INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Université d'Ottawa · University of Ottawa

Impact of Descriptive Versus Evaluative Constructive Feedback on Public Speakers' Performance Self-efficacy

John J. Donohue

Faculty of Education

University of Ottawa

A thesis submitted in partial completion of the requirements of the degree of

Doctor of Philosophy in Education to the Faculty of Graduate and

Postdoctoral Studies at the University of Ottawa

ABSTRACT

The main study consisted of a randomized treatment comparison group design with pre- and posttest intervals. The purpose was to assess the differential impact of two different forms of constructive feedback – descriptive and evaluative – on participants' performance self-efficacy within a growth-oriented appraisal context. It was hypothesized that descriptive constructive feedback would lead to higher observed growth in performance self-efficacy. The results of the main study revealed that although all participants showed evidence of significant improvement from having participated in the study, there was no differential effect between participants in different treatment conditions. Post-hoc analysis of qualitative data revealed contamination between the treatment conditions suggesting that the main effect for feedback condition was confounded. The results are discussed in terms of the methodological challenges facing researchers interested in testing the hypothesis that descriptive feedback is superior to evaluative feedback in enhancing performance self-efficacy. The failure of the randomized treatment group design to yield valid results is framed as a basis for considering developing methodologies in this area.

Dedication

This thesis is dedicated to the memory of my maternal grandfather,

who personified integrity and perseverance.

Leon D. Choffin ("Pops")

1901 – 2000

Acknowledgements

As with all great achievements in life, this Doctoral degree would not have been completed without the support and assistance of many people.

First and foremost, I would like to thank my wife, Deanne, for her unwavering support and confidence in me. The arrival of my daughter, Maggie, provided the final impetus to finish this document.

Thanks to my parents who provided me with limitless support throughout the entire process. My siblings were also there for me when I needed them.

My thesis supervisor, Dr. J. Bradley Cousins, endured my idiosyncrasies for the past seven years. Throughout our time together, Brad personified the type of knowledgeable, collegial mentor that I hope I can be for my future students.

The development of both this research project and my own understanding of educational research have been considerably enhanced by the members of my thesis committee. Professors Pierre Trudel, Edward Drodge, Sergio Piccinin and Edward Hickcox all provided valuable guidance.

Numerous friends also provided motivation and support for me throughout the duration. Prominent among these are Phil Nguyen, Linda Michelin, Douglas Pelchat, Kurt Clausen, Dale Petruka and Greg Michalski. Some of you were in the same boat; others helped paddle through the rough waters; all of you are important to me.

This research could not have been conducted without the enthusiastic participation of members of the Canadian Association of Professional Speakers and Toastmasters International from across Canada. A sincere thank you is due them, also.

TABLE OF CONTENTS

List of Tables

Chapter 3

Chapter 4

List of Figures

Chapter 1 – Introduction & Review of Basic Literature

1.0 Statement Of The Problem

The provision of constructive feedback concerning one's performance with the expressed intent of improving that performance has become the focus of increased interest in educational, psychological and organizational research in recent years. Yet much remains to be understood about this relatively complex, multidimensional phenomenon. Much of the research to date has been designed to ascertain the effects of certain attributes or aspects of constructive feedback. There exists, however, considerable variation in the isolation and identification for study of attributes of constructive feedback and the conceptualization of the effects such feedback is intended to have. In the present study, the effects of the evaluative content of constructive feedback on one's sense of self-efficacy were examined.

The motivation for the present study comes from the observation that most formative appraisal currently being done is benign and a waste of time (Lawton, Hickcox, Leithwood & Mussella, 1986). This is true for both the appraisee and the supervisor. The intent of the present study is to systematically examine one approach to making formative appraisal effective.

1.1 Definition of Terms

In the present study, the focus will be on public speakers' growth in performance self-efficacy. The term *public speaker* refers to individuals who are engaged in the process of performing presentations before audiences. Both professional speakers (members of the Canadian Association of Professional Speakers) and amateur speakers (members of Toastmasters International) were participants in the research.

Growth-oriented appraisal refers to forms of performance appraisal that fall within the context of formative appraisal. This type of appraisal was consistent with clinical supervision

models, detailed by Cogan (1973), Goldhammer (1969) and others (see Cousins, 1995). Growth-oriented appraisal is distinct in purpose from judgmental or evaluative appraisal in that it seeks to improve performance rather than label it as good or bad.

Self-efficacy is the belief in one's ability to perform certain behaviors (Bandura, 1977). This represented the dependent variable of interest in the present study. Specifically, the study examined the self-efficacy of public speakers vis-à-vis ability to perform within a presentation context. This is called *performance self-efficacy* in the current study.

In the performance appraisal context, the term *feedback* is typically defined as information about current or past performance fed back to an individual (appraisee) by someone else (supervisor, coach, peer) (Ilgen, Fisher & Taylor, 1979). Feedback provided with the intent of improving performance can be labeled *constructive* feedback (Baron, 1990).

A second dimension of feedback, and the one of central interest to this study, is the *evaluative content* of the feedback. Evaluative content of feedback refers to its location along a continuum ranging from rich description of performance (low evaluative content) through highly analytical and judgmental feedback (high evaluative content). It is assumed that the intention of feedback (constructive vs. non-constructive) and the extent to which feedback consists of evaluative content (descriptive vs. evaluative) can vary independently of one another.

1.2 Overview of the Study

The current research was restricted to the context of growth-oriented appraisal (i.e., performance appraisal with the intent of improving one's performance). An important question was whether constructive feedback would be more potent if highly descriptive as opposed to highly evaluative in content. Recent theoretical developments in the supervision literature suggest that under specific circumstances descriptive feedback is more likely than judgmental

input to improve one's performance. But despite calls to address this issue (e.g., Tracy & McNaughton, 1989) there has been little empirical research to date.

In order to address this issue, a two-phase research project was undertaken. The first study, the pre-study, included a pilot study to establish the validity of an instrument custom-designed to assess public speaking performance self-efficacy, the Speaker Self-efficacy Rating Scale (SSRS). The purpose of the pre-study was to assess the reliability of the SSRS for use in the main study and to establish norms for the instrument. The main study used a comparative group design to investigate directly the impact of evaluative content of constructive feedback on self-efficacy in growth-oriented appraisal with public speakers.

First, a comprehensive review of the literature directly related to the hypothesis is presented. Next, in order to situate the problem, we will examine the relevant research within the fields of education, psychology and business management. Chapter 2 presents this integrated review of the literature in cognate fields. In Chapters 3 and 4, the methodologies and results of the studies are explicated. Finally, the meaning of the results in the context of the reviewed literature is examined in Chapter 5.

1.3 Significance of the Research

The study seeks to add to the academic knowledge base in a domain where there is a paucity of empirical research. Although much is written about constructive feedback (as we shall see in the literature review) very little systematic research exists. Most of the contributions have been theoretical or have been anecdotal reports on professional practice. The experimental test of a hypothesis concerning evaluative content will be the first of its kind and will therefore be ground breaking in this area. While this is the central focus for the present thesis, the pre-study instrument development and validation activities will also represent a significant contribution to

a literature with few empirical bases. The literature review, which integrates empirical and theoretical knowledge from the domains of education, psychology and business management, also serves as a contribution to our understanding of constructive feedback.

1.4 Review of Basic Literature

The current study explored the impact of descriptive and evaluative forms of constructive feedback on the performance self-efficacy of public speakers. In this review of the basic literature, we will examine the constructs of constructive feedback and self-efficacy. We have several objectives. First, to survey what is known about constructive feedback and self-efficacy, and to integrate the literature from education, psychology and management. Second, to provide some definition to the concepts that make up these constructs. Third, to critique the methods used in empirical work in these areas. Finally, to highlight the gaps in our knowledge, which provide the rationale for the research study that follows. A review of the literature in cognate fields, which help to motivate the study, is presented in Chapter 2.

1.4.1 Constructive Feedback

Constructive feedback, defined by Ilgen, Fisher and Taylor (1979) as *information that is fed back to the appraisee in order to improve performance*, is the essence of the growth-oriented appraisal process. But just how important is feedback to improving performance?

Feedback is critical to any effort to maintain performance, let alone efforts to improve performance. According to DeGregorio and Fisher (1988) in their theoretical piece on alternate methods of providing performance feedback in the field of business management "Research has shown that performance feedback is necessary in order to maintain and / or improve job performance" (p. 605). They also note that traditional forms of supervisory feedback have proved to be less than adequate for developmental purposes.

Larson (1989) also agrees that feedback is essential to performance improvement initiatives. In his theoretical piece on feedback seeking behavior, he states that, "The importance of feedback as a tool for enhancing performance in organizations can hardly be overestimated" (p. 408).

Ilgen and Moore (1987) agree that the cumulative evidence indicates that performance feedback is important for improving performance: "Both theory and practice strongly advocate the availability of specific and timely feedback for enhancing task performance" (p. 401). Ilgen et al. (1979) write about the critical importance of performance feedback. "Feedback about the effectiveness of an individual's behavior has long been recognized as essential for learning and for motivation in performance-oriented organizations" (p. 349). In this conceptual piece that summarizes the research-to-date in the field of psychology on performance feedback, they divide the influences on the effectiveness of feedback into three parts: the source of the feedback, the message, and the recipient.

Of these three factors, the source of the feedback can be the most important. The source of the feedback can even overshadow the effects of the feedback message itself. In the words of Ilgen et al.: "It is often difficult to separate the effects of the feedback from the effects of the source" (1979, p. 350). Here, they are talking about the fact that the messenger may be more important than the message.

As far as the message, or content of the feedback, is concerned, Ilgen et al. (1979) emphasize that the information being fed back to the performer must be meaningful to the performer. There are two main purposes to providing feedback. The first is to direct behavior (e.g., formative appraisal). The second is to provide reinforcement, either punishment or reward,

in a goal-oriented setting. This reinforcement function could be viewed as an incentive for the performers to change their behavior.

Factors related to the recipient of the feedback include the receiver's perception of the feedback, acceptance of the feedback, desire to respond to the feedback, and intended response to the feedback. In general, performers tend to accept as valid feedback from sources that are close to them either spatially or organizationally (e.g., coworkers performing similar tasks, or immediate supervisors or subordinates) more readily than sources that are more distant.

These general comments provide a broad description of the concept of constructive feedback. But what does constructive feedback look like? How may it be described?

1.4.1.1 Process and content issues in constructive feedback

In their guide for school districts developing appraisal systems, Duke and Stiggins (1986) provide a useful distinction between process and content components of feedback that can have an impact on the growth-oriented appraisal process. Process components deal with how the feedback is presented. Examples of process components are the amount of feedback provided and the timing of when the feedback is delivered. Content components deal with what is presented as feedback. Examples of content components are the specificity of the feedback and the extent to which the feedback is descriptive as opposed to evaluative.

These process and content components are not defined so as to be completely independent; there may be substantial overlap between them. For example descriptive feedback delivered in a forceful way can take on distinct evaluative characteristics. This is due to the nature of the construct of interest: feedback. Since there are contextual factors that may impact on either or both the supervisor and the appraisee, it does not seem appropriate to separate these

factors completely. This inability to completely separate the components of feedback does not diminish the usefulness of the distinction for illustrative purposes.

The psychological literature differs somewhat in its assessment of the construct of feedback. One novel area that has been deemed worthy of investigation is feedback-seeking behavior. Larson (1989) bases his theoretical piece about this aspect of feedback on three assumptions about business employee's feedback seeking behavior: 1) employees are motivated to seek out feedback; 2) employees receive feedback through two methods (monitoring the environment and asking for feedback); and 3) that supervisors are reluctant to give negative feedback and will resist giving it.

The essence of the argument put forth is that feedback inquiry is often oriented toward self-verification, or the maintenance of a sense of value to the organization. When circumstances challenge the self-image of the workers, they seek feedback from supervisors that re-confirms their worth to the organization. Larson describes two methods that employee's use to reduce the likelihood of negative feedback: they ask for feedback when negative feedback is least likely to be given (e.g., monitor the mood of the supervisor, ensure that other problems don't compound the problem) and they structure the form of the inquiry so as to elicit positive feedback (e.g., acknowledge weaknesses to disarm the supervisor's identification of these weaknesses).

Larson's (1989) work does provide some interesting theoretical views of feedback-seeking behavior, and his arguments seem valid. Unfortunately no recommendations for practice are provided. While we may gain some understanding into the feedback-seeking behavior of employees, without empirical evidence to back up the assertions, there is little of practical value for the supervisor.

While a burgeoning empirical literature is beginning to contribute to knowledge about process and content components, the focus of the present study is on the concept of the evaluative content of constructive feedback. In the theoretical literature much has been written about this concept. For example, Showers and Joyce (1996) state that their research in peer coaching has shown that, optimally, feedback is to be free of evaluative content. Some would argue, however, that the mere selection of episodes to describe constitutes evaluation. Despite theoretical discourse and rhetorical opinion, little empirical evidence has been generated regarding this concept.

The scholarly literature concerning this concept is reviewed below. First, however, by way of providing context, a review of several related process and content components identified by Duke and Stiggins (1986) is provided. Specifically, two process components and two content components that have surfaced repeatedly in the growth-oriented appraisal literature are examined. The process components were the amount of feedback and the timing of the feedback. The content components identified were the specificity of the feedback and the evaluative content of the feedback. The following review will focus on these components of feedback.

1.4.1.1.1 Amount of feedback

There is considerable debate as to how much feedback appraisee's should receive. Natriello (1984), in an empirical piece on teacher's perceptions of appraisal effectiveness and frequency, reported that teachers prefer more feedback of any kind than what they were currently receiving. They did not care whether the feedback was positive, negative or neutral; they only wanted some form of recognition from the supervisor.

Sullivan and Wircenski (1988) also advocate that teachers desire more feedback on their performance than they receive. In their review article on the effective clinical supervision of

teachers, they report that: "Teachers welcome professional suggestion from their supervisors about improving their teaching but they rarely receive them" (p.34). The authors also mention that some states (in the United States) have even passed legislation to ensure that teachers receive feedback concerning their work. While this legislation does not necessarily refer to formative appraisal, it is encouraging to see that these states deem performance feedback as worthy of inclusion. On the other hand, it is discouraging that they feel they must enshrine the provision of feedback in legislation.

It is not enough just to give sporadic feedback to performers. Burke and Fessler (1983), in their conceptual piece on collaborative approaches to supervision, recommend ongoing feedback within a continuous process. Blair (1991) shares this view, reporting that teachers are most concerned about two things. Their first concern is the infrequency of supervisory visits for observation purposes. Their second concern is the unavailability of people with supervisory responsibility to provide assistance in terms of changing behaviors.

On the other side of the argument are those advocating less feedback. In a study with factory workers, Chhokar and Wallin (1984) found that providing feedback once every two weeks was just as effective as providing feedback once a week in achieving the predetermined goal of increased worker safety behavior. Chhokar and Wallin claim that the results of their study provide conclusive evidence refuting the assertion that more feedback is better. The restricted range of their data set, however, comparing once a week to once every two weeks limits the generalizability of their findings.

Ilgen and Moore (1987) studied the effect of giving the appraisees' control over whether they receive feedback. Using 220 undergraduate psychology students as participants, they explored the impact on performance in a proofreading task of three types of feedback: quality of

performance, quantitative and mixed. Some participants were also given the choice of receiving or not receiving feedback, as they wished, at various times throughout the performance. There was also a control group that received no performance feedback. The dependent measures in this study were the quality of the proofreading and the speed with which the tasks were completed.

The results of this 3 x 2 design (type of feedback by choice) revealed that participants who were given control over whether they received feedback or not showed increased speed in completing the tasks, but there was no impact on the quality of the performance. In terms of the type of feedback received, participants who received quantitative feedback performed more quickly, while those who received quality feedback or mixed feedback showed increased quality of performance on the task.

Ilgen and Moore (1987) conclude that when a performance can be evaluated on more than one dimension, it may be useful to provide separate feedback on each dimension. Appraisees can then choose which dimension of feedback to receive and when to receive it, thus minimizing the amount of time taken to receive and process the feedback. This method would also reduce the occurrence of unnecessary feedback on areas the appraisee does not desire feedback.

One area that Ilgen and Moore (1987) seem to have overlooked is the interval nature of their design. It would have been interesting, and perhaps useful, to have more information on the extent to which those participants in the feedback choice condition followed a pattern in their requests for feedback. For example, did they ask for feedback after every second or third trial? Or did they seek feedback only after trials on which their performance was poor?

The issue of feedback on areas that are not of interest to the appraisee is one that has not been examined to a great extent. In the growth-oriented appraisal model, the scope of the

feedback sessions is typically agreed upon before the beginning of the process (in the preconference stage). Ilgen and Moore raise the interesting issue of whether the appraisee should control the type of feedback as well as the content of the feedback, at least in terms of the specific focus.

Generally, it is believed that the more feedback, the better. Darling-Hammond et al. (1983) go one step further, indicating that even negative feedback is preferred to no feedback at all, an observation corroborated empirically by Natriello (1984). In their theoretical piece on the development of teacher evaluation systems, Darling-Hammond et al. note that it is the infrequency of evaluation that bothers teachers the most, not the nature of the feedback that they receive. Essentially, they indicate that teachers want to know more about what their supervisors think of their work, regardless of whether the feedback is positive or not.

In summary, it is interesting to note that the studies that report that more feedback is better tend to be ones that involve participants that perform highly idiosyncratic tasks (e.g., teachers, speakers). Those who seek to restrict the amount of feedback tend to focus on more easily measured tasks (e.g., units produced in a time interval). The dearth of research specifically focusing on this issue limits our ability to draw firm conclusions either way. While the optimum amount of feedback may not have been identified, it seems clear from Natriello (1984) that appraisees in relatively complex performance domains desire more feedback than they are currently receiving.

1.4.1.1.2 Timing of feedback

The timing of the feedback can impact on the effectiveness of the growth-oriented appraisal process. In general, the process is seen as cyclical, as presented by Cousins (1995). The preconference is followed by observation and then a feedback session, which leads into the next

preconference. The process continues in this fashion. There is no hard and fast rule, though, as to how soon after the observation the feedback session should occur, or how long before the next observation session.

There is considerable debate as to when is the best time to deliver the feedback. Some researchers find that feedback is maximally effective when given immediately after the observation, when the performance is still fresh in the mind of the appraisee. Others believe that the best time for feedback is immediately before the next performance. Still others advocate that some time in between is preferable.

Of those that believe that feedback should be delivered immediately after the performance, McLaughlin and Pfeiffer (1988) are the most direct. In their conceptual piece on improving teacher appraisal, they state outright that immediate feedback is the only course to follow, in order to maximize the learning potential of the appraisal process. "Feedback provided immediately after an observation, when the event is fresh in the minds of both the teacher and the evaluator, has maximum learning potential" (p. 126).

In their survey of 109 teachers, Ovando and Harris (1993) provide confirmatory evidence for this position. One of the five recommendations they make for improving feedback conferences in formative evaluation is specifically that feedback sessions take place immediately after the observation session. The other recommendations dealt either with the nature of the relationship between the teacher and the supervisor, or the overall process (e.g., that there should be a preconference to establish the scope of the evaluation). Sullivan and Wircenski (1988), in their review article on effective clinical supervision of teachers, also recommend immediate feedback to appraisees.

Not everyone agrees that immediate feedback is best. Geis (1986), in an article on formative feedback, states that feedback is best given immediately prior to the next performance. Geis' reasoning is that the appraisee has the chance to use the feedback information immediately, without delay.

To even things out, in their review article on feedback to individuals, Ilgen et al. (1979, p. 354) take the middle ground: "if the activities that occur during the feedback delay time do not interfere with the individual's ability to accurately recall the behavior and associate the feedback with it, the length of the time delayed should have no detrimental effect on the feedback perceptions."

The empirical evidence seems to favor immediate feedback, while opinion pieces seem to hold diverse views. The most comprehensive review article, Ilgen et al. (1979), lends credence to the intervening activities. Ilgen et al. do not say that immediate feedback is not good; they say that delaying feedback is not necessarily bad. While there does not seem to be closure on this issue, the preponderance of the evidence seems to favor the immediacy of feedback.

1.4.1.1.3 Specificity of feedback

The specificity of feedback, whether the feedback details specific aspects of performance or tends to take a more theoretical, abstract tone, is an important issue. Virtually every author reporting on this issue agreed that the more specific the feedback can be in terms of detailing behavior, the more effective it is likely to be in fostering growth in a teacher evaluation context (Acheson & Gall, 1987).

Stiggins (1988), in a re-examination of earlier case study research with teachers (Duke & Stiggins, 1986), also found that appraisees favored specific feedback to more general, theoretical feedback. In addition, Stiggins recommends that the feedback be given small amounts at a time,

rather than large amounts of feedback all at once. Note that this is not a comment on the overall amount of feedback, but rather recognition that all appraisee's have a limit to the amount of information that they can process at a time.

Lawton et al. (1986), in their study of appraisal practices for certified educational personnel in Ontario, found that "the most apparent deficiencies among appraisal practices, as a whole, in the case study boards were ... the non-specific and perhaps irrelevant feedback provided as a consequence of some appraisals" (p. 269). Similarly, in a comparable U.S. sample, McLaughlin and Pfeiffer (1988) found that generalities or theoretical abstractions are meaningless to the appraisee.

