub>	.102	36	.003		

**Table 5:** Tukey-T for Multiple Comparisons Post-hoc Confidence intervals around random factor (Samples) from C/S ratio ANOVA where 1=1st 5 minutes, 2=middle 5 minutes, 3=last 5 minutes, and 4=whole conference.

	Upper	Lower
$\bar{X}_1$ w/ $\bar{X}_2$	.45	-12.01
$\bar{X}_1$ w/ $\bar{X}_3$	11.45	-1.01
$\bar{X}_1$ w/ $\bar{X}_4$	6.66	-5.80
$\bar{X}_2$ w/ $\bar{X}_3$	17.23	4.77
$\bar{X}_2$ w/ $\bar{X}_4$	12.44	-.02
$\bar{X}_3$ w/ $\bar{X}_4$	1.44	-11.02

NB: Confidence intervals including zero between their bounds show non-significant differences.

## Discussion

Results of this study suggested that five-minute samples of supervision conferences in speech pathology analyzed with the MOSAICS are valid representations of events of the entire conference from which the samples were taken. There was a significant difference between clinicians' and supervisors' frequency of use of the initiatory/reflexive ratio with supervisors initiating such sequences significantly more often. Of course, such a finding was not surprising and did not bear on the validity of short-segment samples. There were no significant differences between each of 3 short-segment sampling procedures and the entire conference, in terms of any of the four ratios nor in terms of teaching cycles. This suggested that short-segment samples from the beginning five minutes, last five minutes, and a random five-minute segment from the middle are valid representations of entire speech language pathology supervision conferences analyzed with the MOSAICS. These findings must be applied with caution. This study was based on a small number of subjects, who may not be representative of supervisor-student pairs across the profession. While results suggested that five-minute samples were valid, there remains a possibility of significant differences being discovered in replications of this work with larger numbers of subjects. Additional validity studies of short-segment sampling would certainly be appropriate.

**Table 6:** Results of analysis of variance on six teaching cycles derived from the MOSAICS with four sampling procedures.

Cycles	Source of Variation	df	SS	ms	F	p
1	Total	27	1370			
	Conferences	6	702			
	Samples	3	30	10.1	.29	n.s.
	Error	18	638	35.4		
3	Total	27	1759			
	Conferences	6	1051			
	Samples	3	142	47.2	1.50	n.s.
	Error	18	566	31.5		
4	Total	27	6799			
	Conferences	6	3216			
	Samples	3	369	122.9	.69	n.s.
	Error	18	3214	178.6		
14	Total	27	1362			
	Conferences	6	326			
	Samples	3	153	51.1	1.04	n.s.
	Error	18	883	49.0		
18	Total	27	5115			
	Conferences	6	1516			
	Samples	3	797	265.5	1.71	n.s.
	Error	18	2802	155.7		
19	Total	27	7053			
	Conferences	6	3574			
	Samples	3	117	39.0	.21	n.s.
	Error	18	3362	186.8		

NB: Critical value for F at .05 with 3 and 18 df=3.16

Future studies might include an alternative sampling method such as a series of one-minute samples taken periodically throughout the session, as suggested by Brookshire, Nicholas, and Krueger (1978) in their study of clinical treatment activities.

It was interesting that fifteen teaching cycles did not occur frequently enough to justify statistical analysis. Smith (1978) did not report frequencies of the various teaching cycles, but if some occurred infrequently, five-minute samples might easily render an already infrequent behavior useless for purposes of analysis. Another factor may have been different supervisee clinical experience levels. Supervisees in the Smith (1978) study were mostly graduate students, while those in the current study were all undergraduates. The majority of teaching cycles in the current study were simple exchanges involving three or fewer total moves, sequences which one might expect with relatively inexperienced supervisees. Several of the remaining 21 cycles found by Smith (1978) could be described as more complex pedagogical exchanges having four or more total moves. This observation is conjecture at best, since neither study manipulated level of supervisee experience. An investigation comparing supervisee experience with complexity of teaching cycles would be an interesting contribution to the supervision literature.

---

Hopefully the results of this study will contribute to changes in both the practice and study of supervision. Reliable coding with the MOSAICS necessitates users listening to a taped conference three times, occasionally stopping the tape to catch up on written coding. Therefore, a twenty-minute conference might require one hour or more to code. A five-minute segment can be coded in approximately eighteen minutes. Supervisors can now allocate the minimal time required to code and analyze five minutes of their conferences and, in so doing, learn about themselves, their students, and their conference interactions. Surely the information obtained is well worth the effort.

Researchers may not feel comfortable yet in using the MOSAICS to code short-segment samples from large numbers of subjects. Perhaps replication studies should corroborate these results before short-segments are used for research purposes. Certainly, when that is done, the enormous amount of work and associated costs in this type of supervision research can be substantially reduced.

#### Acknowledgments

The authors wish to thank Kristine Larsen, Karen Klingenberg, and Karen Grauer for their assistance in collecting data for this study.

#### Footnotes

(1)The four ratios were actually calculated with numerator added to denominator prior to division (i.e. — A/E ratio =  $A/(A+E)$ ).

#### References

Bloom, L. & Lahey, M. (1978). *Language development and language disorders*. Toronto: John Wiley and Sons, Inc.

Boone, D. & Prescott, T. (1972). "Content and sequence analysis of speech and hearing therapy". *Asha*, 14, 58-62.

Brookshire, R.; Nicholas, L; & Krueger (1978). Sampling of speech pathology treatment activities: an evaluation of momentary and interval sampling procedures, *Journal of Speech and Hearing Research*, 21, 4, 652-667.

Bruning, J. & Kintz, B. (1977). *Computational handbook of statistics*. Glenview, Illinois: Scott, Foresman and Company.

Casey, P. (1980). The validity of using small segments for analyzing supervisory conferences with McCrea's adapted system. Doctoral Dissertation, Indiana University.

Culatta, R. & Seltzer, H. (1976). Content and sequence analysis of the supervisory session, *Asha*, 18, 8-12.

Glass, G. & Stanley, J. (1970). *Statistical methods in education and psychology*. Englewood Cliffs, New Jersey: Prentice Hall, Inc.

McCrea, E. (1980). The relationship between supervisor offered conditions of empathic understanding, respect, facilitative genuineness and concreteness upon supervisee self-exploration in individual conferences in speech pathology. Doctoral Dissertation, Indiana University.

Schubert, G. & Laird, B. (1975). The length of time necessary to obtain a representative sample of client-clinician interaction, *Journal of the National Student Speech and Hearing Association*, 3, 26-33.

Smith, K. (1978). Identification of perceived effectiveness components in the individual supervisory conference in speech pathology and on evaluation of the relationship between ratings and content in the conferences (Doctoral Dissertation, Indiana University, 1977). *Dissertation Abstracts International*, 39, 680B.

Smith, K. & Anderson, J. (1982). The relationship of perceived effectiveness to verbal interaction/content variables in supervisory conferences in speech language pathology. *Journal of Speech and Hearing Research*, 25, 252-261.

Underwood, J. (1973). Interaction analysis between the supervisor and the speech and hearing clinician (Doctoral Dissertation, University of Denver, 1973). *Dissertation Abstracts International*, 34, 2995B.

Weller, R. (1971). *Verbal-communication in instructional supervision*. New York: Teachers College Press, Columbia University.