In the psychological and management literature, the same holds true. Although given only passing mention, the message provided by DeGregorio and Fisher (1988), Latting (1992) and Schneier, Beatty and Baird (1986) is clear: the more specific the feedback, the better in terms of fostering growth.

In sum, the empirical evidence, while somewhat thin, does favour providing highly detailed specific feedback to appraisee's in the growth-oriented appraisal process. No evidence was located to suggest that non-specific feedback would be effective.

1.4.1.1.4 Summary of related issues

While the theoretical literature concerning these three components of feedback (amount, timing, and specificity) seems rich, it fails to provide consistent and clear guidelines. In general, we can conclude that more feedback is better than less, that feedback provided immediately after performance is better than delayed feedback, and that the more specific the content of the feedback, the better. The empirical literature concerning these issues is somewhat thin, and

unfortunately fails to provide consistent results. We will now examine the literature concerning the focus for the proposed research: the evaluative content of the feedback.

1.4.1.2 Evaluative content of feedback

The evaluative content of feedback, or how descriptive or evaluative the feedback is, is the main focus of the current research study. Evaluative content, in this instance, refers to the location along a continuum from strictly descriptive feedback to strictly evaluative feedback. There are a number of issues at play here, not the least of which is the epistemological basis for the argument.

1.4.1.2.1 Epistemological considerations

The question of whether the feedback given to appraisee's should be predominantly descriptive or evaluative opens a Pandora's box of opinion grounded largely in deeply philosophical and epistemological issues. Tracy and McNaughton (1989), in their theoretical piece on clinical supervision, provide a useful framework for understanding this issue. They differentiate between neo-progressive and neo-traditional approaches to growth-oriented appraisal. This differentiation is captured in the form of a continuum in Figure 1.1 and is linked to underlying epistemological considerations.

Hunter (1988a, 1988b) characterizes the essence of the neo-traditional approach to growth-oriented appraisal, at least within the educational milieu. Hunter has long been a proponent of the expert-novice model of research, where the expert supervisor compares the behavior of the novice supervisee to a research-based model of effective teaching. The supervisor provides corrective feedback to the appraisee on how to remedy any discrepancies from the model.

Researchers / Scholars	Tracy & McNaughton (1989)	Cogan (1973) Goldhammer (1969) Duke & Stiggins (1986) Kilbourn (1990)	Hunter (1988a) Hunter (1988b) Minton (1979)
Appraisal Orientation	Neo-Progressive	Collaborative	Neo-traditional
Epistemological Orientation	Interpretivist	Revisionist traditionalist	Traditionalist

Approach to Growth-Oriented Appraisal

Figure 1.1. Continuum of approaches to growth-oriented appraisal process.

Hunter believes that "weakness in teaching cannot be hidden from a competent evaluator" (1988a, p. 49) and that improving teacher performance – and thus student achievement – is a matter of applying a research-based model of effective teaching in which certain "regularities" in performance are understood. This orientation is grounded in logical empiricist or positivistic epistemological assumptions. The individual performer's goals are not part of the process; the 'expert' supervisor is assumed to be knowledgeable and capable of both detecting problems with performance and helping to remedy them. The preconference is a superfluous luxury in Hunter's version of growth-oriented appraisal; the appraisee's underlying intent is not important, only their actual performance. In neo-traditional approaches, the 'expert' supervisor's duty is to judge the performance of the appraisee against the research-based model of effective teaching, and to provide remedial information (feedback) to the 'novice' teacher. The feedback is judgmental by nature, since it focuses on deviations from the model of effective practice (deficiencies).

Minton (1979, as cited in Tracy & McNaughton, 1989) supports the highly structured version of supervision promoted by Hunter. Minton, a student of Hunter's, agrees that the expert-novice structure is the only viable form of assessment. Minton's version of supervision differs only superficially from Hunter's. The rejection of contextual factors, in favor of a research-based model of effective teaching, is again the central theme.

There has been criticism of Hunter's model. McGreal (1988) acknowledges that the effective teaching literature has merit. The problem arises with Hunter's extension of the work and the fact that the relationships identified in the effective teaching literature are correlational, not causal. McGreal concludes that the relationship between the model of effective teaching

proposed by Hunter and the impact on student learning has not been proven, at least to McGreal's satisfaction.

At the other end of the spectrum lies a perspective argued by Tracy and McNaughton (1989), who believe that the idiosyncratic nature of teaching prevents any supervisor from understanding all of the underlying reasoning for the appraisee's performance. Their approach is consistent with an interpretivist epistemology. The growth-oriented appraisal process concerning the appraisee is best served by investing in the appraisee complete control of the process, from the issues to be explored through the methods of obtaining performance data (observation), to the decision as to whether the performance meets the desired standard. The appraisee is both the center of the appraisal process and in control of it. Growth in performance, from this perspective, is inextricably context-bound.

While neo-traditional and neo-progressive perspectives mark opposite ends of the continuum, in practice, most growth-oriented appraisal processes fall somewhere between these extremes. Proponents of more collaborative forms of appraisal, such as those purported or explicated by Cogan (1973), Goldhammer (1969), Duke and Stiggins (1986) and Kilbourn (1990) provide examples of middle ground approaches.

According to Tracy and McNaughton (1989), the roots of clinical supervision can be found in the work of Goldhammer (1969) and Cogan (1973). Both stress the importance of the preconference for establishing the collegial relationship and setting the scope of the appraisal process. And both stress the importance of the collaborative assessment of the observation data. According to Pavan (1986), the preconference is wrongly dismissed by neo-traditional adherents as unnecessary. Pavan claims that the preconference is essential to understanding the appraisee's

reasons for particular behaviors. This understanding is not important to neo-traditionalist adherents; the behavior and the expected outcome of the behavior are the only important factors.

Kilbourn (1990) provides an instructive example of how the growth-oriented appraisal process can look when the supervisor and appraisee commit themselves to the process. In the case study that he presented, the form of the feedback is not strictly descriptive, but it is definitely more descriptive than evaluative. Kilbourn stresses the importance of the agreement between the supervisor and appraisee as to which topics are off-limits and which are to be focused upon in the appraisal process.

In addition, Kilbourn (1990) notes the difficulty of maintaining the collegial relationship throughout the course of the appraisal process. When addressing sensitive topics, the relationship between the supervisor and appraisee is strained. The importance of providing descriptive feedback, with a minimum of evaluative commentary, is evident in these situations. It seems obvious that the added stress that evaluative commentary would bring to these already tense situations would certainly not be helpful.

These collaborative forms of appraisal stress the sharing of responsibility for, and ownership of, the appraisal by both the appraisee and the supervisor. Such approaches have become fashionable lately (Cousins, 1995; Poole, 1994) because of the failure of more supervisor-centered approaches to promote growth. The shift from a supervisor-centered to more collaborative forms of appraisal has been accompanied by a shift in the form of the constructive feedback delivered to the appraisee. In more collaborative forms of appraisal, the nature of the feedback shifts away from the heavy-handed judgmental neo-traditional approach. However, neither is it likely to be completely descriptive as would be the ideal case from the neo-

progressive perspective. Rather, collaborative forms of appraisal are more likely to blend descriptive and evaluative feedback in a collegial manner.

On a deeper level, though, more fundamental issues are at play. The differences among the neo-traditionalist and neo-progressive approaches to appraisal reflect fundamental epistemological differences in their approaches to knowledge and knowing. As indicated above, the regimented, rational, objective approach favoured by the neo-traditionalists reflects a positivist epistemology, accompanied by the hypothetico-deductive research paradigm that has dominated inquiry in the social sciences for several decades. Within this paradigm, knowledge is context-free, and as such can be "packaged" and implemented in different contexts.

The neo-progressive approach, on the other hand, is consistent with an interpretivist epistemology, which reflects a constructivist, individualistic view of knowledge. Within this paradigm, knowledge is context-bound, and as such is considered non-valid once removed from the context within which it is created. This relatively recent and decidedly distinct view of knowledge has provided new and innovative forms of understanding inaccessible under the traditionalist paradigm.

The traditionalist and interpretivist paradigms reflect rather extreme views of knowledge, views that are parallel with neo-traditionalist and neo-progressive approaches to appraisal. These paradigms seem completely incompatible with one another. At a fundamental level they each reject the very assumptions about reality and knowledge that form the basis of the other perspective. Some reconciliation can be found in a "revisionist-traditionalist" view of knowledge put forth by scholars such as Cousins and Simon (1996), Huberman (1989, 1994), and Louis (1995). Within this view, the fundamental rejection of the traditionalist body of knowledge by the interpretivist paradigm is tempered by the recognition that both approaches have

contributions to make to our understanding. For example, the deep understanding possible in more qualitative approaches can help explain the more generalizable knowledge generated through more quantitative approaches. This compromise position, which draws from both of the extreme views of knowledge, can be positioned towards the middle of the continuum in Figure 1.1. Epistemologically, it assumes that certain regularities and uniformities are generalizable but that the importance of context, interpretability and adaptability cannot be denied.

1.4.1.2.2 Empirical research on evaluative content of feedback

On the issue of the evaluative content of feedback, the idiosyncratic nature of such professional activities as teaching or public speaking is one reason why more descriptive forms of feedback are being favored, at least in theory. The appraisee shares in the ownership of the appraisal because his or her involvement in the process is more integral in this approach as compared with other traditional approaches. While an integral component of neo-progressive approaches, increased involvement of the appraisee is inconsistent with neo-traditional approaches to appraisal.

Neo-traditional approaches to appraisal support the expert-novice, research-based model of supervision. The appraisee is the object of the evaluation, not a participant. The role of the appraisee is as a source of data. This observation data, and other data collected by the supervisor, is analysed by the supervisor. The supervisor reports the results of the analysis to the appraisee, and provides guidance in terms of rectifying performance that does not match the model.

Kilbourn (1990), in his case study of a teacher and supervisor, cites the importance of providing descriptive rather than evaluative feedback to the appraisee. Kilbourn warns that the improvement-oriented intent of the appraisal process must remain the focus; it is easy to drift to evaluative rather than descriptive feedback, but this shift can destroy both the relationship

(between supervisor and appraisee) and the process. In Kilbourn's view, a trusting relationship between the supervisor and the appraisee is essential to the success of the process.

In their review of the development of peer coaching applications, another form of growth-oriented appraisal wherein the appraisee and supervisor are peers interested in assisting each other, Showers and Joyce (1996) report that the innate tendency of people to drift from descriptive to evaluative feedback resulted in their decision to remove altogether the verbal feedback component of the process. Participants in their workshops now provide written feedback that can be edited for evaluative content, a method that Showers and Joyce report works well. Their observations underscore the difficulties inherent in trying to provide strictly descriptive feedback.

The reason for this change is that the evaluative nature of the verbal feedback delivered to appraisees was reported to have destroyed the collaborative nature of the process – even when the supervisor consciously tried to avoid being judgmental! This represents a refinement of their earlier work (Joyce & Showers, 1980, 1982; Showers, Joyce & Bennet, 1987), in which they report that informal verbal feedback, even within the context of peer coaching, does not consistently foster growth. "Unstructured feedback – that is, feedback consisting of an informal discussion following observation – has uneven impact. Some persons appear to profit considerably from it while many do not" (Joyce & Showers, 1980, p. 384).

Showers and Joyce (1996) maintain that eliminating verbal feedback can remove the expectations associated with neo-traditional forms of supervision – the appraisee's expectation of "first the good news, then the bad" (p. 15). This assertion is based on the assumption that written feedback can be restricted to pure description more completely than verbal feedback. However, while this may be the case, it might be noted that the observer's mere decision about what to

report in and of itself may constitute judgment. The shift from evaluative to descriptive feedback represents a shift from the traditionalist towards the interpretivist approach to appraisal. How far along this dimension one should move – from evaluative towards descriptive – in order to optimize growth is still open to debate. Showers and Joyce (1996) report that removing the evaluative aspect has not decreased appraisee growth at all, which leads us to ask whether evaluative feedback is important at all?

1.4.1.3 How to deliver descriptive feedback

We now turn to the issue of providing descriptive feedback. Observed above regarding the work of Showers and Joyce (1996) is the notion that providing descriptive feedback is easier said than done. One method that has come into favor has been the use of videotape technology. Quigley and Nyquist (1992) report that the use of video to provide feedback in growth-oriented appraisal with public speakers allows appraisees to adopt a role similar to that of the observer, helping to reduce the evaluative slant of the feedback itself. This observer role "creates a learning opportunity whereby students see how others might receive their performance" (p. 326).

The use of video feedback in speech performance development is also well documented. Karl and Kopf (1994) report, "videotaped feedback has become a widely used educational technique in business communication programs" (p. 213). In their experiment on feedback seeking behavior among management students in a presentation skills course, Karl and Kopf (1994) found that low performers (those who received low performance ratings in their presentations) were less likely to seek video feedback than high performers. But a significant confound of this study was that there were no controls for the importance of the course to the student.

Video feedback has also been used to stimulate reflection with music teachers. Berg and Smith (1996, p. 32) report that video "encourages teachers to reflect more deeply on their teaching by providing the opportunity to review a particular segment a number of times." This finding is echoed by Biggs (1980), who found that video may provide increased insight into growth issues: "This [video feedback] may prove valuable as an indication of *how* a strategy failed or succeeded rather than *whether* it did or not" (p. 582).

Boyce, Markos, Jenkins and Loftus (1996) studied grade 3 and 5 students receiving peer, teacher, or teacher-directed video feedback on a performance task. Their findings indicate that, for the older students, video feedback provided more improvement in performance than the other two methods. For the younger children, teacher feedback provided the most improvement in performance. Boyce et al. caution that this may be due to the failure of the younger children to focus on the improvement task, being distracted by the novelty of viewing themselves on video.

The issue of the impact of using 'new' technology is important. Karl and Kopf (1994) report that repeated self-viewing tends to reduce the self-focus of appraisees. Smith (1988), in a report from the field, suggests that over time appraisees will forget about the video equipment. Smith recommends that the video equipment be placed at the back of the room and be set up so that there is no need for an operator. This reduces the distraction to the performer. Smith also recommends that video monitors of the performance be positioned so that the performer cannot see themselves during their performance; a process feature that can be very distracting to the performer.

In summary, the use of video to provide descriptive feedback shows promise. The use of video can reduce some negative evaluative connotations of feedback. But caution must be taken

when introducing new technology, because the distraction can reduce the benefits. Berg and Smith (1996) also note that video feedback can be used to stimulate reflection.

1.4.1.4 Summary of research on constructive feedback

Findings from recent educational and management research and theory seem to be somewhat aligned with the neo-progressive position that the idiosyncratic nature of performance requires more responsive, individualized appraisal methods than have been utilized in the past. The more traditional, regimented, neo-traditional approaches seem not to serve the purpose of growth-oriented appraisal processes very well.

Unfortunately, the available evidence does not clearly indicate what is the best way to provide constructive feedback; that is, feedback provided with the intent of improving performance. Specifically, it seems that there are serious gaps in knowledge about the optimal balance of the evaluative content of feedback. The neo-traditional and the neo-progressive positions, at opposite ends of the continuum presented in Figure 1.1, both seem extreme. Empirical evidence suggests that the neo-traditional approach of providing strictly evaluative feedback has not effectively facilitated appraisee growth in the past. The neo-progressive approach of providing strictly descriptive feedback has not been tested empirically to date.

It can be argued given the emphasis on the development of self-evaluative capacity that this approach holds promise, but empirical evidence is wanting. The collaborative approach, where the appraisee and supervisor work together as partners to facilitate appraisee growth, seems to be growing in popularity, but constructive feedback under this approach, too, suffers from limited systematic scrutiny. We now turn to considerations that more directly bear on the sorts of effects that might be expected of constructive feedback, specifically self-efficacy.

1.4.2 Self-efficacy

The concept of self-efficacy (Bandura, 1977, 1995, 1997; Bandura, Adams & Beyer, 1977) was advanced as a synthesis of previous research on behavioral change. While Bandura (1977) believes that behavior is cognitively driven, he maintains that these cognitive drives are best changed through behavioral means. In his words, "the apparent divergence of theory and practice can be reconciled by postulating that cognitive processes mediate change but that cognitive events are induced and altered most readily by experience of mastery arising from effective performance" (Bandura, 1977, p. 191).

In general, self-efficacy can be thought of in terms of the model presented in Figure 1.2 (adapted from Bandura, 1977). In this model, the person who performs the behavior (performer) and observes an outcome of the behavior also experiences two forms of expectation: efficacy expectations and outcome expectations. Efficacy expectations relate to the performer's perceived self-efficacy (belief in one's ability to perform the behavior), while outcome expectations relate to the performer's expectations that the successful performance of the behavior will produce certain outcomes. Efficacy and outcome expectations are differentiated because they have different functions and effects on performance: "individuals can believe that a particular course of action will produce certain outcomes [outcome expectation], but if they entertain serious doubts about whether they can perform the necessary activities [efficacy expectation] such information does not influence their behavior" (Bandura, 1977, p. 193).

The focus of the present review will be on expectations of self-efficacy, because research has indicated that outcome expectations are less useful in predicting performance: "Experimental research strongly suggests that self-efficacy is a more powerful predictor of behavior than either

outcome expectations or past performance" (Sherer et al., 1982, p. 664). The first question to answer is: From where do these efficacy expectations come?

Figure 1.2. Framework of self-efficacy (adapted from Bandura, 1977).

1.4.2.1 Sources of efficacy expectations

Bandura (1977, 1995, 1997) cites four different sources of efficacy expectations: performance accomplishment, vicarious experience, verbal persuasion, and emotional arousal. Performance accomplishment "provides the most influential source of efficacy information because it is based on experiences of personal mastery" (Bandura et al., 1977, p. 126). Performance accomplishment refers to personal experience of the performer, the performer's own experience performing the behaviors.

Vicarious experience refers to observing others perform behaviors. The viewing of others successfully performing the behavior builds in the viewers the belief that they, too, can perform the behavior successfully. "Seeing others cope with threats and eventually succeed can create expectations in observers that they too should be able to achieve some improvement in performance if they intensify and persist in their efforts" (Bandura et al., 1977, p. 126).

More recent developments, such as the widespread and affordable availability of videotape equipment, present new opportunities for combining these two sources of efficacy expectations. A compilation of successful performances by a single performer can be produced. This compilation video can combine the benefits of both personal experience and vicarious experience, and may be a potent resource for developing efficacy expectations.

Verbal persuasion, while widely used due to its ease and ready availability as a source of efficacy expectations, is believed to produce weaker expectations than experience-based sources. Interestingly, emotional arousal has been found to influence efficacy expectations, at least in threatening situations: "People rely on their state of physiological arousal in judging their anxiety and vulnerability to stress. Because high arousal usually debilitates performance, individuals are apt to consider themselves more able when they are not beset by aversive arousal than when they are tense and viscerally agitated" (Bandura et al., 1977, p. 126).

These sources of efficacy expectations are not defined so as to be mutually independent; they may combine to produce increased or decreased efficacy expectations. Research into the dependent nature of efficacy expectations has not been abundant to date; no scientific evidence is available. However, recent research in teacher efficacy has contributed to our understanding of this construct.

1.4.2.2 Teacher efficacy

The construct of teacher efficacy has been defined in terms of Bandura's (1977) model for efficacy, which was presented earlier in Figure 1.2. There seems to be general agreement as to the efficacy expectations component of Bandura's model, which has been labeled as personal teaching efficacy. The outcome expectations component of Bandura's model, however, has produced varied results.

Ross, Cousins and Gadalla (1996), in their empirical study of within-subject predictors of teacher efficacy, define personal teacher efficacy as "an individual teacher's expectation that he or she will be able to bring about student learning" (p. 386). A related concept, general teaching efficacy, refers to the belief that students are teachable. These two types of efficacy are

independent of one another, and only weakly correlated. General teaching efficacy appears to be related to the outcome expectations portion of Bandura's (1977) model.

Gibson and Dembo (1984) conducted a survey of elementary teachers in order to develop an instrument to assess teacher efficacy, as well as to provide construct validation for the variable(s). Their 30-item survey was administered to 208 teachers, and the results of the factor analyses conducted revealed two related but relatively independent factors. The two factors were labeled personal teaching efficacy, which included items related to their ability to bring about student learning (representing efficacy expectations, Bandura, 1977), and teaching efficacy, which included items related to whether students are teachable (representing outcome expectations, Bandura, 1977).

Soodak and Podell (1996), in a re-examination of Gibson and Dembo's (1984) two-factor construction of teacher efficacy (personal teaching efficacy and teaching efficacy), found that there were actually at least three factors (which they labeled personal efficacy, outcome efficacy, and teaching efficacy) to be found in a modified version of the instrument. This finding confirmed earlier work by Woolfolk and Hoy (1990), which indicated that Gibson and Dembo's (1984) construct of teaching efficacy could be further subdivided into outcome efficacy, which represents the belief that student learning outcomes could be attributed to their actions, and teaching efficacy, which refers to the belief about the influence of outside factors (such as the home, heredity, and other influences outside the control of the teacher) on the impact of teaching.

The results of Soodak and Podell's (1996) study show that the subcomponents of the construct teacher efficacy have not yet been clearly identified. Personal teaching efficacy seems to be a fairly solid construct. However, the outcome-related variable teaching efficacy seems to require more study.

1.4.2.3 The measurement of self-efficacy

Bandura's (1977) conception of self-efficacy is as a situation-specific belief that *could* generalize to other situations (Bandura et al., 1977). Sherer et al. (1982) developed the Self-efficacy Scale as a "measure of self-efficacy that is not tied to specific situations or behavior" (p. 664). They developed the 23-item scale, comprised of two subscales (General self-efficacy with 17 items and Social self-efficacy with 6 items), which have been found to correlate with other personality measures. The scale was developed using a sample of undergraduate psychology students (n =376), and the original factor structure was confirmed using a second sample of 298 students. The General self-efficacy subscale was found to account for 26.5% of the variance, while the Social self-efficacy subscale accounted for 17% of the variance. The reported factor loadings underlying the results seem to support the results, the reliability estimates (Cronbach's alphas of .86 and .71 on the two subscales, respectively) are quite high, and the predicted relationships with other measures were found, with moderate correlations that do not undermine the validity of the scales. Sherer et al. (1982) assert that the General self-efficacy subscale measures self-efficacy without reference to any specific behavioral domain, while Social self-efficacy subscale items measure efficacy expectations in social situations. The instrument seems to adequately measure non-situational self-efficacy.

Sherer et al. (1982) caution that their self-efficacy scale is not intended to replace more focused measures designed to assess specific dimensions of self-efficacy. They recommend that more specific wording of the questions on these behavior-specific measures can provide accurate estimates of performers' self-efficacy. In their words, "The self-efficacy scale is not intended to replace more specific measures that assess expectations for specific target behaviors. When dealing with target behaviors in unambiguous situations, more specifically worded questions or

direct behavioral measures are likely to provide the most accurate estimates of an individual's efficacy expectations" (Sherer et al., 1982, p. 671).

Vispoel and Chen (1990) conducted a review of purported measures of self-efficacy. Their purpose was to identify and review measures of self-efficacy reported in the educational and psychological literature, in order to provide a guide for others who seek to assess self-efficacy or to develop new measures of self-efficacy. The Self-efficacy Scale as developed by Sherer et al. (1982) was identified as one of the best resources for both measuring general self-efficacy and for use in developing and validating other measures of self-efficacy.

Sherer and Adams (1983) performed further work on the construct validity of the Self-efficacy scale developed by Sherer et al. (1982). Using 101 introductory psychology students, they assessed the relationship between scores on the Self-efficacy Scale (Sherer et al., 1982) and several other established psychological tools (e.g., the MMPI). The results of the study showed the expected correlations between the Self-efficacy scale and the other scales. Sherer and Adams (1983) conclude that the data add support for the construct validity of the Self-efficacy Scale, while several unexpected correlations require further study before any conclusions can be drawn.

These findings, while not troublesome, suggest that the construct of self-efficacy may be difficult to define conclusively. Care must be taken when asserting that an instrument measures self-efficacy, in whatever form. The construct must be carefully defined, as well as the measure. An observation, related to the public speaking literature, is that outcome expectations and the actual outcome of the performance are not mentioned anywhere. This may be related to the results seen in the teacher efficacy literature. The outcome construct *teaching efficacy* seems to be less clearly defined than the efficacy expectation construct personal teacher efficacy. This

suggests that measuring outcome expectations, and outcomes, may be more difficult than measuring efficacy expectations.

1.4.2.3.1 The measurement of public speaking self-efficacy

While specific measures of self-efficacy in various performance domains have been developed, in the area of public speaking the focus to date seems to have been on performance-anxiety rather than efficacy as a performer. A review of the literature on assessing public speaking ability generated multiple studies that purported to measure ability to some extent; however, very few provided any meaningful statistics that would lend credence to their true reliability and validity. These studies tended to use qualitative pass / fail type measures; indications that the behavior exceeded, met, or failed to meet the required level.

One empirical study that did report meaningful results was that of Ellis (1995), who developed the Self-perceived Public Speaking Competency Scale (SPSCS) from the competencies identified by Morreale (1990; Morreale et al., 1991). The refined instrument consists of one factor that accounted for 32% of the variance in Ellis' (1995) study of 155 undergraduate students. Three separate administrations of the instrument (to the same sample at weeks 1, 4 and 16 of the semester) produced reliability estimates of .89, .82 and .84. The SPSCS seems to be a good measure of public speaking competence, and provides reliable items for adaptation to a self-efficacy study; the items were originally written for use with 5-point Likert scales.

Willmington, Benson, Neal and Steinbrecher (1993) studied the communication skills of beginning teachers in teacher certification courses and developed their Public Speaking Rating Form. This form, consisting of Likert-scale ratings of various components of public speaking, is quite similar in content to Ellis' (1995) instrument. It also addresses two unique points not

addressed by Ellis but reported in the anecdotal literature: the use of filler words (and non-words) and general vocal quality. Unfortunately, as they report no statistics, the reliability of these items is unknown at present. They do, however, provide a starting point to address these issues in a systematic manner.

In summary, the literature on assessing public speaking ability and efficacy is quite thin. The focus in the literature has been on public speaking anxiety. The literature that exists, however, is quite promising in terms of adaptation for use in other research. Again, there is no mention of measuring either outcomes or outcome expectations in the literature that was found. The development of tools to assess outcome expectations and outcomes of public speaking is an area that could benefit from further investigation.

1.4.3 Summary of Literature Review

The provision of constructive feedback for the purpose of improving performance is a complex undertaking. The body of knowledge on feedback for summative purposes is extensive, while the body of knowledge for formative purposes is relatively insubstantial. Where direction is given, there seems to be little application to practice. Where results from practice inform the guidelines, there is a distinct lack of substance.

In summary, the enhancement of performance efficacy expectations seems to hold promise as a tool for helping improve performance (the goal of formative appraisal and constructive feedback). The teacher efficacy literature has provided some insight into the construct of self-efficacy. While some instruments have been developed to assess various aspects of self-efficacy, there is a distinct gap when it comes to assessing self-efficacy in public speaking.

Three main reasons support the use of a measure of performance self-efficacy, as opposed to outcome or outcome expectation measures, in the present study. First, the existing literature supports measures of performance. That is, existing instruments have been validated as measures of performance. These instruments can logically be adapted for use as measures of performance self-efficacy, provided adequate validation and reliability assessments are conducted. There is no basis for assuming their validity as measures of outcomes or outcome expectations.

A second reason for using a measure of performance self-efficacy is that the measurement of outcomes related to public speaking, or even outcome expectations, in an empirical study would be very difficult to implement. Nevertheless, a number of factors emerge as being relevant and difficult to control. For example, the size of the audience for the presentation will probably have an impact on the outcome and outcome expectations. Secondly, the number of times the audience has heard the same presentation will probably have an impact. Although it could be argued that these factors may also impact on performance self-efficacy, there is justification for using performance self-efficacy: other measures of performance have been devised. No measures of outcome or outcome expectations in the area of public speaking were found.

The third reason for using a measure of efficacy expectation, as opposed to outcomes or outcome expectations, is that efficacy expectations are arguably the best predictors of performance. According to Sherer et al. (1982), "experimental research strongly suggests that self-efficacy is a more powerful predictor of behavior than either outcome expectancies or past performance" (p.664).

One contribution the present thesis will make will be the development and validation of an instrument for measuring performance self-efficacy. While this is not the central focus for the thesis it is a necessary step needed in order to empirically study constructive feedback in the context of public speaking. A second contribution is the integrated review of the literature from the fields of education, psychology and business management with respect to constructive feedback.

1.5 Research Question

The preceding literature review has been focused on the issue of the extent to which evaluative content in feedback explains performance improvement in growth-oriented (formative) appraisal. The gaps in knowledge highlighted in the preceding sections lead us to question what impact descriptive versus evaluative forms of constructive feedback have on performance self-efficacy within the context of a growth-oriented appraisal process. This question is grounded in a much deeper theoretical issue concerning the relative merits of neo-progressive versus neo-conservative approaches to performance appraisal. The available evidence is far from conclusive.

In the present study, the impact of the evaluative content of constructive feedback on self-efficacy in the public speaking performance domain is of central interest. It is hypothesized that the impact on performance self-efficacy of predominantly descriptive feedback will be more powerful than that of predominantly evaluative feedback. The rationale supporting this hypothesis is grounded in recent theoretical developments concerning neo-progressive approaches to performance improvement. Specifically, it is argued that a focus on developing the performers capacity for self-critique through the provision of descriptive feedback will lead to higher levels of performance self-efficacy.

H_1: Participants who receive descriptive feedback will report significantly greater improvement in their performance self-efficacy than participants who receive evaluative feedback

The crux of the argument is that the development of one's capacity for self-evaluation and reflection will give rise to the development of a stronger sense of self-efficacy and confidence and ultimately improved performance. A secondary rationale for the hypothesis comes from the observed failure of traditional forms of appraisal, which have focused on evaluative feedback, to promote growth. The available evidence indicates that traditional forms of appraisal have not had the expected or desired impact. Having said that, however, it must be recognized that there currently exist no evidence indicating that non-traditional methods (descriptive feedback) would necessarily be better. The thesis will add significantly to the theoretical literature in this respect.

The hypothesis is tested in the domain of public speaking performance. Specifically, the study focuses on public speaking performance efficacy expectations, as opposed to outcomes (impact on the audience) or outcome expectations.

Chapter 2 – Review of Related Literature

2.0 Introduction

In this chapter, an integrated review of the literature that bears directly on the current study from the fields of education, psychology and management is presented, in order to provide context for the research question. As a starting point, a distinction between formative and summative performance appraisal is offered. Next, a general framework for growth-oriented appraisal developed by Cousins (1995) is presented. This framework is presented to help conceptualize the present thesis. A review of research on public speaking, the field of interest in the present study, is presented. Next, summaries of research in fields related to personnel evaluation are presented. Links to the current study are established in the section summaries, justifying their inclusion.

2.1 Performance Appraisal

The two fundamental functions of performance appraisal are evaluation for personnel decision-making and accountability, on the one hand, and supervision for professional growth and development on the other. Historically in education, for example, a clear distinction between these two functions in either policy or practice has not been made. Moreover, performance appraisal policies have been shown to be relatively benign in their contributions to, and effects on, either personnel decision-making or professional development. A large-scale survey of performance appraisal practice in Ontario school boards, for example, showed a dismal picture of the utility of appraisal policies (Lawton et al., 1986). Similar conclusions have been reached in surveys in the U.S. (Stiggins, 1988).

Within the past 15 years, private and public sector organizations have begun to formally differentiate the two functions in organizational performance appraisal policies and standard

operating procedures. Growth-oriented or formative appraisal procedures have been identified as distinct activities in organizational practice. However, additional research on the effectiveness of growth-oriented appraisal is required to demonstrate its usefulness for organizations.

2.1.1 Formative / Summative Distinction

The distinction between formative and summative forms of performance appraisal lies in the intent of the appraisal process: Formative forms of appraisal provide information useful for improving performance, while summative forms of appraisal provide information useful for decision-making (e.g., hire / fire, promotions). Currently, systems for evaluating teachers focus on accountability issues almost to the point of excluding professional growth (Duke, 1990a).

In the educational milieu, there are two main schools of thought: those who believe that formative and summative evaluation must be separate entities in order to achieve their goals (e.g., Popham, 1988), and those who believe that one evaluation system can serve both formative and summative purposes (e.g., Hunter, 1988a). Within this first group, there are again two streams of thought: those who believe that formative evaluation should not be combined with summative evaluation, and those who don't consider formative evaluation a form of evaluation at all (e.g., Scriven, 1987, 1988). In the next section, I will review the educational, psychological and management literature on the formative / summative distinction in performance appraisal.

2.1.1.1 Educational literature

2.1.1.1.1 Educational literature supporting the distinction

Popham (1988) typifies those who believe that teacher evaluation must employ different methods to achieve its distinct goals. In fact, in this conceptual piece, Popham insists that the two types of evaluation are so different that different individuals should carry them out in order to enhance the separation. "In spite of its prevalence, the blending of formative and summative

teacher evaluation represents a grave conceptual error. Both formative and summative evaluation are important functions, but these two teacher evaluation tasks must be carried out separately by different individuals" (p. 58).

In Popham's (1988) view, the intent of formative evaluation is the improvement of teacher's skills. This increase in teacher skill is expected to result in improved student learning. No summative consequences should attend formative evaluation: "There should be no tenure or termination decisions associated with formative teacher evaluation; it is exclusively improvement focused" (p. 58). The primary reason for this separation is that teachers would be reluctant to reveal their weaknesses in a summative evaluation context. They would be reluctant to take performance risks, which some consider to be essential to performance improvement (e.g., Cogan, 1973; Goldhammer, 1969).

Summative evaluation is viewed as a completely separate entity. The purpose of summative appraisal is to establish the value of performance (good or bad). Its primary uses are to identify incompetent teachers and to recognize excellence. Summative performance evaluation provides support for personnel decision-making. Although judgment is the main focus, elements of performance improvement are often part of the process. In the case of teachers performing at less than acceptable levels, for example, a process of remediation would be involved. If behavior is un-remediable, then summative evaluation provides support for the termination of the teacher. Obviously, in a situation where formative and summative functions are not distinguishable, teachers would be reluctant to reveal areas where they feel they need improvement at the risk of suffering evaluative career-related consequences.

McQuarrie and Wood (1991) also advocate the separation of formative and summative appraisal processes. Formative appraisal (supervision, in their words) should be used to provide

support for teacher learning and application of new or refined skills. They write that formative appraisal should be used to "help and support teachers as they adapt, adopt, and refine the instructional practices they are trying to implement in their classrooms" (p. 93). According to McQuarrie and Wood, a primary use of summative appraisal is to ensure the competence of the teacher. In this strict sense, there is no overlap between formative and summative appraisal. Summative appraisal is used "to make judgments about a professional under review" (p. 94).

In their survey of 109 elementary and secondary teacher's perceptions of appraisal, Ovando and Harris (1993) report that the focus of teacher evaluation is shifting from summative towards formative evaluation. They stress that formative and summative evaluation should be separate entities because the purposes are distinctly different. In their words, "the purpose of formative teacher evaluation is to provide constructive feedback about their teaching performance, not to make administrative decisions" (p. 301).

In summary, Popham (1988) provides a rationale for the separation of formative and summative forms of evaluation. The basis for the separation is that the purposes of the appraisal (improvement versus accountability) will impact the teacher's participation in the process. Teachers will be less likely to participate freely in formative appraisal at the risk of revealing weaknesses that may be used in a summative context, unless there is explicit separation between the two appraisal purposes. McQuarrie and Wood (1991) and Ovando and Harris (1993) provide support for this separation.

Scriven (1987, 1988) agrees that formative and summative forms of appraisal must be separate, but for different reasons. According to Scriven, evaluation consists of rendering judgment about the merit or worth of the object of the evaluation (the evaluand). As there is not necessarily a judgment about the worth of the behavior in formative appraisal, it does not qualify

as evaluation. To Scriven, it makes sense that evaluation (summative evaluation) and formative advice-giving or coaching be separate entities. Otherwise, teachers would not participate. "Without this separation, it is unreasonable to expect teachers to go to formative advisers about their weaknesses" (p. 114).

Scriven (1987, 1988) argues that formative evaluation is not evaluation because it does not necessarily render judgment about the merit or worth of the behavior. While this may seem more of a semantic argument than a philosophical one, the underlying issue of the rendering of judgment on merit or worth (*evaluating*, to Scriven) is important. What is interesting about Scriven's argument is that the judge is always assumed to be external or separate from those being judged. There is no acknowledgement that the evaluation of performance merit or worth can take the form of self-evaluation. Further comment on this theme is provided below.

2.1.1.1.2 Educational literature opposing the distinction

Not everyone agrees that formative and summative appraisal should be separate entities. Hunter (1988a), in a theoretical piece on successful teacher evaluation, stresses that there is no need to separate formative and summative appraisal for two reasons. First, a skilled evaluator can perform both functions. Second, summative evaluation is the ultimate goal: determining whether the teacher is capable of performing the task.

According to Hunter, "Weakness in teaching cannot be concealed from a competent evaluator" (1988a, p. 49). Essentially, teacher evaluation consists of comparing observed behavior to a pre-conceived model of effective teaching (developed from the literature on effective teaching). The evaluator determines whether the behavior conforms to the model or not. If the behavior does not conform closely to the model, two options exist. Remedial advice can be

given in order to correct improper technique, or the teacher can be terminated. Successive evaluations follow the same model.

Hunter (1988b), in response to criticism of her views on teacher evaluation, attempts to reconcile her views with those who advocate the separation of formative and summative forms of evaluation (e.g., Popham, 1988). Here, she states that the summative evaluator is in the perfect position to provide formative advice to those being evaluated because he or she can provide the motivation for improvement: "The intent to grow can be stimulated as a result of supervision by someone who has the power to make a final evaluation and who has collected ongoing data to support final evaluation" (p. 275). Interestingly, Hunter says that the different expectations of teachers undergoing formative and summative evaluation reinforces her argument that they should be combined activities. "This fact provides provocative evidence that supervision and evaluation should be marriage partners, not divorced activities" (p. 279).

The highly prescriptive, expert-novice approach to evaluation espoused by Hunter (1988a, 1988b) has become the norm in education. Unfortunately, the success of this approach has been mixed. Standardization of teaching behavior may have been achieved to some extent. The concurrent standardization and, implicitly, improvement of student learning (the desired impact) has not been seen, at least to any great extent. Empirical tests of this relationship are scarce.

2.1.1.1.3 Summary of the educational literature re: distinction

No answer to the question of whether formative and summative evaluation should be separate or conjoined has been provided. The theoretical literature has proponents lined up on both sides of the argument. Scriven's position may be thought of as middle ground. He argues

that formative evaluation is not evaluation at all. However it seems reasonable to conclude that Scriven sees performance appraisal as a judgment oriented accountability mechanism.

2.1.1.2 Psychological and management literature re: distinction

In the psychological and management literature, there does not seem to be the same chasm between those who insist on the separation of formative and summative appraisal and those who do not. Stephan and Dorfman (1989) provide a distinction between appraisal purposes that seems to correlate with the formative / summative distinction seen in the educational literature. Administrative appraisal purposes serve to inform decision-making with respect to salary, merit increases, promotions, demotions, demotions and transfers. Administrative appraisal purposes would represent summative appraisal. Developmental appraisals serve to facilitate "performance improvement through job-related feedback provided by supervisors in a helpful manner" (p. 27) and would represent formative appraisal.

In their empirical study, Stephan and Dorfman (1989) attempted to determine whether these different forms of feedback would add to produce greater results than either one individually. They explored the impact of different forms of feedback on a computer data entry task performed by 72 female undergraduate students. The results of their study showed that each form of feedback produced an increase in performance on the task (data entry), but not differentially.

Stephan and Dorfman's results also showed that while administrative feedback increased performance on behaviors directly related to rewards, developmental feedback increased performance on a wider variety of behaviors. The only effect of combining the forms of feedback was to increase the negative emotions directed towards the supervisor (the one who delivered the

feedback). The results of this empirical study seem to indicate that administrative (summative) and developmental (formative) forms of appraisal should be separate functions.

Kruger (1985), in a conceptual piece on personnel evaluation in business, talks about the separation of formative and summative evaluation in terms of evaluating behavior and evaluating results. Kruger calls for the use of more subjective measures when evaluating behaviors, utilizing information sources such as self-report and peer appraisal in the process. Kruger recommends more objective measures when evaluating results, limiting the information sources to supervisor observation. Thus, Kruger supports the distinction between formative and summative evaluation.

Fletcher (1986), in a conceptual piece on the effects of performance appraisal, criticizes traditional forms of performance appraisal in business management. In Fletcher's view, managers fail to identify weaknesses in appraisee's behavior for fear of hostile or defensive reactions. In order to prevent the conflict that arises when both appraisal and formative feedback are offered, Fletcher recommends separating these functions.

It seems that the psychological literature on personnel evaluation supports the separation of formative and summative forms of evaluation. Both the empirical (Stephan & Dorfman, 1989) and theoretical (Kruger, 1985; Fletcher, 1986) literature support this separation. As in the educational literature, however, there is a distinct paucity of empirical research on this issue in the psychological literature.

2.1.2 Formative / Summative Distinction Summary

A shift towards separating formative and summative forms of evaluation in education has begun. A similar separation is seen in the psychological literature, although without the opposition to this move that has been seen in education. It will be interesting to see what results the separation produces, given time.

The purpose of including this review is to provide an understanding of the difference between formative and summative forms of appraisal. In the current study, the emphasis is on formative appraisal. Specifically, growth-oriented appraisal processes as outlined by Cousins (1995). In the next section, a detailed exploration of the nature of growth-oriented appraisal processes is provided.

2.2 Growth-oriented Appraisal Process

In this section, growth-oriented appraisal processes are outlined. The educational, psychological and management research relevant to the components of the process are reviewed. Finally, gaps in knowledge are highlighted.

2.2.1 Introduction to Growth-oriented Appraisal

The concept of the growth-oriented appraisal process is rooted in the models of clinical supervision originally developed by Goldhammer (1969) and further explicated by Cogan (1973). In general, growth-oriented appraisal can be thought of as an ongoing process, consisting of a pre-observation meeting to establish the goals of the process, an observation period, and a post-observation meeting to interpret the observations and determine how to move forward. Cousins (1995), alluding to the ongoing nature of growth-oriented appraisal, states that "the [appraisal] process is best conceived as nonlinear and cyclical" (p. 201). Rather than one-shot evaluation, it is a continuous process geared toward long-term performance improvement.

2.2.2 Growth-oriented Appraisal Process Framework

In Cousins' (1995) general framework (presented in Figure 2.1), the growth-oriented appraisal process is presented as a cyclical process of preconference (preparation), observation (data collection) and postconference (feedback), situated within a larger framework that includes factors that influence the appraisal process and the impact of the appraisal process. In the next

sections, the components of this framework are examined. The relevant literature from education, psychology and management is reviewed within each section.

2.2.3. Factors That Influence the Formative Appraisal Process

In the Cousins (1995) framework (Figure 2.1), several factors that can influence the appraisal process in an educational context are detailed. These factors consist of appraisee, supervisor, and organizational characteristics that can have an impact on the preconference, observation period and postconference components of the process. These factors can influence both the growth-oriented appraisal process itself, and indirectly, the impact of the appraisal process. In addition, there seems to be considerable dependence among the three categories of factors. Some of the characteristics associated with the appraisee include desire to receive constructive feedback, desire for growth, and knowledge of self. Characteristics associated with the supervisor include training and the provision of time to carry out the process. Organizational characteristics include administrative support and policy history with respect to formative evaluation.

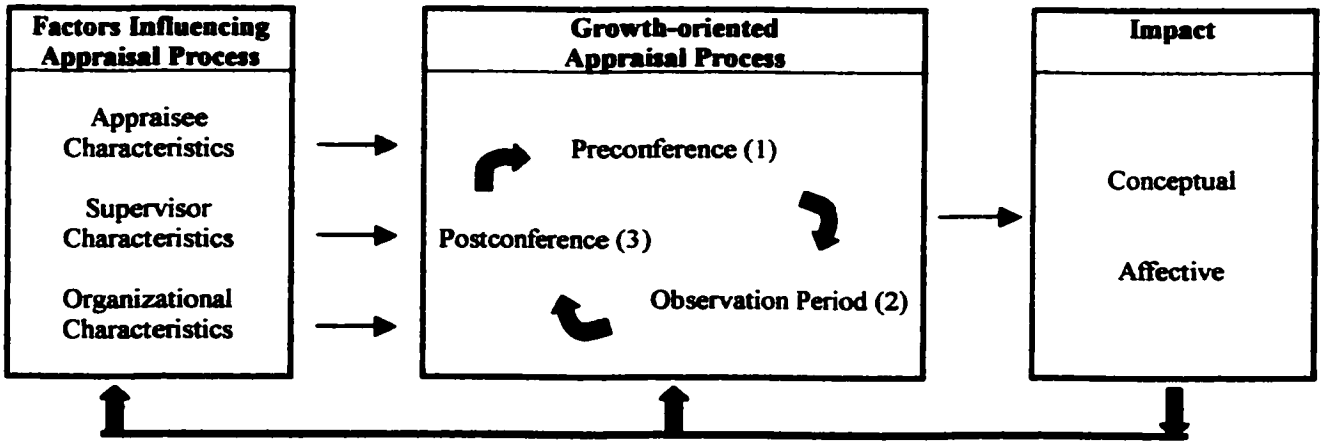


Figure 2.1. General form of growth-oriented appraisal processes (adapted from Cousins, 1995).

But what impacts do these appraisee, supervisor and organizational characteristics have on the appraisal process itself? We have clues as to the characteristics that are more often

associated with effective appraisal than others, although empirical evidence is wanting. In the following sections, we will review the educational, psychological and management literature relevant to these factors that can influence the growth-oriented appraisal process.

2.2.3.1 Appraisee characteristics associated with effective formative appraisal

In a conceptual piece on teacher evaluation, Barber (1990) writes that in order to see improvement, appraisees must first admit that there is room for improvement and that they want to improve. "To improve teaching through formative evaluation, a teacher must first admit that he or she is doing something less than perfectly and that the teacher's behavior can be improved by one or a combination of the many techniques of formative evaluation" (p. 216). This awareness of the need to improve, or the existence of a gap between actual and desired performance, forms the basis for entering into the growth-oriented appraisal process.

Barber (1990) also notes that the admission of the existence of this performance gap requires trust between the teacher and the supervisor. Without this trust, that the supervisor has the best interests of the teacher in mind, it would be unreasonable to assume that the teacher would admit that there were problems with their performance. "Such an admission requires much trust between the evaluator and the teacher being evaluated. Clearly, the formative evaluator must have the best interest of the teacher at heart before the teacher will admit that improvement is needed or possible" (p. 216).

Burke and Fessler (1983) also cite the identification of the need for growth as a crucial component of a collaborative appraisal process. The other three components in their model are: feedback, internalization of the results, and the development of an action plan from the results of the process. In their view, one of the biggest issues to be faced in the process is that of convincing the appraisee that the focus of the process is on their concerns and their

improvement; thus, the emphasis on the internalization of the results. The success of the process depends on the appraisee accepting the feedback (results) as valid and helpful, and committing to developing and following the action plan. If the appraisee does not accept the feedback in this way, the process collapses.

Duke (1990b) also writes about the importance of identifying the growth goal and committing to the process. Implicit in identifying a growth goal is the admission that growth is possible and, presumably, desirable. In Duke's view, there are four main activities that constitute professional development goals. They are breaking routines, changing perspectives, examining assumptions, and reading challenging materials.

Breaking routines correlates with the behavioral changes that are the basis of most formative appraisals. Changing perspectives is probably beyond the scope of most formative appraisal processes, as are examining assumptions and reading challenging materials. These activities seem to be components of a more comprehensive program of professional development (as outlined by Duke, 1990a, 1990b; Duke & Stiggins 1986, 1990) of which formative appraisal can be seen as one aspect.

The concepts of admitting the existence of a gap and commitment to the process mesh with Cousins' (1988) recommendations in his piece on implications for practice from research on evaluation utilization with school principals. Evaluation utilization refers to the actual outcomes of the evaluation process, such as changes in behavior, as opposed to the results of the evaluation itself. Some of the most important factors in the success of the process include user commitment to the evaluation process and receptiveness to evaluation. User commitment to the process refers to involvement in the actual process of the evaluation, while receptiveness to evaluation refers to willingness to be a part of the evaluation process and to accept the results of the process.

Cousins (1995) conducted a survey of persons with responsibility for teacher supervision in Ontario (n = 152) on their views of a variety of factors associated with collaborative performance appraisal. The results of this study indicated that respondents agreed that teachers were aware of their strengths and motivated to improve. Respondents did not agree that teachers were aware of their weaknesses, nor that they should be the ultimate judges of their own performance. An interaction effect was found with respect to perceptions about whether heavy workloads intrude on the appraisal process. Respondents affiliated with secondary panels agreed that this was a factor, while respondents from elementary panels tended to disagree that this factor was important.

Cousins (1995) admits that there were limitations to this study. These limitations involve both the sample and the instrument. First, although the sample represents a wide geographic area and many school districts, it was obtained through convenience sampling methods. This limits the generalizability of the findings. Second, rigorous investigation of the validity and reliability of the instrument was not carried out prior to implementation. Although the items were developed from a thorough review of the literature, the results may be biased. As an exploratory study, however, it does provide direction for further systematic investigation.

Beck and Seifert (1983) reinforce the importance of trust between the appraisee and the evaluator in their guidelines for instructional supervision. They add that not only must the supervisor be viewed as having the teacher's interests in mind, but the supervisor must also be seen as competent to help the appraisee improve. This second factor, the perceived competence of the supervisor to help the appraisee, is an important issue that certainly bears more investigation.

In his review of literature on collaborative performance appraisal, Cousins (1988) notes that the motivation of the appraisee is frequently cited as an important factor in selection of personnel for participation in growth-oriented appraisal. The more motivated the candidate is to become a better performer, and to participate actively in the appraisal process, the more likely the process is to succeed. This seems intuitive, but does not seem to have been tested empirically.

Darling-Hammond, Wise and Pease (1983), in their review of teacher evaluation literature, also report that the motivation of the appraisee is crucial to the success of the process. Appraisal without the full cooperation and investment of the appraisee is a waste of time. In their words, "Effectively changing the behavior of another person requires enlisting the cooperation and motivation of that person, in addition to providing guidance on the steps needed for improvement" (p. 314).

McLaughlin (1990) writes that the entire outcome of the appraisal process is contingent on the support of the teacher: "Teacher evaluation will be no more effective than the extent to which teacher's support it" (p.404). Here, McLaughlin is referring to the entire teacher evaluation genre, not just formative appraisal. The message is no less relevant, though. The commitment of the appraisee to the process is essential to the success of the process.

In Brock's (1981) view, there are three main factors that determine whether formative appraisal is likely to succeed. First, teachers who see student learning as a function of their teaching ability are more likely to improve in a formative appraisal process than those who see the two as more distinct and unrelated. Secondly, teachers who receive more favorable evaluations (as opposed to unfavorable) are more likely to improve. Thirdly, teachers who think they are better than their students rate them are more likely to improve.

The first assertion, that teachers who relate student learning to their teaching are more likely to improve, seems obvious. If there were no such relationship, there would be no need to refine teaching skills, as they have no impact on student learning. In truth, this would raise the question of how one can assess teaching, if not in terms of student learning. Advocates of the effective teaching movement (e.g., Hunter, 1988a, 1988b) base their entire rationale on the linkage between teacher behavior and the resultant impact on student learning. Others, such as Popham (1988) argue that the effective teaching literature does not conclusively link the behaviors with student learning.

The second assertion, that teachers who receive more favorable evaluations are more likely to improve, on the surface seems simplistic. There may be more at work here; for example, the motivation of the participants may be higher when they receive more favorable evaluations. It seems incautious to make this blanket statement without more empirical evidence and a clearer definition of exactly what the construct is. The third assertion, that teachers who think they are better than their student ratings indicate, suffers from the same lack of clarity and empirical support as the second assertion.

Duke and Stiggins (1986), in their guide to teacher evaluation based upon their extensive empirical work, emphasize that the single most important factor in teacher evaluation is the teacher. The teacher characteristics that they deem most important include instructional competence, personal expectations, openness to suggestions, orientation to change, subject knowledge, and experience. Of these, the most relevant to the current work are personal expectation, openness to suggestion and orientation to change.

Personal expectations refer to how the teacher accounts for student success (i.e., the relationship between teacher behavior and student learning), their approach to professional goal

setting, and their reaction to professional development opportunities. Openness to suggestion refers to the belief that useful insights are available from a variety of sources. Orientation to change refers to the teachers' belief that they can succeed in new endeavors, that they have a need to succeed, and that they are committed to the process. Teachers who link student learning to their behavior, believe in the goal setting process, are open to and interested in professional development activities, are open to suggestion, and have a positive orientation towards change are most likely to achieve positive results from participation in formative evaluation processes.

Stiggins (1988), drawing from case study research, reports that the appraisee characteristics associated with successful appraisal include openness to constructive criticism, a positive orientation towards risk-taking and change, and high personal performance expectations. These characteristics are similar to the ones presented in Duke and Stiggins (1986). Again, the recommendations emphasize that the appraisee must admit the need for improvement in performance.

Duke and Stiggins (1990), in their theoretical piece on moving beyond minimum competence to professional growth, reaffirm their earlier views on the characteristics of appraisee's associated with successful appraisal. Although most of the characteristics are the same as in the earlier piece, there is an interesting addition. They refer to the appraisee's prior experience with evaluation. This may be a refinement of the more broadly focused factors from their earlier work labeled openness to suggestion and orientation towards change (Duke & Stiggins, 1986). It seems intuitive that one's previous experience with an activity, be it positive or negative, would impact on one's future behavior when confronted with a similar activity. Intuitive or not, it certainly bears consideration when designing an appraisal system.

In their case study of a classroom teacher and supervisor engaged in formative appraisal, Grimmet and Crehan (1990) reported two essential appraisee factors that helped determine the success of a formative appraisal process. First, the appraisee, not the supervisor, must identify the issue to be dealt with in the process. This seems to ensure the interest and commitment of the appraisee. Second, the appraisee must feel secure enough to admit the shortcoming and to ask for help. Once these conditions are met, the stage is set for formative appraisal and professional growth. Unfortunately, as Grimmet and Crehan point out, "these conditions occur all too rarely today" (p. 234). Grimmet and Crehan foreshadow the potency of capacity for self-critique and evaluation as integral to the formative process.

In summary, the main appraisee characteristics that are associated with successful growth-oriented appraisal are a desire for growth, commitment to the appraisal process and the identification of areas where growth is needed. Not explicitly addressed but likely to be foundational to these observed factors is the concept of capacity for self-evaluation.

2.2.3.2 Supervisor characteristics associated with effective formative appraisal

McLaughlin (1990) writes about the need for the supervisors and appraisees to undergo training together, for more reasons than just to learn the skills of appraisal. The common training sessions allow the participants, both supervisors and appraisees, to come to a consensus understanding of the parameters of the appraisal process. "Joint training sessions comprise important opportunities to develop common understanding about evaluation, common expectations for the process and outcomes, and common language to talk about all aspects of the process" (p. 408). Another key component in McLaughlin's view is the commitment of the supervisor to the process. This can be seen in the willingness to participate in the joint training sessions, among other ways.

In a theoretical piece, Blair (1991) writes about the infrequency of observation and the failure of supervisors to make time for discussion with appraisees. Teachers, says Blair, see the principal as the arbiter of when and where appraisal takes place, which may or may not be the case, and assign blame for infrequent observation to the principal. "They express concern, however, about the infrequency of supervisory visits and the unavailability of principals and other supervisors to assist with teaching refinements" (p. 103). In Blair's opinion, the blame may be appropriately placed: "Administrators avoid the practice of supervision if at all possible" (p. 102). Unfortunately, no empirical evidence is offered to support this statement.

Haefele (1993) also raises the issue of time for appraisal. Supervisors may feel that there is not enough time, either to properly appraise the performance or to give due feedback to the appraisee. This feeling of a lack of time for the process may be interpreted by the appraisees as a lack of commitment to the process by the supervisor, which may have an undermining effect on the process.

Kilbourn (1990) provides a prime example of the amount of time required for effective appraisal. In his exemplary case study of a teacher and supervisor working together in a growth-oriented appraisal process, Kilbourn explicates the vast amount of time and effort required of both the teacher and the supervisor. This detailed account of the 13-day intensive process undertaken by these dedicated individuals to assist the teacher in improving his skills provides keen insight into the process and content of constructive feedback in a growth-oriented appraisal process.

The amount of time and effort required of the participants is well documented, and exhaustive. At the end of the process, Kilbourn notes that the teacher seemed to find the feedback sessions worthwhile, although extremely time-consuming and stress producing. The

stress was mainly caused by attempts to change teaching style in response to the feedback sessions. Interestingly, Kilbourn notes that the supervisor was not as positive about the feedback sessions. The supervisor found the sessions draining due both to the intensity of the sessions and a feeling of frustration that progress was not as rapid as expected. Further, the supervisor found that some of the topics within the feedback sessions were outside the scope of the appraisal process, for example personal issues. Uncertainty as to how to respond, or even whether to respond, to these topics also caused stress for the supervisor. In summary, Kilbourn (1990) shows how involved, in terms of both time and effort on the part of both the supervisor and the appraisee, the growth-oriented appraisal process can be.

To summarize, it seems evident that one of the supervisor characteristics that is related to successful appraisal is a willingness to commit time and effort to the process. This time and effort can include both time for the actual meetings and observation, as well as for more innocuous activities such as improving the collegial relationship between supervisor and appraisee.

Brock (1981) writes about evaluation-based teacher development. In Brock's view, the most important characteristics of effective supervisors are that they adopt a collegial attitude towards the appraisee, and that they see the partnership of appraisee and supervisor as collaborative rather than expert-novice. This collegial approach to the process, rather than the more prescriptive stance adopted in more traditional appraisal models, is essential to the success of the growth-oriented appraisal process.

Darling-Hammond et al. (1983), in their review of teacher evaluation literature, note that teachers report dissatisfaction with the frequency of appraisal. Interestingly, they note that although the infrequency of the evaluation caused dissatisfaction among teachers, whether the

feedback was positive or negative was not important to the teachers. This indicates that the systematic provision of feedback from supervisors to teachers is more important to teachers than the provision of positive feedback in those feedback sessions.

Duke and Stiggins (1986) also report on the important characteristics of supervisors in the teacher appraisal process. The supervisor characteristics that they report as being most important are credibility, patience, and trust. Credibility refers to the supervisors' expertise in both teaching and the subject area of the teacher. As well, the supervisors' familiarity with the appraisee, the school district, and the supervisor's own classroom experience all effect their credibility.

Patience is expressed in terms of the supervisor's willingness to wait for the appraisee to feel comfortable performing the new behaviors before reassessing their impact, and the supervisor's willingness to try different approaches if the first attempt doesn't prove fruitful. Finally, trust refers to both the teachers belief that the information shared in the appraisal will be kept confidential and that the supervisor and appraisee are truly partners in the process, with the same end-goals in mind. Once again, the issues of trust and confidentiality are key to the success of the appraisal process.

Stiggins (1988) also talks about the important attributes of the supervisor. Primarily, they should be perceived by the appraisees as a credible source of sound ideas for improvement and trustworthy, in terms of keeping confidential the shared information. In addition, the supervisor should strive to establish a track record of having helped others improve, and be able to model or demonstrate the behaviors that need to be changed. Interestingly, this last attribute brings a new dimension to the fore: the supervisor in a growth-oriented appraisal process being demonstrably capable in the performance domain. This raises the question of whether the supervisor needs to

be superior to, or at least equally as good as, the appraisee in the performance domain. The credibility of the supervisor may be a serious issue. Further research into this issue would seem to be warranted.

The issues of trust and commitment are echoed by Haefele (1993), who talks about the requirement that the appraisee have faith in the appraisal system and the supervisor. "If teachers are to be motivated to achieve high levels of performance and seek training when needed, they must have faith in the evaluation process, results of the process, and, in particular, the competence of the evaluator who implements the process" (p. 26).

The psychological and management literature seem to devote less time to the characteristics of supervisors that underlie effective appraisal. One empirical study that does get at such factors concerned the performance evaluation of nurses. Ferris, Judge, Rowland and Fitzsimmons (1994) report on social and situational context characteristics that impact on the appraisal process. They report that higher assessments result when supervisors and appraisees have similar attitudes and values, while there is no apparent link between higher assessment and demographic characteristics. In their study, involving 95 nurses and 28 supervisors, they investigated the differential impact of a variety of factors including affect, experience, physical distance, and demographic variables. The results of their study revealed that the experience of the supervisor correlated positively with higher ratings. Interestingly, the other relevant finding was that supervisor affect for appraisees (whether the supervisor liked them or not) correlated positively with better appraisals (i.e., higher affect was associated with higher rating).

The results of this study seem to raise more questions than they answer. The fact that more experienced (in terms of job-related activity, not appraisal) supervisors gave higher ratings than less experienced supervisors raises the question of why this occurred. Was there some form

of bias in the supervisor sample, or was there some confound to the study that produced this result? It is difficult to know in the absence of more information on the research program. This unexplained result does raise the question of the validity of the research.

The second finding is perhaps more troubling, in that it raises the issue of the validity of the appraisal process. If there is substantial bias in the underlying mechanism, it will be difficult if not impossible to convince participants that the process is worthwhile. The growth-oriented appraisal process, rooted in trust between the appraisee and the supervisor, cannot thrive when such obvious bias is evident.

In summary, supervisor characteristics that are associated with effective appraisal are the training of both the supervisor and the appraisee, as well as the provision of time for the process and the establishment of trust between the supervisor and appraisee. The inclusion of the provision of time for the process assumes that the supervisor has control over the amount of time that can be made available. From the Kilbourn (1990) study we can see that the growth-oriented appraisal process can be extremely time-consuming and stressful to both the supervisor and the appraisee. These factors should be taken into account when designing formative appraisal processes.

2.2.3.3 Organizational characteristics associated with effective formative appraisal

McLaughlin (1990) writes that teachers' suspicions and cynicism about the purposes of appraisal are certainly justified. Teachers consider appraisal to be in their interest only if it is seen to be embedded in a much broader effort on the school district's behalf that involves the investment of time, money and effort. As well, McLaughlin cites the involvement of teachers in the process as an important factor in determining the success of the process. "The involvement of

teachers, in summary, contributes at least three essential factors to the successful development and implementation of a teacher evaluation effort: greater teacher commitment to evaluation and motivation to participate fully; greater tolerance for the inevitable mistakes and flaws in the developing system; advice and assistance for teachers during the course of development that may be crucial to the system's long-term viability" (p. 407). Interestingly, although McLaughlin (1990) consistently refers to the inclusion of various stakeholder groups in the process, there is no mention of what many consider to be the primary stakeholder group: the students.

In their survey of 109 high school and elementary teachers, Ovando and Harris (1993) report the most important factors influencing the evaluation process. Chief among these were the attitude of the administration towards the process, the establishment of a collegial relationship between the supervisor and the appraisee, and the provision of time to carry out the appraisal process.

Stiggins (1988), in a case study of a school district, offers recommendations for the effective implementation of formative appraisal processes. His main recommendation is that the school district must be seen to be intrinsically involved with and supportive of the process. This involvement and commitment should take the form of district policies that reflect the importance of the process, the inclusion of clauses in the collective bargaining agreement that reflect the commitment to growth, and the provision of resources (personnel, equipment, etc.) for carrying out the appraisal process. Another key component is the provision of time for preconferencing, observing, and postconferencing. When all of these factors are in place, the context for effective appraisal has been provided. While this does not guarantee effective appraisal, it at least allows for it to happen.

Duke and Stiggins (1986), in their district guide for developing effective growth-oriented appraisal systems, outline three key organizational factors. The first factor consists of three types of administrative documentation: state law, district policy, and contractual obligations. The extent to which these documents reflect a commitment to, and support for, the appraisal process at all levels of the administration goes a long way towards developing the organizational climate most likely to produce positive appraisal results, and to foster commitment to and trust in the process from the appraisee viewpoint. Also included here is the history of labor relations in the district, and specifically how appraisals have been handled in the past. It is important to note that, although the codification of these principles is important, it is equally important that the implementation of these policies be seen on an ongoing basis.

The second factor is the amount of time spent on appraisal. This includes time spent designing and implementing the system, convening goal-setting conferences, conducting pre-observation conferences, conducting observation, and conducting post-observation conferences. In addition, the inclusion of informal classroom visits and feedback sessions, customizing the process for individuals, and the inclusion of other sources of information (e.g., peers, students) can be beneficial.

The third factor is the availability of resources to aid the process. This can include time off from regular duties to visit other classes in order to observe other teachers' behavior, technical assistance from consultants and/or experts, and the availability of audio and video resources. Also important are access to information retrieval systems, the development of both organized and informal staff development activities, and the implementation of peer mentoring programs.

In their survey of appraisal policies of school boards across Ontario, Lawton et al. (1986) conclude that the essential problem with existing appraisal systems was that the burden for carrying out the exhaustive amount of work involved lay with the administrators on a local level. This overload was found to be largely responsible for the reported tendencies towards token preparation for the process due to time constraints, the use of standardized performance criteria rather than individualized criteria, the reliance on more limited amounts and varieties of data, and less than desired frequency of appraisal. Obviously, the solution to this overload is not to be found at the local level; there needs to be commitment of time and resources at the district and provincial levels, also.

McLaughlin (1990), in a conceptual piece on implementing teacher appraisal, also writes about the importance of overt support at the organizational level. In McLaughlin's view, the teacher's justified mistrust of traditional appraisal systems can only be changed through the investment of time, money and effort on the part of the school district. Only when the district is seen to be involved can appraisees be expected to participate fully in the process. "Teachers considered evaluation meaningful and in their professional self-interest only when it is embedded in a broader district effort to improve educational practices" (p. 405).

The provision of resources to support the appraisal process is integral to the credibility of the organizational efforts. From the appraisees' perspective, the organization must focus on investing in the process of appraisal, rather than on the outcomes of the process. If the organization focuses on the outcome, the tendency to hide problems emerges. If the organization focuses on the process, participants feel freer to take risks and attempt to master new skills.

Within the psychological and management literature, Wildman and Niles (1987), in their guidelines for professional growth, list autonomy as one of the most important conditions that

facilitates professional growth. They base this finding on their review of the psychological and management literature. By autonomy, they mean that workers feel that they are free to experiment in terms of both behavior and problem solving, without fear of repercussions for mistakes. This is in opposition to the traditional, hierarchical personnel development models found in less progressive organizations. Wildman and Niles feel that the development of similar autonomous cultures in the school system may produce similar results.

In summary, the organizational characteristics that seem to be associated with effective appraisal are the provision of resources for the process and the overt commitment of the organization to the process. Ways to demonstrate this organizational commitment include, but are not limited to, the inclusion of clauses in policy documents, the provision of resources (both money and equipment) for carrying out the process, and the provision of time to implement the various components of the process (preconferences, observation periods, and postconferences).

2.2.3.4 Summary of factors that influence the formative appraisal process

In the preceding sections, we have reviewed the major categories of factors that can influence the growth-oriented appraisal process. These factors have been categorized into appraisee, supervisor and organizational characteristics. In general, appraisee characteristics associated with effective appraisal processes included a desire for growth, a commitment to the appraisal process, and the identification of areas where improvement is needed.

Supervisor characteristics associated with effective appraisal included joint training with the appraisees, and the provision of time for the process. The provision of time for the process is a somewhat controversial factor, because this may be beyond the control of the supervisor. Nevertheless, given the results of Kilbourn (1990), it is important. Finally, the psychological and management literature reports that the correspondence between the values and attitudes of the

supervisor and appraisee can have an impact on the process. Organizational characteristics associated with effective appraisal include district policies that reflect commitment to the appraisal process and the provision of time, money and other resources to carry out the appraisal process.

As can be seen from the literature reviewed, the factors that impact on the growth-oriented appraisal process are complex and intermingled. For example, take the issue of the provision of time for the process. Even if the school district policy documents the need for time for the process, this does not guarantee that it will be provided in practice, given the multiple competing priorities that must be juggled by administrators.

We now turn our attention to the educational, psychological and management literature that deals with the components of the appraisal process itself and the impact of the process. The components of the process are the preconference (preparation phase), observation period (data collection phase), and postconference (feedback period). The impact, or outcome, of the appraisal process can be either conceptual or affective (Cousins, 1995).

2.2.4 Characteristics of Effective Preconferences

In the preconference, the supervisor and the appraisee meet to determine the professional growth goals, as well as the methods of observation and the formalities for the communication of feedback. As the growth-oriented appraisal process is typically ongoing and cyclical, there can be long-term goals that extend beyond one supervision cycle, as well as short-term goals that are addressed in fewer or even one cycle.

Acheson and Gall (1987), in their book on techniques in the supervision of teachers, identify several activities that should be carried out at the planning conference stage of the appraisal process. In their view, the supervisor and appraisee must first identify the teacher's

concerns about instruction, and then translate these concerns into observable behaviors. Next, they identify procedures for improving the instruction, and then set goals for improving the teacher's behavior. Obviously, Acheson and Gall advocate a highly collaborative approach to appraising performance, as the identification of the goals and the actual goal setting are done by the appraisee and supervisor together.

Cousins (1995) provides guidelines for effective collaborative evaluation as a result of his empirical study of educational personnel with responsibility for supervision of teachers in Ontario. In terms of the preconference, he reports that the literature indicates that appraisees should be given specific criteria that the appraisal will focus on, and that the appraisal should only focus on a small number of criteria at a time. In addition, the goals of the appraisal process, which should be negotiated by the supervisor and appraisee, should be as closely aligned with the organization's goals as possible.

The results of Cousins' (1995) survey did not support all of the theoretical guidelines presented. Specifically, respondents indicated that they did not agree that appraisee's growth goals needed to overlap with school and district priorities. They did agree that it is important that the supervisor have enough information to set expectations for the teacher, that the number of criteria should be limited, and that the goals of the process should be collaboratively arrived at. The results of this study indicate that personnel with responsibility for supervision in the Ontario educational system are, generally, in favor of formative appraisal for the sake of improving the appraisee's performance, whether or not this aligns with the organizational objectives.

McGreal (1988), in a theoretical piece on enhancing instruction, reinforces the importance of collaborative goal setting in the preconference. The involvement of the appraisee

at this stage allows both the supervisor and appraisee to focus the appraisal on a narrow set of behaviors, as well as building trust and cooperation between the two.

Lawton et al. (1986) provide some characteristics of effective appraisals from their large-scale survey of Ontario teachers, principals, superintendents, trustees and directors. In their view, effective performance appraisals tend to have preconferences, and the preconferences that occur tend to be longer than in less effective appraisals. In addition, within these preconferences, there tends to be clear communication of the purpose of the appraisal. Finally, during the preconference, the participants develop concrete plans for the appraisal. One limitation to the work of Lawton et al. is that they failed to distinguish between formative and summative forms of appraisal. Thus, their findings are difficult to interpret in the context of formative appraisal.

Hunter (1988a, 1988b) has a somewhat different view on the preconference. In Hunter's view, the preconference is unnecessary, because a competent supervisor is capable of identifying weaknesses through direct observation of teacher performance. Input from the appraisee is not needed, as the supervisor can compare the performance with the research-based model of effective teaching to determine where improvement is required. In Hunter's words, "Except in rare cases, pre-observational conferences are unnecessary" (p. 46).

In summary, the preconference, which serves as the planning and/or goal setting portion of the process, is either an integral part of the growth-oriented appraisal process (Lawton et al., 1986) or a superfluous luxury (Hunter, 1998a, 1988b). Although the literature suggests that appraisee goals should align with school and district priorities, the respondents in Cousins' (1995) empirical work dispute this claim. The goals of the process should be collaboratively set, and a concrete plan for the process should be agreed to by the supervisor and appraisee.

2.2.5 Characteristics of Effective Observation Periods

The observation period is the data-collection or information generation phase of the growth-oriented appraisal process. Although direct observation of the performer in action by the supervisor is the most common method of collecting observation data (Cousins, 1995), other methods are also used. In fact, Cousins' empirical research with Ontario school personnel with responsibility for supervision revealed that multiple data sources were recommended.

Gordon, Meadows and Dyal (1996) in their survey of 100 principals from each of Alabama, Pennsylvania and Washington report that observation is an expected part of the appraisal process. The 148 respondents indicated that formal observation of teachers was mandated at either the state or local level in 94% of the cases. Teachers were typically observed at least twice a year, with only 8% of the respondents reporting only one observation per year, and the vast majority reporting that all teachers are included in the process, not just those that need improvement or that are untenured. Finally, more than half of the respondents reported that the observations lasted at least one full instructional period, while only 3% reported that observations lasted less than half of a period.

In a theoretical piece on teacher evaluation, McGreal (1988) lists a number of methods in addition to direct observation that can be used for collecting data for use in feedback sessions: parent evaluation, peer evaluation, student evaluation, artifacts (e.g., tests, notes), student performance data, and self-report. McGreal (1988) also notes that using data from more sources increases the reliability of the overall process.

Brock (1981) advocates the use of audio and video recordings of teacher performance, in addition to student ratings, as sources of data for improving performance. The purpose of using these multiple sources of data is to provide as broad and accurate picture of the appraisee's

behavior as possible. This information is provided in hopes of inspiring the appraisee to want to improve their performance.

Campbell and Lee (1988), in their theoretical piece on the use of self-appraisal in performance evaluation, note that self-appraisal is becoming more acceptable as a data source for appraisal. Specifically, they see it as useful in two situations. First, self-appraisal can serve as additional information when the subjective self-appraisal data aligns with more objective data sources. Second, when the information sought is not available through other sources (for example, self-efficacy beliefs: beliefs about one's ability to perform a behavior). In particular, Campbell and Lee (1988) cite self-appraisal as a potentially powerful tool when the objective of the appraisal is to stimulate reflection on behavior and receptivity to suggestions from others.

Darling-Hammond et al. (1983) list students, peers and self-assessment as potential data sources for appraisal purposes. They also note that, in order to provide the most useful information for appraisal purposes, the data must not only help the supervisor and appraisee identify areas of weakness, but also provide information about potential remedies for these weaknesses. Darling-Hammond et al. also report that self-evaluation is being used more frequently, and that a combination of self-evaluation and goal-setting can "promote self-reflection and motivation toward growth and change" (p. 308). Once again, self-reflection, or capacity for self-critique, is identified as an important component in the formative appraisal process.

Duke and Stiggins (1986) provide detailed information on the data that they feel should be collected as part of effective performance appraisal processes. First, the performance criteria and standards to be utilized for the appraisal process should be tailored to the context and abilities of the individual appraisee. They should also be collaboratively negotiated with the

appraisee. They list the examination of teacher records (e.g., lesson plans, tests, assignments, grading practices, comments made to students, and student notes) as important sources of information. Duke and Stiggins (1986) caution that measures of student achievement (e.g., test scores) should not be used as data sources, as they are not precise enough measures, nor has the link between teacher behavior and student achievement been made clearly enough to justify their use.

Interestingly, Duke and Stiggins (1986) stress that the purpose of classroom observation is to provide *description* of what is occurring in the classroom during the observer's presence. The interpretation of this information is carried out collaboratively with the appraisee at a later time (during the postconference). The purpose of providing this descriptive information is to enable appraisees to have input to the evaluation of the data, thereby developing their capacity for self-critique.

They also comment positively on the use of more subjective forms of appraisal, such as self-assessment, student assessment, and peer assessment. Duke and Stiggins note that, although these data sources may not be rigid enough for use in a summative evaluation context, they can provide useful information for formative purposes. The use of these forms of appraisal also reinforces the importance of helping appraisees develop their capacity for self-critique.

Acheson and Gall (1987) list some aspects of instructional behavior that should be observed. In particular, they advocate that students' time on task, the amount of time they spend actively engaged in academic activities, is an important variable to consider. In addition, investigation of the teacher's planning can reveal rich information for formative appraisal. They also recommend the use of correlates of effective teaching, which form the basis of Hunter's (1988a, 1988b) system of teacher appraisal, as a basis for observation. Extra-classroom activities

that can yield useful information include the collegial relationships the appraisee forms as well as their engagement in other professional development activities.

Stodolsky (1990) provides some interesting ideas on observation in a theoretical piece on classroom observation. The main issue raised concerns validity versus reliability in observation. Great pains have been taken to ensure reliability in observation, while relatively little effort has been expended in addressing validity.

Although "observations may be particularly well suited for the purposes loosely labeled 'improvement', because they provide a base for discussion of actual classroom teaching by individual practitioners" (p. 177), Stodolsky cautions that it is less useful for examining underlying constructs such as thought processes and feelings. These less easily observable constructs require more highly inferential assessment systems. The more inference required, the more training required (of the observer), and the less reliable the observation.

In terms of number of observations, Stodolsky (1990) recommends that more is better. She goes further, stating that the same number of minutes of observation spread over a longer period of time (i.e., shorter, more frequent observation periods) is preferable to longer, less frequent observations. The reasoning is that because of the idiosyncratic nature of teaching, generalizability to other appraisees is not an issue. More important is the larger picture, a broader understanding of the specific behaviors of the appraisee.

Within the psychological and management literature, Schay (1993) provides a review of performance models used in U.S. government departments. In this conceptual piece, Schay reports that appraisees in the government service are unhappy with the subjective nature of their appraisal processes. Typically, this appraisal is based on the input of one person: the supervisor. Schay also reports on a survey of federal employees that revealed the majority would like their

supervisor to have input into their appraisal, but they also wanted data collected from other sources.

Nearly half of the respondents to the survey indicated that they would like to see forms of self-evaluation incorporated into the appraisal process (Schay, 1993). Other data sources identified included peer evaluations, and evaluations from both internal (persons from other government departments) and external customers. Schay concludes that formative appraisal systems may benefit from increasing the number of data sources they incorporate. "The next generation of performance appraisal systems may follow the multirater approach which relies on input from various sources, including internal and external customers. This approach holds promise, particularly if it is used for feedback rather than evaluation" (p. 667). Unfortunately, Schay fails to elaborate on the survey methodology, which brings into question the validity of the results. Schay also fails to differentiate between formative and summative forms of appraisal.

Tziner and Latham (1989) report the results of their empirical study exploring the differential impacts of two rating instruments on the outcome of the appraisal process. The participants in their field study were 20 managers and 125 workers at an airport. The managers used either a behaviorally anchored rating scale or a trait-based rating scale to record performance observations. The workers were familiarized with the rating mechanism to be utilized before the appraisal period began. The dependent variables were measures of job satisfaction and organizational commitment. The results of the study revealed that both job satisfaction and organizational commitment increased after the feedback on performance was provided. Interestingly, workers who received feedback based on the behaviorally anchored rating scales showed a significantly greater increase in job satisfaction and organizational commitment than workers who received feedback based on the trait-based rating scale. Schay

claims that the reason for this observed difference was that the behaviorally anchored rating scale explicitly defined the performance requirements for both the worker and the supervisor. Trait-based rating scales tend to be more generic in nature, leading to uncertainty in the appraisee as to what specific behaviors and levels of performance are desired. Schay (1993) concludes that behaviorally anchored rating scales are more desirable than trait-based rating scales, because they provide clearer parameters for both the supervisor and appraisee.

In summary, the characteristics of the observation period, or data-collection portion of the growth-oriented appraisal process, which are associated with effective performance appraisal include the use of multiple sources of data and multiple observations, where possible. These multiple sources of data can include self-report, student and peer assessment, audio and video recordings, and artifacts (e.g., tests and notes). This finding was consistent across the educational, psychological and management literature reviewed. Multiple observations of shorter duration are preferred over fewer observations of longer duration. Little detail is offered as to how observations should be conducted, although Duke and Stiggins (1986) report that the purpose is to provide description, rather than evaluation, of the behaviors that occurred.

2.2.6 Characteristics of Effective Postconferences

The postconference is typically the point in the growth-oriented appraisal process where the supervisor and appraisee meet to discuss the observation. The postconference is the primary locus for giving and receiving constructive feedback.

Acheson and Gall (1987) recommend the following format for the postconference. First, the supervisor provides objective observational data to the appraisee. The appraisee analyzes the observations, and together they interpret the meaning. The appraisee is responsible for deciding

on modifications to their behavior in order to improve performance; the supervisor merely supports these decisions.

Acheson and Gall (1987) also provide recommendations for supervisors to improve conferences. They suggest that the supervisor attempt to paraphrase the appraisee's comments, in order to enhance understanding; that the supervisor ask clarifying questions when appropriate; that they avoid giving specific advice; and that they give specific praise for performance and growth. These activities are all geared towards developing the appraisee's capacity for self-critique.

Cousins (1988), in an empirical study of principal's use of their own performance appraisal results, reports that supervisors should try to include unexpected results in the postconference. By unexpected results, Cousins' refers to observational data and interpretations that do not support or follow logically from the preconference specifications. These unexpected results may provide enhanced stimulation for growth within the process, as well as more information for discussion in the postconference. These results may also trigger the appraisee to engage in more reflection on their performance.

Drawing from their empirical and theoretical work on teacher evaluation, Duke and Stiggins (1986) provide some guidelines for the feedback portion of the growth-oriented appraisal process. Important factors include the specificity of the feedback, the frequency of the feedback, the timing of the feedback, and whether the supervisor offers evaluative commentary or merely description of behavior. In terms of the specificity of the feedback, the type of feedback that encourages growth addresses very specific behaviors in detail, rather than more general commentary. Duke and Stiggins recommend that feedback be provided "with sufficient regularity to allow them [appraisee's] to track their improvement" (p. 33), which, although

vague, seems to indicate more frequent feedback than current appraisal processes provide for. As far as the timing of the feedback is concerned, Duke and Stiggins (1986) indicate that feedback as soon as possible after the observation is best. Finally, they are in favor of collaborative interpretation of the data, rather than a prescriptive monologue from the supervisor.

Haefele (1993) also discusses the postconference. He recommends that the focus of the postconference be on the appraisee's strengths and weaknesses. This is not the way it occurs in practice, however, according to Haefele (1993). "Postobservation conferences should be a regular part of the evaluation process and should focus on the strengths and weaknesses of the teacher's performance ... apparently this is not what happens. Many teachers complain about the lack of helpful feedback in these conferences" (p. 26).

Kilbourn (1990) also provides guidance with respect to the postconference. As noted earlier, his in-depth case study of a teacher and a supervisor participating in a multi-week growth-oriented appraisal process specifically focused on helping the teacher improve his teaching performance. One of the main lessons from this case study was that, in order to maximize the effectiveness of the appraisal process, the feedback should be limited to the areas clearly defined by the participants in the preconference or planning session for the appraisal process. This also reinforces the importance of the preconference and the collegial relationship between the supervisor and the appraisee.

Some of the other important factors gleaned from Kilbourn (1990) were again the specificity of the feedback, the frequency of the feedback, the timing of the feedback, and whether the supervisor offers evaluative commentary or merely description of behavior (similar to Duke & Stiggins, 1986). In terms of the specificity of the feedback, a lot of time seemed to be spent (wasted?) by the teacher and supervisor in clarifying the terms being used by one or the

other. Limiting the feedback to the agreed upon areas, and making that feedback as specific as possible, would certainly help reduce the amount of time lost to clarification. This seems especially important when the amount of time and feedback available to appraisees is at issue.

Kilbourn (1990) does not directly address the issue of frequency of feedback and timing of the feedback sessions. In the case study presented the supervisor and appraisee meet every day as soon as possible after the observation. The only factors that interfere seem to be their respective schedules. Because they are so closely involved in the process, two interesting issues emerge.

First, because they met every day, progress did not seem to occur in a predictable manner; rather they seemed to be tired of the process on some days, and to have renewed energy on others. Perhaps the pressures of working so intensely on the process, combined with the regular stresses of their daily lives, combine to produce these highs and lows. Obviously, the supervisor and appraisee do not operate in a vacuum and outside pressures would impact on the process.

The second issue is the timing of the feedback. The postconference was typically held as soon after the observation session as schedules permitted. This immediacy of the postconference seemed to produce tension between the two that may have been alleviated by a longer post-observation interval. Essentially, the appraisee was sometimes still "in the moment" after a less-than-ideal class, and the immediate feedback from the supervisor seemed to add to the tension. The immediacy of the feedback seemed to help the appraisee in terms of being able to recall his behavior during the session. Unfortunately, the added tension may not have been worthwhile, and since the supervisor was providing the observation from notes taken during the observation session, recall may not have been as big an issue as tension.

The issue of whether the supervisor offers evaluative commentary was also not directly addressed by Kilbourn (1990). In the case study, evaluative feedback is clearly given by the supervisor, as well as guidance as to how the appraisee should go about changing his behavior. Kilbourn notes that, as the case study was conducted *in vivo*, there was implied responsibility on the part of the supervisor to ensure the students in the class received the best instruction possible. The appraisal process was a secondary issue, although the supervisor and appraisee were very committed to the process. Kilbourn does address this dilemma, of who is responsible for what in the process.

In the ideal case, the teacher is solely responsible for the classroom and students. The supervisor is responsible to the teacher, for providing feedback and commentary on what happened in the observation session, and even providing advice as to how to improve the classroom behavior. This split, between the teacher and the supervisor and their respective responsibilities, may be difficult to implement in practice. Ultimately, though, teachers need to be the judges of what is done or is not done as a result of the feedback, because they have the responsibility for the classroom. Here, Kilbourn (1990) provides further support for the continuing theme of developing the appraisee's capacity for self-critique.

Lawton et al. (1986) found disturbing results in their survey of performance appraisal policies in the Ontario school system. They report that, of the boards that had performance appraisal policies, only 61% mandated a postconference. This represents only 41% of school boards in Ontario. In addition, only 43% of school boards that have performance appraisal policies (32% overall) required that a plan of action be developed from the appraisal process. A potential confound to their results is found in the fact that they found little consistency between responding school boards in terms of distinguishing between formative and summative

evaluation, as well as between evaluation and supervision. It would be useful to have broadly accepted terminology in the field, in order to clarify communications.

In summary, characteristics of effective postconferences include having a postconference that occurs soon after the observation (Duke & Stiggins, 1986). The data provided should be as specific as possible (Duke & Stiggins, 1986; Kilbourn, 1990) and the appraisee should be allowed to evaluate the data (although there is no clear indication that collaborative evaluation of the data is better or worse).

2.2.7 Outcomes of Growth-oriented Appraisal Processes

The impact, or outcome, of the growth-oriented appraisal process can be either conceptual or affective (Cousins, 1995). Conceptual outcomes refer to changes in the thoughts of participants in the process. Examples of conceptual outcomes of the process are: new knowledge, adjusted focus, new skills, and changed classroom practice. Affective outcomes refer to changes in the attitudes of the participants in the process. Examples of affective outcomes are: effects on morale, commitment, and job satisfaction. Cousins (1995) notes that there can also be unexpected and/or indirect outcomes, and that these can be either positive or negative. An example of a positive outcome that is not necessarily expected could consist of teachers' taking the appraisal process more seriously after a good experience. An unexpected negative outcome could be anxiety resulting from the identification of a problem that cannot be resolved (Natriello, 1990).

Other authors have dealt with unexpected or indirect outcomes of growth-oriented appraisal. As reported above, Kilbourn (1990) notes that the process is highly stressful for both participants. Handled in the wrong way, this could cause serious damage to the relationship between appraisee and supervisor. In the case study, the participants repeatedly face the obstacle

of one or the other being highly stressed, either due to the nature of the process or outside factors that impact on the process either directly or indirectly. Kilbourn (1990) surmises that the collegial approach to the process, as well as the appraisee's right to determine the parameters of the feedback sessions, saves the relationship from collapsing outright.

Stiggins (1988) notes that there are two main influences on the impact that the performance appraisal process has. The first influence is the performance data, and the second is the feedback provided to the appraisee. With respect to the performance data, the better the quality of the data, the more likely there will be impact. With respect to the feedback, the same holds true; the better the feedback provided, the more likely the appraisal process will have an impact on the appraisee. In using the term 'better' Stiggins refers to characteristics of the performance data and feedback that conform to the guidelines presented earlier in this review.

The appraisee's ability to evaluate his or her own performance is another important component. The concepts of 'better' data and 'better' feedback are related to the appraisee's ability to evaluate them. The supervisor can assist in the evaluation, but the ultimate responsibility rests with the appraisee (Kilbourn, 1990).

The key message to take from this section is that there are many possible outcomes of the growth-oriented process, specifically conceptual and affective outcomes. In addition, the possibility of unexpected or indirect outcomes exists. It is important that those who choose to undertake growth-oriented appraisal make every effort to ensure that they provide support for all outcomes of the process, not just the obvious ones.

2.2.8 Summary of Growth-oriented Appraisal Process

The Cousins (1995) framework can be taken as a fairly generic version of growth-oriented appraisal processes, which have grown from models of clinical supervision developed

from the work of Goldhammer (1969) and Cogan (1973). The general form of the process is as an ongoing cycle of a preconference, observation period, and postconference, or feedback session. Obviously, there are many variations on this theme.

There are many factors that can influence the process, some of which are appraisee, supervisor and organizational characteristics. The outcomes of the process can also feed back into the next preconference and observation period. One emergent theme has been the development of the appraisee's capacity for self-critique. Although interesting observations can be made from the available literature, there is a lack of systematic empirical research into the characteristics of effective growth-oriented appraisal in the educational, psychological and management domains.

This section was included in order to familiarize the reader with growth-oriented appraisal processes, the factors that influence these processes, and the outcomes of these processes. The continually emergent theme of self-reflection and self-critique supports the rationale for the hypothesis provided in Chapter 1.

Next we will examine the literature on public speaking, the performance domain of interest in the current study.

2.3 Public Speaking

The vast majority of the literature available on public speaking seems to be devoted to the treatment of public speaking anxiety. The main distinction made is between state and trait anxiety. State anxiety refers to anxiety related to actual public speaking performance, while trait anxiety refers to generalized anxiety. Other factors related to public speaking that have been studied, although not to the same extent as anxiety, include preparation and performance outcomes.

Menzel and Carrell (1994) explored the relationship between preparation and performance in public speaking using 119 public speaking students. The performance criteria were rated on five-point scales and included eye contact with the audience, voice quality, gestures / movement, energy, and thought content (a combination of organization and structure of the speech). The preparation criteria assessed was the total time, in minutes, spent on nine different types of preparation (e.g., silent rehearsal, oral rehearsal, research) for the speech they were to present. Anxiety of the speaker was also assessed using written survey instruments. The final measure was the subjects' self-reported grade point average at college.

The results of this study revealed that the strongest correlational relationship existed between grade point average and both the cumulative performance criteria and thought content. This means that students who had a higher grade point average tended to score higher on both the performance measures and the measure of thought content. The strength of this relationship was not exceptional (0.25), accounting for only 6% of the variance in the model.

The authors conclude that the more time spent in preparation, the better the performance. In addition, they note that the type of preparation was also related to performance. The more realistic the practice sessions (e.g., out-loud performance, before an audience) the better the performance. In summary, the authors "recommend rehearsal before an audience as a way of improving speech delivery" (p. 24).

Watt (1995) explored the effects of a public speaking course on the communication skills of business students. Of the 19 students (14 undergraduate students, 5 graduate students) who completed the program, 13 were female. Participants completed a pre- and post-course 37-question self-rating form that assessed their speaking ability. The results of the study supported the main hypothesis that student self-ratings of their speaking ability would increase after they

completed the course. The results did not support the two additional hypotheses. First, the group of female students produced statistically significant increases in their scores, while the male students did not increase significantly. This finding was contrary to the second hypothesis that there would be no significant difference between males and females. The third hypothesis, that graduate and undergraduate students would not perform differently, was also rejected. The undergraduate students as a group showed statistically significant increases in their specking skills, while the graduate students did not.

Watt (1995) concludes that the reason that males did not perform as well as females is that males have better developed communication skills than females, and thus their improvement is not expected to be as great. The explanation for the graduate students not performing as well as the undergraduate students is that the graduate students have both more professional and educational experience. These conclusions are unacceptable for two main reasons.

First, the small sample size and unequal cell sizes prevent any serious statistical analysis being done on the data presented. There were twice as many females as males, and almost three times as many undergraduate students as graduate students. Second, the analyses presented do not take into account the pre-existing ability of the students. A more appropriate analysis would have been to use the pre-test as a covariate to remove this error variance.

Ellis (1995) studied the relationship between public speaking anxiety, self-perceived public speaking competence, and teacher immediacy for students who were classified as high, medium or low in terms of communication apprehension. Teacher immediacy is defined as communication behavior that reduces the physical and/or psychological distance between people (e.g., the use of humor, praise and feedback). Students enrolled in communication courses (n = 97) volunteered to participate in the study. The instruments used were the Personal Report on

Communication Apprehension (McCroskey, 1970), Ellis' (1995) Self-perceived Public Speaking Competence Scale (SPPSC scale), and written tests of public speaking anxiety and teacher immediacy.

The main hypothesis was that there would be a negative relationship between public speaking anxiety and public speaking competence at all three test points (pre-course, mid-course, and post-course). This hypothesis was supported at all three time intervals. The second hypothesis was that high communication apprehension (CA) students would perceive less improvement over the course of the semester than lower CA students. This hypothesis was not supported. In fact, the opposite was found: students with high CA reported more improvement than students with lower CA. The third hypothesis investigated the relationship between teacher immediacy and changes in public speaking apprehension for students in different CA groupings. A moderate correlation (0.46) was found between the teacher immediacy scores and a decrease in public speaking apprehension for high CA students, but not for the other CA groupings.

The results of Ellis's (1995) study support previous research that indicates perceived speaking competency is a good predictor of public speaking anxiety. In addition, the findings regarding teacher immediacy and public speaking anxiety provide evidence that teachers of public speaking should be aware of their immediacy behaviors, and try to increase their immediacy behaviors when dealing with high CA students.

Campbell (1995) performed an experiment to assess the effectiveness of Ellis' (1995) Self-perceived Public Speaking Competence (SPPSC) scale and McCroskey's (1970) Personal Report on Communication Apprehension. Students in the experimental group ($n = 56$) were enrolled in a public speaking course. Control group A consisted of students ($n = 71$) who had previously taken a public speaking course, but were not currently enrolled in such a course.

Control group B consisted of students (n = 62) who had not taken, and were not currently taking, a public speaking course. The hypothesis was that the experimental group would show an improvement in their scores on the SPPSC scale and the PRCA from pre- to posttest. Participants in the control groups were not expected to show any change from pre- to posttest. All participants completed the instruments before the beginning of the public speaking course and again after the course ended.

The results of the study were interesting. All participants improved their score on the PRCA, an assessment of their communication apprehension. This overall decrease in apprehension was not expected, as only the experimental group had received any organized training and practice between the pre- and posttests. The second part of the hypothesis, that only experimental participants would show an increase in their scores on the SPPSC scale, was also not supported. Both experimental and control group B participants increased significantly in their scores on the instrument, while control group A participants showed no increase or decrease.

Campbell (1995) admits that these findings raise concerns about either the validity of the instruments or of the research project that produced these results. Evidence that there may have been internal validity problems with the research design is presented. Specifically, students were giving presentations in classes other than the public speaking course. This practice and experience may have contributed to the students not actively engaged in improving their public speaking skills (the control groups) reporting essentially the same improvement in their communication apprehension and public speaking competence as the experimental group.

Carlson and Smith-Howell (1995) provide an assessment of the reliability and validity of a select group of public speaking evaluation instruments. The participants were categorized into three groups with varied levels of experience in teaching and evaluating speeches:

communications instructors (n = 18), graduate students in communications with no teaching experience (n = 19), and undergraduate students (n = 21). The instruments used were three commonly used evaluation forms (labeled A, B and C), each of which focused mainly on speech content issues, but did include some performance factors. The communication instructors were also allowed to use their own personal instruments in place of the third instrument. The performances to be rated were two videotaped performances by students; one rated by "experts" as an A-level speech, the other a C-level speech.

Each participant watched the prepared videotaped speech performances twice, rating the performance each time with a different rating form. There were two different prepared videotaped speeches; thus each participant gave four ratings, two each for each speech. A modified Latin-square type design meant that some participants used the same form more than once, while others used a different form each time they evaluated a speech.

The results of the study revealed that the three commonly used evaluation forms showed a high reliability across forms, time intervals and raters. Carlson and Smith-Howell continue with an interesting discussion of the order-effect on the scores across instruments, which are minor in any event. Equally importantly, the results indicate that the three instruments exhibited construct, content and predictive validity. Construct validity was demonstrated by the fact the instruments measure the established elements of speech (content and delivery). Predictive validity was exhibited by the fact that the evaluator ratings of the speeches corresponded to "expert" ratings of the speeches (the "better" speech was consistently rated higher than the other).

Carlson and Smith-Howell's claim that content validity was demonstrated by virtue of the fact that the raters consistently rated the A-level speech higher is arguable. This result indicates

that the participants rated the "better" speech higher, and provides evidence of predictive validity, but does not address the issue of the content of the instruments. Content validity refers to whether the items on the instrument accurately represent the performance domain of interest (Crocker & Algina, 1986), in this case public speaking. Content validity was not demonstrated in this study.

In summary, the results of Carlson and Smith Howell's (1995) study revealed that the three instruments under examination demonstrated good reliability across raters, time intervals and forms. This bodes well for the reliability of similarly developed instruments. The validity results are encouraging, although not conclusive. Although the construct and predictive validity appear to be acceptable, more information and analysis of the content validity would seem prudent.

Newburger, Brannon and Daniel (1994) investigated whether viewing videotape of their own speeches would reduce participants' self-rated public speaking apprehension. An experimental group (n = 56) of undergraduate students enrolled in an introductory public speaking course completed the public speaking apprehension instrument before and after the course. During the course they performed four speeches that were videotaped and reviewed after each performance by the class as a whole. Control group participants (n = 112) were enrolled in the same program and completed the same instrument pre- and post-program, the only difference being that they were not videotaped. The hypothesis was that participants in the experimental group would exhibit reduced public speaking apprehension as a result of reviewing the video of their performances.

The results of the study did not support the hypothesis. In fact, the control group participants exhibited a significant reduction in their public speaking apprehension scores, while

experimental group participants did not. The authors conclude that, contrary to their hypothesis, the review of the video after presentation in their public speaking course actually inhibited the reduction of public speaking apprehension experienced by control group participants. Their explanation revolves around the method of review – the entire class and instructor were present when the speech was reviewed. This group review seems to have served to heighten, not reduce, the participants' public speaking apprehension. One explanation for this finding is that the participants felt increased anxiety because they were not the only ones to judge the performance. Rather, the entire class and instructor observed, and judged, the performance as well.

An interesting point that was not addressed by Newburger et al. (1994) was the large difference in the sample sizes. The fact that there were twice as many control group participants as experimental immediately raises concerns about the results. The authors failed to provide any statistics other than summaries of the analyses, which prevents further examination of the data.

A series of articles by Morreale (1990; Morreale et al., 1991; Morreale & Hackman, 1994) addresses the development of a valid and reliable speech evaluation form, to wit, The Competent Speaker: Eight Public Speaking Competencies and Standards / Criteria for Assessment. These eight competencies were developed from the literature on communication competency in response to a call for a standardized and validated evaluation tool for national distribution. Of the eight competencies identified, five addressed content issues of the speech (e.g., topic selection, appropriateness for the audience, supporting material, organization of speech, and use of language) while three addressed performance issues (use of vocal variety, pronunciation and articulation, and physical behaviors). The evaluation instrument consisted of behavioral descriptions for each of the eight competencies, with an assigned value (1, 2, or 3) to the level of achievement exhibited by the speaker.

Content validity is claimed based on the extent of the literature review that informed the development of the instrument, and the reputation of the committee that developed the instrument. Reliability, both between raters and between competencies, has been examined and shown to be favorable.

The main problem with the instrument is that it uses subjective ratings of observed behaviors, despite its claim to objectivity. In some areas, this does not present a major obstacle. For example, Competency 2 refers to whether the speaker provides a thesis statement in the speech. The three levels of rating this are a) speaker provides a clear statement of thesis, rated as Excellent and assigned a value of 3 points; b) thesis provided but lacking substance, focus or clarity, rated as Satisfactory and assigned a value of 2 points; and c) no identifiable thesis, rated as Unsatisfactory and assigned a value of 1 point. These categories seem clearly defined and fairly easily determined.

In other areas, however, there are problems presented by the measurement scale provided. For example, Competency 8 addresses the physical behaviors that support the message. An Excellent rating is assigned when the speaker demonstrates excellent posture, gestures and facial expressions that support the message; a Satisfactory rating for satisfactory use of posture, gestures and facial expressions that do not interfere with the message; and Unsatisfactory for unsatisfactory use of posture, gestures and facial expressions that are incongruent with the verbal intent, and that distract the audience to the point that the speakers message is lost. These categories seem to be less well defined, with considerable overlap between the categories.

The competencies seem to adequately address the many facets of speeches that should be addressed in an evaluation instrument. Unfortunately, despite the reliability and validity claims,

there seem to be problems with the scale developed from the competencies. For unreported reasons, the authors decided to assess each competency with only one scale. Perhaps the use of more subscales to directly assess the individual components of each competency would yield useful results.

Hugenberg and Yoder (1994) present the argument that any communication evaluation system that fails to take into account both the speaker and the listener fails. The presence of a "competent speaker" is nullified by the absence of a "competent listener". Although they do not provide solutions for the failure of Morreale (1990, Morreale et al., 1991) to include the receiver of the message in the assessment of competence, they do provide other criticisms of the evaluation instrument. Their criticism focuses on three areas: the ability of evaluators to discriminate between the competency levels described in the instrument, the generalizations from the rater to the audience, and the cultural narrowness of the instrument.

In reference to the discrimination of the competency levels described in the instrument, Hugenberg and Yoder attack the lack of definition for vague terms such as "above average", "high", and "exceptional". They argue that the terms are ill-defined, when they are defined. For example, they state, "there are no universal standards for appropriateness, much less 'exceptional' appropriateness" (p. 10).

The second issue addresses the generalization from the rater to the audience. The argument is based on their claim that the assessment should involve the speaker and each audience member (the receivers of the message), not just one person's view of the audience's perceptions. In this case, no solutions are offered as to how to implement this recommendation, nor do Hugenberg and Yoder address the issues of reliability and validity associated with these numerous evaluations.

The final issue is concerned with the cultural narrowness of the competencies. Here, Hugenberg and Yoder directly attack the validity of the competencies in terms of their relevance and adaptability across a diverse cultural landscape. They claim that at least some of the competencies reflect a decidedly biased western approach to communication. For example, appropriate dress is something of an enigma even within different parts of North America according to Molloy (1971, 1977). Another example raised is the issue of eye contact. In North America, eye contact is considered to be essential to good communication. In many cultures, however, it can be both negative and insulting. Again, no solutions are offered for addressing these concerns.

The essence of the argument presented by Hugenberg and Yoder is that communication competence is both situation- and participant-dependent. It is not possible to be a competent communicator, only to have a competent communication. Because there are so many context-dependent variables, it is only possible to learn behaviors and skills that may enhance our probability of having a competent communication, from the point of view of both the speaker and the listener.

In summary, the literature on public speaking focuses, for the most part, on reducing the anxiety of the performer. Efforts to develop guidelines for assessing public speaking performance have been critiqued for a lack of specificity. Of primary interest to the present study is the fact that no studies were found that addressed the impact of the performer on the audience of the presentation. Hugenberg and Yoder (1994) address this issue in theoretical terms. Unfortunately, they offer no answers to the question of how to assess the impact on the audience. This may be a result of the difficulties encountered in developing guidelines for the assessment of public speaking.

Two other areas of interest were conspicuous in their absence. No research was found that explored performance self-efficacy with public speakers. Also, no research was found that explored the impact of evaluative content of feedback on performance in formative appraisal of public speakers.

This section was included in order to provide context for the public speaking component of the current study. The review of the literature provides motivation for the development of a valid and reliable measure of performance self-efficacy for use with public speakers.

2.4 Trends in Fields Related to Personnel Evaluation

It seems prudent to note recent developments in fields related to personnel evaluation that seem to impact on the current study. Thus, we will briefly review some of the research in the areas of student assessment, counselor evaluation, and program evaluation focusing on the issue of who provides the evaluation of the data. In addition, a brief review of literature related to self-reflection is provided, in order to provide some perspective on this issue for discussion in the final chapter.

2.4.1 Student Assessment

Student assessment is an area that has embraced the concept of self-evaluation, at least to some extent. Traditional paper-and-pencil tests of factual knowledge have come to be replaced, or at least supplemented, by more inclusive, less objective forms of assessment. These styles of assessment include self-evaluation and portfolio assessment (McMillan, 1997).

McMillan (1997), in a textbook on classroom assessment for effective instruction, reports that classroom, or student, assessment has experienced a shift from "objective" testing that occurs at the end of instruction towards more subjective assessment that occurs during the

instruction. Traditional objective testing typically requires the student to demonstrate the ability to recall facts in isolation.

More recently adopted forms of assessment strive to assess the ability to construct meaning from factual knowledge, as well as the ability to apply factual knowledge in contextually diverse situations. These forms of assessment also attempt to address the need to prepare students for increasingly complex workplaces (Earl & Cousins, 1995; McMillan, 1997; Stiggins, 2001). McMillan also notes that the emphasis on assessment is shifting away from the instructor providing "the correct answer." The emphasis is moving towards the students, in groups and individually, determining the validity of their answers, with guidance from the instructor. In addition, these new assessment processes allow the students to explore the process by which they arrived at their answer.

While embracing the change towards more subjective forms of assessment, and the shift towards students helping judge the value of their work, McMillan (1997) notes that there is still a need for objective forms of testing. McMillan stresses that the form of assessment must be matched to the purpose of that assessment. He is not advocating allowing students to determine their entire mark in a course; rather he points out that self-assessment techniques can be used to augment more traditional forms of assessment. In fact, they may be more appropriate in some cases.

Popham (2000), in a textbook on educational measurement, also discusses the use of self-assessment, specifically portfolio assessment. Popham defines a portfolio as "a systematic collection of one's work" (p. 299). He notes that, while the use of portfolios is a fairly recent development in education, it has long been used in other fields. In fact, in some fields portfolios are the main method by which skill is demonstrated. These fields include art, photography,

architecture, and modeling. An important characteristic of portfolios is that they must be constantly updated, to reflect the owner's development.

Many contemporary writers on student assessment and evaluation emphasize that the key benefit, and objective, of portfolio assessment is the enhancement of the students' ability to evaluate their own work – self-evaluation (Earl & Cousins, 1995; McMillan, 1997; Popham, 2000; Stiggins, 2001; Wilson, 1996). During the portfolio conference, wherein the instructor and student discuss the portfolio contents, the instructor should encourage the students to supply an appraisal of their own work. Moreover, students should be encouraged to compare their recent work with earlier work, in order to enhance their sense of accomplishment and to practice their appraisal skills.

There are drawbacks to portfolio assessment, as well as other forms of self-evaluation. First, there is the problem of generalizability. Even if the guidelines for evaluating the portfolio are standardized, the criterion used to assess those standards by each individual probably won't be equivalent. Second, and probably most importantly, is the issue of time. Portfolio assessments and conferences take time, and lots of it, to do properly.

This brief review has demonstrated that the concept of self-evaluation is alive and seen as integral to the teaching and learning process (McMillan, 1997; Popham, 2000) as is the extent to which students can develop their capacity to apply what they have learned. The extent of its use and the nature of its use in the classroom assessment literature are beyond the scope of this paper. What is important is that the capacity for self-critique is now understood to be pivotal in the learning process and that this corroborates the position of Tracy and McNaughton (1989).

2.4.2 A Special Case of Student Assessment: Counselor Evaluation

Next, we briefly review some relevant literature from the field of counselor education that has looked in systematic ways at the concept of feedback. Carney, Cobia and Shannon (1996a) note that there is very little research available on the evaluation methods used with counseling students. One area that has seen growth in the literature is that of portfolio assessment. They catalogue two important components of portfolio development that enhance its usefulness in evaluation. First, they recommend that students be allowed input into at least some of the contents of the portfolio. This input enhances the students' engagement in the evaluation process. Second, they recommend including self-reflection components in the portfolio. These components are focused on the student's self-evaluation of their performance, progress and thought processes. These components may include the use of journals and practicum evaluations. Other, more specific self-evaluation tools can be customized by instructors, such as situational questions. The main benefit of the portfolio, in Carney et al.'s (1996a) view, is that it involves the student in the evaluation process. A major drawback, though, is the amount of time required to successfully develop a portfolio assessment. It is a very time consuming process, both for the student to develop the contents and the supervisor to evaluate the contents.

Baltimore, Hickson, George and Crutchfield (1996a) also recommend portfolio assessment for documenting growth with counselors-in-training. One of the main purposes of the portfolio is to systematically present information for self-reflection on performance, as well as to serve as a record of growth.

Interestingly, Baltimore et al. (1996a) recommend that the portfolio serve as both a formative tool for the student and a summative tool for the supervisor. This is most interesting because it highlights the formative / summative distinction made earlier and presents a dilemma.

If the appraisee includes only material that makes them appear competent, of what summative value is the portfolio? And if the students know that the portfolio will be used for summative purposes, why would they include material that highlights their weaknesses?

Alschuler (1996) is one who does not like the idea of portfolio assessment. He prefers the term portfolio education. Alschuler points out the relative strengths and weaknesses of portfolio methods from the field of education, as there was no equivalent literature in counseling to review. While he concludes that there are both advantages and disadvantages to portfolio assessment, his argument lies in the realm of summative evaluation. The well-documented difficulties in establishing reliability in portfolio assessment in education preclude their use in the summative evaluation of counselors, in accordance with the ethical standards of the profession. Alschuler, however, does support their use for purely educational (formative) purposes.

Carney et al. (1996b) and Baltimore et al. (1996b) take issue with Alschuler's (1996) argument against the use of portfolios in summative evaluation. Carney et al. (1996b) agree with Alschuler (1996) that portfolios can and should be an integral part of the multiple assessment methods used in counselor evaluation. Their argument is that the validity and reliability of portfolios is established by the agreement of well-qualified evaluators using well-developed evaluation criteria.

Baltimore et al. (1996b) take a different tack arguing that the reliability of traditional forms of assessment may need more careful examination. They also note that the lack of literature on the use of portfolios in counseling evaluation should not preclude their use. Perhaps, they suggest, some members of the establishment content in their belief that they have perfected counselor training are threatened by the emergence of new methods for assessing this training.

Hart (1994) presents developmental guidelines for implementing effective supervision of counselors. Beginning counselors require a high degree of support from the supervisor, but over time, this reliance on the supervisor for guidance is reduced. More advanced supervisees can begin to select behaviors and topics for discussion in the feedback sessions. The supervisees take progressively more responsibility for the appraisal as they grow in skill. At this point, the supervisor becomes more of a colleague or consultant than a teacher.

Benshoff (1994) distinguishes between the terms peer consultation and peer supervision with counseling trainees. In Benshoff's view, peer consultation is a more appropriate term for the process, as the peer is not in a supervisory role. Rather, the peer acts in the role of a consultant, whose suggestions and ideas may be rejected or accepted as the appraisee sees fit. The term peer supervision implies that the peer is in a form of expert / novice relationship, which would be contrary to the objective of obtaining critical and supportive feedback. The benefits of peer consultation include decreased dependence on "expert" supervisors and increased responsibility on the part of the supervisees for assessing their own skills and their colleagues.

In this form of peer consultation, students work in dyads or triads, and utilize traditional supervision activities like goal-setting, video review and case consultation. The difference here is that the emphasis is on helping each other achieve their pre-determined goals, rather than evaluating each other's performance. The peer consultants take greater responsibility for providing critical feedback and support to each other. Benshoff also notes that the added responsibility within a group setting fosters enhanced understanding of how to evaluate themselves. This improvement has been documented by Benshoff and Paisley (1993), and is apparently a byproduct of collaborating to make the process work.

Hosford and Johnson (1983) report on their experimental study on the use of self-observation, self-modeling and practice without video feedback. The participants in the study were counseling students, and the study examined the impact of these forms of training on the students interviewing behaviors. Participants in the self-observation condition (n = 3) reviewed video of their own performance in five practice counseling sessions. This procedure has been shown to be differentially effective in improving performance depending on how successfully the behaviors are performed on the video. Participants in the self-modeling condition (n = 4) reviewed videos of their own performance. The videos were edited so that only positive experiences or successfully performed behaviors were included; non-successful or negative behaviors were removed. Participants in the final condition (n = 4), practice without video feedback, received written feedback from supervisors on their behavior. The dependent variables of interest varied for the three conditions, but were the frequency and duration of behaviors identified by supervisors as distracting and clearly communicated to the participants in each condition. Examples of these behaviors include lack of eye contact, word repetition, fidgeting, and the like. The stated objective was to reduce the frequency and duration of these behaviors to zero.

The results of the study indicated that participants in all three conditions reduced the frequency and duration of the behaviors dramatically in almost all cases. The only procedure that extinguished the behaviors completely, however, was the self-as-model condition. They conclude that self-as-model forms of video feedback, although harder to deliver, are more effective in improving performance. Their rationale for this argument is that the provision of video feedback that contains negative behaviors or failures can cause some trainees more anxiety and reduce the effectiveness of the video feedback in improving performance.

The preceding review of the literature on counseling education has revealed that there is little research available on the evaluation methods used with counseling students (Carney et al., 1996a). There is a trend towards using portfolios for formative evaluation of counseling students, although this movement is not without its critics, who focus their sights on the issues of validity and generalizability. Even the proponents of portfolios admit that the use of portfolios requires a great commitment of time by both appraisees and supervisors. Another development in counselor education is the use of peers as consultants to assist in skill development, instead of the more traditional supervisors (program instructors). Empirical work on the use of video-feedback reveals that self-modeling is marginally more effective than either regular video-feedback or written feedback without video.

2.4.3 Program Evaluation

That self-evaluation is seen as integral to student evaluation and counselor training is beyond question. Similarly, the field of program evaluation has seen a shift from an objective, judgmental format to a more responsive, stakeholder-service approach wherein those with a vital interest in the program, and the evaluation of the program, are included in the evaluation process as more than just sources of data.

Collaborative evaluation has been defined as trained evaluators working with program practitioners and other stakeholders in the application of principles of systematic inquiry to solve practical social or organizational problems (Cousins, Donohue & Bloom, 1996). Collaborative, participatory, and empowerment forms of evaluation have taken many different forms in the last 30 years. For our purposes, we will focus on the extent to which those responsible for program management and implementation participate in the process, specifically the decision-making aspects of the evaluation.

Stakeholder based evaluation may be rooted in Stake's (1976) original vision of responsiveness in evaluation. This early form of collaborative evaluation involved stakeholders (defined as persons with an interest in the object of the evaluation, such as participants, beneficiaries, etc.) in the evaluation process on a limited basis. Previously, stakeholders were considered to be sources of data only. In stakeholder-based forms of evaluation, stakeholders are typically involved in consultative tasks such as scope setting and interpreting findings on a local level, rather than more complex tasks such as data analysis, instrument development, or technical decision-making in evaluation.

More recently Patton (1994) describes "developmental evaluation," with a focus on the evaluation process and its relationship to program practitioners. Developmental evaluation is the essence of formative evaluation, wherein the evaluator is actually instrumental in assisting in the development of the program. The roles of evaluator and stakeholder may be blurred, as both focus on all aspects of the evaluation. In this scenario, the decision-making (i.e., evaluative judgments about the program) is shared between the evaluator and the stakeholders.

Patton (1997) also talks about decision-making in his book on utilization-focused evaluation. His emphasis is on the development of evaluation results that the stakeholders will use, and one way of ensuring this is to involve the stakeholders in the process at every level. While Patton clearly does not advocate sloppy evaluation practice, he is interested in evaluators advising stakeholders, guiding them towards results and conclusions but not imposing external views. This sharing of the control of the evaluation extends to all levels, including decision-making.

Participatory evaluation, as conceptualized by Cousins and Earl (1992), advocates joint participation in most evaluation activities. The reason for this collaboration is to enhance the

utilization of the results of the evaluation. Decision-making is one of the many traditionally evaluator-controlled activities that are shared in this approach to evaluation.

Cousins (in press) traces the development of what he terms participatory evaluation, wherein trained evaluators interact on an intimate level with stakeholders to create evaluation knowledge collaboratively. Cousins distinguishes between two forms of participatory evaluation: practical and transformative. The difference between these forms of participatory evaluation lies with the objective of the process. The aim of practical participatory evaluation is to inform decision-making, be it program, policy or organizational. The aim of transformative participatory evaluation is to increase the participants' understanding of the context for the object of evaluation, and to foster the amelioration of social inequity and chance self-determination through participation in the evaluation process. In Cousins' words, the aim "is to empower people, through participation, in the process of constructing their own knowledge, and their understanding of connections among knowledge, power and control" (p. 7).

Although the thrust of Cousins' chapter is to enhance understanding of evaluation utilization in participatory forms of evaluation, there is overlap with our interests: the control of decision-making within the evaluation. Cousins reports that proponents of practical and transformative forms of participatory evaluation believe that control of decision-making should be shared between the evaluator(s) and the stakeholders. Care must be taken, however, to avoid transferring control of the evaluation to the stakeholders at the expense of the quality of the evaluation.

On a related note, Cousins et al. (1996) report the results of their large-scale international survey of evaluation practitioners throughout North America, indicating that the evaluators both

believe in and perform evaluations where they maintain control of the process. Other authors go further, advocating participant control of the evaluation process (Fetterman, 1994).

The reason for this finding, according to Cousins et al. (1996), is that the majority of the respondents in their survey reported on evaluation projects of a summative nature, where a judgment of merit or worth was expected. Evaluators may have retained control of decision-making in order to reduce the impact of stakeholders' self-serving bias. Had the sample included more respondents involved in formative types of evaluation, the results may well have been as expected; that is, indicative of more shared control of the process.

Patton (1997) explains that the essence of collaborative forms of evaluation (within this umbrella may be included all the types of evaluation reviewed above) is an approach that emphasizes the role of the evaluator as an enabler of the process, rather than an independent arbiter of good and bad. This enabling includes sharing the decision-making power with the participants. In Patton's words, "what these approaches have in common is a style of evaluation in which the evaluator becomes a facilitator, collaborator, and teacher in support of program participants and staff engaging in their own evaluation. While the findings from such a participatory process may be useful, the more immediate impact is to use the evaluation process to increase participant's sense of being in control of, deliberative about, and reflective on their own lives and situations" (1997, p. 98).

In summary, we can see that a shift from objective, evaluator-controlled decision-making towards more subjective, collaborative decision-making has been seen in the field of program evaluation. This shift is grounded in the principle that program practitioners need to become judges of their own programs. They need to develop their capacity for critiquing such programs. Again, we have strong corroboration of current theory in formative performance appraisal that

advocates the provision of descriptive, rather than evaluative, feedback to appraisees. Both seek to enhance the capacity for self-critique and evaluation of those involved in the evaluation process.

2.4.4 Self-reflection[1]

Schön (1983, 1987) presents a potentially useful framework for understanding the impact of different forms of feedback on appraisees through his work on self-reflection. Self-reflection has been defined by Bullard (1998) as the process of learning from one's experience. This definition is based on the concept of reflection on practice developed by Schön. In the following section, the early work of Schön is presented. Next, some criticism of Schön's work is reviewed, and finally, a recent empirical study relevant to the present study is presented. The purpose of this review is to provide an overview of the concept of self-reflection in the context of the current study. The review provides context for the discussion of results later in the document.

Argyris and Schön (1975) presented the concept of reflection on practice as a model of professional knowledge that ran counter to much of the theory developed through the application of the positivist, technical-rational model that dominated inquiry for centuries (see Schön, 1983, for a more thorough history of this matter). Argyris and Schön (1975) assert that professionals, or technical experts, possess both 'espoused theories' and 'theories in use' that guide their behavior. 'Espoused theories' are those explanations for their behavior that people claim allegiance to when asked to explain their behavior. 'Theories in use' are the true, usually unconsciously known theories that actually guide behavior.

Argyris and Schön (1975) further explain that 'theories in use' can only be constructed from careful observation of a person's behavior and attempting to discern patterns. This is true

---

[1] It should be noted that this section was added after the current study was complete, and thus, unfortunately, was not used to inform the research. The section is included in order to provide context to guide future research.

because the 'espoused theory' and the 'theory in use' for the same behavior may in fact be incompatible with each other, unbeknownst to the individual. These theories are all based on theories of action that relate behaviors to outcomes. That is, if a certain behavior occurs in a certain environment, a certain outcome will result. The environment, or context, within which the behavior occurs, is extremely important to the theory of use. The issue of the importance of context represents a radical departure from the positivist models of professional knowledge, which frame knowledge as distinct from context. This context-free knowledge can be applied with little regard for either the context from which the knowledge was drawn or the context within which it is to be applied.

Argyris and Schön (1975) also present the concepts of single- and double-loop learning. Single-loop learning involves basic cause and effect relationships: what behavior(s) will produce what outcome(s) within a given context. This is considered to be low-level learning. Double-loop learning involves questioning the context that frames single-loop learning. Further, double-loop learning involves questioning the utility of outcomes, and their relative value, rather than simply determining how to achieve those outcomes. This higher-order learning, double-loop learning, is the desired form of learning for reflective practice. In the present study, single-loop learning would be characterized by insights from participants regarding methods they could use to produce specific outcomes within their presentations. Double-loop learning would be characterized by insights about the importance and desirability of various outcomes of their behaviors within their presentations, as well as the impact of these behaviors on the audience for their presentations.

Schön (1983) presents a refinement of his earlier work with Argyris, moving beyond theories of action and use to the concepts of reflection in and on action. By reflection in action,

Schön refers to the practice of deciding on behaviors while in the midst of action. This reflection in action can consist of conscious analysis of possible behaviors and their expected outcomes, a weighing of the desirability of these various outcomes, and a decision as to which behavior to perform, all in real-time. This process can be repeated several times in the course of one 'event', resulting in multiple comparisons of behaviors and outcomes, producing sequenced behavior patterns that may be automatic (well-practiced) or carefully thought out during the 'event'. In the present study, this could be characterized by insights into the ongoing decision-making that participants make during their presentations. For example, the presenter may make a conscious decision to perform a portion of their presentation differently based upon their observation of the audience's reaction to earlier portions of the presentation.

By reflection on action, Schön (1983) refers to the practice of reviewing past behaviors and determining whether the resultant outcomes were optimal. In addition, by reflecting on action we can explore alternative behavior patterns that might produce more optimal outcomes. It is important to note that immediacy of events may be the determining factor in whether an individual can engage in reflection in or on action. Schön uses the example of a baseball player in a close game and a teacher faced with a student's behavioral problem. The baseball player may have a very short amount of time in which to determine his next behavior (e.g., which pitch to throw), while the teacher may be able to postpone a behavioral decision because she has a 4- or 8-month term in which to deal with the student.

Schön (1987) notes that performers can be coached in how to improve their performance, but they can't necessarily be taught how to improve. "Students learn by practicing the making or performing at which they seek to become adept, and they are helped to do so by senior practitioners who ... initiate them into the traditions of practice" (p. 16). This statement reflects

Schön's view that performers must be allowed to determine for themselves the relationships between the means they employ and the outcomes they desire. It should be noted that Schön does limit the scope of his framework to understanding forms of practices that are considered more idiosyncratic or intuitive, such as artistry and education.

There has been criticism of Schön's work. Grimmet (1989) feels that Schön's work, in comparison to other work on reflection in education, is a general model that may be of use in teacher education. Grimmet states that Schön's work offers an explanation of who or what a reflective practitioner is, but not how they are reflective. Grimmet calls for a more detailed explanation of how the process can work in a specific situation, rather than the more generic model of what the process seems to be across various settings as offered by Schön.

Perhaps this desire for concrete guidelines for implementing reflective practice is antithetical to Schön's thrust, which is to describe how the process seems to unfold. The constructivist epistemology underlying Schön's work argues against providing a specific generalizable framework for use in different settings.

In the realm of teacher education, Grimmet and Crehan (1989) report on their case study of a supervisor and teacher engaging in formative appraisal. The intent of the study was to examine the conditions that impact the teacher's ability to reflect on practice within the appraisal process. The main issue that arose was that of who identifies the issue and frames the context for examining it within the appraisal process. According to Schön (1983), "reflectively transforming experience occurs when teachers [appraisees] show evidence of naming the things they will attend to and framing the context they will use to attend to them" (p. 218). Grimmet and Crehan's results provide support for this assertion. They found that collegiality between the supervisor and teacher is an important component of the process, but the most important factor

was that the appraisee (teacher) controlled the scope and context of the process. That is, the appraisee was allowed to identify the issue(s) to be dealt with in the appraisal process, as well as determining the context for dealing with the issue(s). Unfortunately, as demonstrated by Grimmet and Crehan, these optimal conditions are not easy to achieve.

Harris (1989) also raises some concerns with Schön's ideas. From a practical standpoint, Harris notes that there are no support structures in place for facilitating reflection. Support from others, or at least the opportunity to discuss ideas with others, seems to be an important component of reflection. Harris's second criticism addresses the dichotomy Schön seems to set up between extant knowledge (e.g., the work of Dewey, Piaget, etc.) and reflection-in-action. Harris believes that this extant knowledge can inform reflection in tandem with personal experience, a link that does not seem to be made in Schön's work. In addition, the importance of recording and disseminating experiences so that others can learn from them is also seen missing from Schön's work, according to Harris (1989). This recording and dissemination (in academic journals, etc.) will help ensure that advances made will not be lost because of poor communication.

Peters (1991) presents a systematic approach to developing self-reflective capacity in a cyclical model that includes four logically related steps. The first step is to describe the problem in detail. The second step is to analyze the behavioral cause(s) of the problem, and the third step is to theorize alternative actions that may produce more optimal outcomes. The fourth and most important step is to select one of these theories and act on it. The outcome of this action provides feedback to the initial step of the process – determining the problem. If the initial problem is no longer a concern, a new problem can be selected. If there are still issues to be resolved, then there is more information to feed back into the cycle. Peters (1991) states that applying this

systematic approach to reflection provides the link needed to integrate thought and action: "Thought and action are thus integrated through reflection" (p. 89). This cyclical model of reflective practice is similar to the growth-oriented appraisal process outlined earlier (e.g., Cogan, 1973; Goldhammer, 1969; Cousins, 1995). In contrast to the interactive model of clinical supervision, individuals engaging in the self-reflective process can complete multiple iterations of the process as quickly or as slowly as they desire. Another advantage to the self-reflective process is that no disclosure of weaknesses to others is required.

Bright (1996) suggests that current reflective practice is not very effective. In Bright's view, the purpose of reflection is to raise the performer's awareness of factors that influence behavior choices. Bright also emphasizes the importance of involving others in one's reflective practice. According to Bright, others may be able to provide insight that the performer cannot access on their own: "this information may be hidden from [the performer] but may be observable by others, which suggests the potentially important role of others in the reflective process and its objective of increasing awareness" (p. 171). This point addresses a potential weakness of the cyclical model detailed by Peters (1991): Is external feedback important to the reflective process? Bright seems to think so.

Bright also notes that the very nature of reflective practice produces knowledge that is context-bound. Thus, knowledge developed in one context may not be generalizable to other contexts. This forms the basis for the statement that, although reflection can result in both single- and double-loop learning, double-loop learning represents the more important level of learning. The tightly-focused, 'means to achieve an outcome' type of learning represented by single-loop learning is more context-bound and thus less important. The less tightly focused, bigger-picture type of learning represented by double-loop learning, where the focus is on the value and utility

of the outcome, is regarded as more fundamental learning. In Bright's words, "it holds the promise of not just improving practice, but of changing it in fundamental ways" (p. 183).

Marita, Leena and Tarja (1999) attempted to demonstrate empirically that nurses could improve their interviewing skills through reflective practice. Nineteen female nurses participated in the study, in which they were videotaped performing an interview with a real patient, and then self-evaluated their own performance using the video as a stimulated recall device. The nurses performed both oral and written evaluations of their performance. The nurses received information on effective communication skills and reflective practice in both written form and in the form of a presentation by an instructor before the study began. The nurses had access to the written information throughout the study. The nurses repeated the same procedure 6 months later for comparative purposes. Expert raters analyzed the written evaluations and transcripts of the oral self-evaluations to assess the level of self-reflection evident in the evaluations. The raters assessed the self-evaluations in terms of Mezirow's (1981) seven levels of reflectivity. The first 4 levels of reflectivity represent levels of consciousness about oneself, while the other 3 levels represent critical consciousness. Critical consciousness refers to reflection on the other four levels of consciousness, and is meant to represent a deeper level of reflection.

The results of the analysis of the first session revealed that the nurses provided much more detailed written analyses than verbal analyses. The nurses also had no difficulty identifying problem behaviors, but that was the extent of the reflection being done. Essentially, the nurses raised their awareness of the problems they were encountering, but there was little evidence that they were reflecting on this new awareness (the deeper levels of reflection). The analysis of the second session results revealed that, while the nurses felt that they had improved their

interviewing skills, and they improved modestly in their ability to identify problem behaviors from the first session, there was again little evidence of deeper levels of reflection.

The results of this study indicate that self-reflection aided by video, as implemented by Marita et al. (1999), was effective in increasing the participants' lower-level reflectivity (as defined by Mezirow, 1981). It was not effective in increasing the participant's higher-level reflectivity, or critical consciousness. One problem with the methodology employed was that there was no explanation of what expertise the raters had in assessing the level of reflectivity evident in the participants' self-evaluations. Obviously, this expertise is crucial to the interpretation of the findings. There may have been evidence of deeper reflectivity in the participants' self-evaluations that went unnoticed.

The results reported by Marita et al. (1999) indicate that the provision of video feedback to participants stimulates lower-level self-reflective ability. This form of feedback did not seem to enhance higher-level reflectivity. The methodological problems outlined above indicate that it would be prudent to attempt to determine whether higher-order reflectivity is enhanced in the reflective process that participants are involved in.

In summary, Schön's (1983, 1987) conception of self-reflection provides a useful framework for understanding the impact of different forms of feedback on appraisees involved in growth-oriented appraisal processes. Essentially, the provision of descriptive feedback to appraisees could be seen as a means of providing the appraisee with an external view of their behavior, which has been shown to be important in self-reflection. This would allow the appraisees to assess their own behavior, or self-reflect. The importance of the appraisee being in control of the process, as recommended by Tracy and McNaughton (1989), is reinforced by the findings of Grimmet and Crehan (1989).

The results of Marita et al. (1999) also support the use of this framework for enhancing our understanding of the different forms of feedback. The fact that the appraisees could improve their ability to self-reflect, at least at the lower levels of reflectivity (Mezirow, 1981), with the use of video feedback supports the idea that appraisees can learn to improve their ability to self-reflect when provided with descriptive feedback. In addition, the results showed that performers improved their self-assessed performance through video feedback.

2.4.5 Summary of Trends in Fields Related to Personnel Evaluation

The previous reviews of the literature with respect to student assessment, counselor evaluation and program evaluation, have revealed trends that bear discussion in the context of constructive feedback and formative appraisal. In student assessment and program evaluation, there is a marked trend towards increasing the extent of the participation of the objects of evaluation (e.g., the students, program participants). This increased participation includes, but is not limited to, involvement in the actual evaluation of the data. This trend was also seen in the counseling evaluation literature, where the use of portfolios for assessment purposes is coming to be accepted. The objective of using portfolios according to Carney et al. (1996a) is to involve the counseling student in the evaluation process. Again, we see that the importance of the development of the capacity for self-critique is supported.

The section on self-reflection was included to provide context for the discussion section of the document. The concept of self-reflection is used in the discussion of directions for future research into the evaluative content of constructive feedback.

The purpose of including this section is to show that the trend towards self-evaluation is not an isolated event. Rather, it can be seen in several related yet distinct fields, from student evaluation to program evaluation. The existence of this trend across multiple domains, coupled

with the paucity of research into the nature and consequences of the evaluative content of feedback, supports the exploration of the hypothesis as outlined in Chapter 1.

2.5 Summary of Related Literature

The preceding review of related literature is meant to provide context and support for the hypothesis presented in Chapter 1. The distinction between formative and summative forms of performance appraisal is provided to situate formative appraisal processes within the broader context of personnel appraisal. The detailed exploration of Cousins' (1995) model of growth-oriented appraisal processes provides context for the appraisal paradigm used in the current study. The review of the literature on public speaking provides perspective on the state of research in this field, as well as some motivation for developing a measure of performance self-efficacy.

The summary of research in fields related to personnel evaluation is provided in order to show that the trend towards self-evaluation is not an isolated occurrence, but rather a trend that seems to span several related fields within the domain of evaluation. Finally, the section on self-reflection is provided in order to provide background for the discussion section of the document.

Chapter 3 – Pre-Study

In the next two chapters, the methods used in the study are described, followed by the results of the study, and a brief summary of the study.

3.0 Overview of Research Design

The overall design of this multi-phase research project is presented in Table 3.1. The main study is preceded by a preliminary study designed to develop and validate a measure of performance efficacy. This pre-study included a pilot test phase in which the face validity of the measure of performance self-efficacy is assessed.

Table 3.1

Overview of the Research Project

Phase of Study	Purpose
Pilot Study (Chapter 3)	To establish the face validity of the Speaker Self-efficacy Rating Scale (SSRS) for use in the Pre-study and Main Study
Pre-Study (Chapter 3)	To assess the reliability of the SSRS and to establish normative distributions
Main Study (Chapter 4)	To test the impact of two forms of constructive feedback on SSRS scores in an experimental setting

3.1 Pilot Study Phase of the Pre-study

The purpose of the pilot study was to assess the face validity of the Speaker Self-efficacy Rating Scale (SSRS). The SSRS was developed as a measure of performance self-efficacy in the domain of public speaking performance. The SSRS was to be used as the main data collection instrument in the following studies.

3.1.1 Method

3.1.1.1 Participants

Five Ottawa-area members of Toastmasters International (TI) chapter executives and five members of the national executive of the Canadian Association of Professional Speakers (CAPS) were recruited by direct telephone contact to volunteer their services for this phase of the study. There were 9 men and 1 woman in the sample; the participants ranged in age from late twenties to early fifties. The participants were all known to the researcher, and were recruited specifically because of their extensive experience in public speaking.

3.1.1.2 Instrument

A copy of a customized consent form for participation in a University of Ottawa research project is presented in Appendix A1. The data collection instrument used was a custom-designed instrument, the Speaker Self-Efficacy Rating Scale (SSRS). The SSRS integrated items from other instruments as a purported measure of performance self-efficacy. The items on the SSRS were adapted from the works of Sherer et al. (1982), Wilmington et al. (1993) and Ellis (1995). The SSRS consists of three subscales and a demographic data section. A copy of the questions on the SSRS subscales is presented in Table 3.2. The questions for the demographic data section are presented in Appendix A2.

The three subscales of the SSRS are: general self-efficacy (GSE), social self-efficacy (SSE) and performance self-efficacy (PSE). The items on the three subscales consist of a statement and a 5-point Likert-scale response section. Respondents were to indicate on the Likert-scale the extent to which they believe the statement reflects their behavior. The options on the Likert scale were: *Never*, *Rarely*, *Sometimes*, *Frequently* and *Always*. There was also an option for the respondent to indicate that the statement was "Not Applicable." The GSE and SSE

Table 3.2

Speaker Self-efficacy Rating Scale (SSRS) Questions

Subscale	#	Item
General Self-Efficacy	**1.**	When I make plans I am certain I can make them work.
	2.	One of my problems is that I cannot get down to work when I should. **(R)**
	3.	If I can't do a job right the first time I keep trying until I can.
	4.	When I set important goals for myself I rarely achieve them. **(R)**
	5.	I give up on things before completing them. **(R)**
	6.	I avoid facing difficulties. **(R)**
	7.	If something looks too complicated I will not even bother to try it.
	8.	When I have something unpleasant to do I stick to it until I finish it.
	9.	When I decide to do something I go right to work on it.
	10.	When trying to learn something new I soon give up if I am not initially successful. **(R)**
	11.	When unexpected problems occur I don't handle them well. **(R)**
	12.	I avoid trying to learn new things when they look too difficult for me. **(R)**
	13.	Failure just makes me try harder.
	14.	I feel insecure about my ability to do things. **(R)**
	15.	I am a self-reliant person.
	16.	I give up easily. **(R)**
	17.	I do not seem capable of dealing with most problems that come up in life. **(R)**
Social Self-efficacy	**1.**	When I'm trying to become friends with someone who seems uninterested at first, I don't give up easily.
	2.	If I meet someone interesting who is hard to make friends with I'll soon stop trying to make friends with that person. **(R)**
	3.	I do not handle myself well in social gatherings. **(R)**
	4.	If I see someone I would like to meet I go to that person instead of waiting for him or her to come to me.
	5.	I have acquired my friends through my personal abilities at making friends.
	6.	It is difficult for me to make new friends. **(R)**
Performance Self-efficacy	**1.**	I have excellent posture when giving a presentation.
	2.	I have difficulty using appropriate gestures when giving a presentation. **(R)**
	3.	Generally, I move smoothly from idea to idea within my presentation.
	4.	I use appropriate facial expressions.
	5.	I use variety in pitch effectively to enhance my message
	6.	Maintaining eye contact is a problem with me. **(R)**
	7.	Some audience members have difficulty hearing me. **(R)**
	8.	I use variety in my rate of speech effectively.
	9.	I have trouble articulating words clearly. **(R)**
	10.	I make very few, if any, pronunciation errors.
	11.	I find it difficult to avoid using filler words and pauses. **(R)**
	12.	I maintain a suitable vocal quality during my presentations.

(R) - item to be reverse-scored

subscales contained 17 and 6 items respectively, while the PSE contained 12 items. Scores on each item within a scale were added to yield a scale score. Several items were worded in the negative and were therefore reverse-scored (these items are marked with an '**(R)**' after the item in Table 3.2). The values assigned to the items were: Never = 1, Rarely = 2, Sometimes = 3, Frequently = 4 and Always = 5.

The maximum possible score on the GSE subscale was 85, on the SSE subscale 30, and on the PSE subscale 60. The subscale score was divided by the number of items on the subscale, yielding a mean score (out of 5) that was readily comparable with other subscale means. The items of the GSE and SSE subscales were randomly ordered and presented as a single 23-item scale labeled General Views about Self, while the items for the PSE subscale were randomly ordered and presented as a separate scale labeled Views about Performance as a Speaker. Items for the GSE and SSE were adapted from the work of Sherer et al. (1982). The PSE subscale was adapted from items arising from the work of Ellis (1995) and Willmington et al. (1993).

There was also a demographic data section (labeled Background Information) that consisted of items intended to assist in understanding the makeup of the sample without being unnecessarily intrusive. The demographic data section items are presented in Appendix A2.

3.1.1.3 Procedure

The researcher contacted each participant by telephone, and the intent of the study was explained. All participants agreed to review the SSRS and to assess the face validity of the instrument. Specifically, they were asked to assess whether the questions were clear and easily understood, and whether the PSE subscale logically and reasonably covered the range of performance behaviors expected of public speakers. Depending on the participant's location,

either a printed copy or a fax of the consent form and the SSRS instrument were delivered to the participants, and a deadline of two weeks for submitting responses was established.

3.1.2 Results of the Pilot Study

Of the 10 people contacted, seven replied within the two-week timeframe established (4 members of TI and 3 members of CAPS). Of these seven, only three had suggestions for improving the instrument; the other four indicated that they thought the instrument met the requirements of the study. The suggestions for improving the instrument were incorporated into the revised SSRS where appropriate, which included adding four new questions and rewording several other questions on the PSE subscale. The revised PSE subscale questions are presented in Table 3.3. The only suggestion for the demographic data section was that the maximum number of presentations per year should be increased from 50 to 100. This was accomplished by adding two response options for the number of presentations per year: "50-100" and "100+". The only suggestion not acted upon was to reword all of the questions so that they would be positive in nature. This suggestion, although advanced by several respondents, would not be in keeping with established test-development theory (Crocker & Algina, 1986).

3.1.3 Summary of the Pilot Study

The results of the pilot study provided support for the use of the SSRS (in its revised form) in the pre-study phase of the research project. Suggestions for changes were incorporated where appropriate, and the revised SSRS provides a potentially useful tool for assessing performance self-efficacy in the public speaking forum.

Table 3.3

Revised PSE Subscale Questions

#	Item
1.	I have correct posture when giving a presentation.
2.	I have difficulty using open and varied gestures when giving a presentation. **(R)**
3.	I move smoothly from idea to idea within my presentation.
4.	I use dynamic, varied and natural facial expressions.
5.	I vary my vocal pitch to suit the material and maintain audience interest.
6.	I have difficulty maintaining eye contact. **(R)**
7.	Some audience members have difficulty hearing me. **(R)**
8.	I vary my rate of speech to suit the material and maintain audience interest.
9.	I have difficulty articulating words. **(R)**
10.	I make very few, if any, pronunciation errors.
11.	I find it difficult to avoid using filler words and pauses. **(R)**
12.	I maintain a resonating, dynamic vocal quality, fully supported by breath during my presentations.
13.	I increase volume and slow down on key words to emphasize their meaning in a sentence.
14.	I use the entire stage or podium during my presentation.
15.	I speak from behind a lectern. **(R)**
16.	I have difficulty pausing during my presentations.

(R) – item to be reverse-scored

3.2 Pre-study (survey)

The purpose of the pre-study was to establish the reliability of the SSRS, and to obtain normative data for the SSRS from members of both professional and amateur speaker organizations in Canada. To this end, a large-scale survey methodology was employed, with participants recruited from across Canada.

3.2.1 Method

3.2.1.1 Participants

A sample of 949 members of TI and the population of 150 professional members of CAPS were recruited for participation in the pre-study. Within the TI participants, each province (except Québec) and the Yukon Territory were represented (the territory of Nunavut was not in existence at the time of the study, and the representative of the sole club in the Northwest

Territories did not respond to the invitation to participate). The 949 surveys distributed represent roughly 25% of the national membership of approximately 4000. Of the 949 surveys distributed to TI members, 323 usable responses were received by the deadline, yielding a 34% response rate. Of the 150 surveys mailed to CAPS members, 51 usable responses were received within the 2-month deadline, also yielding a 34% response rate. Therefore the overall response rate was 34%. Within the CAPS participants, members of all seven chapters across the country were represented.

Within the TI participants, of the 323 responses received, only 223 provided evidence of their province of residence.[2] Of the 223 who indicated their province of residence, none indicated that they were from either Newfoundland or the Yukon Territory. Due to the anonymous response nature of the survey, it is not possible to determine where the 100 respondents who did not indicate their province of residence were from. There was no reason to believe that non-respondents differed from respondents, especially given that the response rate for CAPS participants was identical.

3.2.1.2 Instrument

A cover letter from the national president of CAPS was used for the CAPS participants (copy presented in Appendix A3). A generic cover letter was used for the TI participants (copy presented in Appendix A4). A revised version of the SSRS based on the results of the pilot study served as the data collection instrument for the pre-study. The SSRS was presented as a four-page booklet, with the appropriate cover letter (TI or CAPS) on the front and the three subsections of the SSRS presented in order on the ensuing pages. The combined GSE and SSE

[2] The participants province of residence was determined by examining their club number (obtained in the Background Information section of the SSRS) and matching it with the master list of TI clubs obtained from the TI website. This information was used only to determine what province the participant resided in. There was no attempt, nor intent, to identify individual participants.

subscales are presented in Appendix A5, while the PSE subscale and the demographic data section are presented in Appendices A6 and A7 respectively.

3.2.1.3 Procedure

The procedure differed for the two groups (CAPS and TI) as the groups themselves differ greatly in their organizational structure. For CAPS participants, the president of the national association provided a covering letter of support for the project, and a nominal fee was charged to cover expenses to include the SSRS in their monthly mailing to the 150 professional members of the association. A professional member of CAPS is considered to be a "full-time" public speaker, who must have "20 fee paid engagements during the past 12 months OR $25,000 in speaking income in the same period" (CAPS, 1999). Each professional member of CAPS received the SSRS printed on blue paper, to distinguish it when returned from the TI version which was printed on white paper, and the cover letter explaining the purpose of the project (sample provided in Appendix A3) in the regular monthly mailing of the association, as well as a stamped, addressed reply envelope. A deadline of two months from the initial mailing date was established for completed surveys to be returned. A reminder notice was included in the regular monthly mailing of the association one month after the initial mailing. The reason for the long delay between mailings and the 2-month response deadline was the schedule of mailings by the association; it only mails to its members once a month.

The structure of the TI organization precluded the use of a similar procedure for recruiting participants. The researcher obtained telephone contact numbers for the TI clubs across Canada using the TI website (**www.toastmasters.org**). The researcher contacted 64 randomly selected local club presidents from all areas of Canada except Québec and the Ottawa area, and 26 responded positively that they and their members would be interested in

participating in the research project. Of the 38 presidents contacted who did not respond positively, most (35) did not respond at all. The three who did respond indicated that they would not be able to proceed without approval from their area and / or district governors, and the timelines they provided for receiving this approval (upwards of 2 months) prevented their inclusion. Québec TI clubs were not included because a French version of the SSRS was not available. Ottawa-area TI clubs were excluded because these were the population from which the sample for the main study was to be drawn. Several of the local club presidents across the country were also district officers, and volunteered to forward additional copies of the SSRS to other interested clubs in their area. A deadline of 2 months after the initial mailing to the local club presidents was established for the return of completed surveys. This deadline was established to correspond with the timeline for CAPS participants. A follow-up letter was sent to the TI contacts one month after the initial mailing, to remind them of the response deadline and to ask them to convey this reminder to the members who had received the surveys.

It should be noted that, while perhaps not the optimum method of recruiting participants, given the structure of the TI organization, it seemed to be effective without being overly intrusive. Each participant received a package consisting of the SSRS printed on white paper, a generic cover letter of introduction from the researcher outlining the research project (sample provided in Appendix A4), and a stamped, addressed reply envelope. The returned survey information was entered into SPSS, and the indicated analyses were performed.

3.2.2 Results of the Pre-study (survey)

The summary statistics for the GSE subscale, the SSE subscale, the PSE subscale and the demographic data section are presented in Appendices A8, A9, A10 and A11 respectively. The results of the GSE, SSE and PSE subscales are presented as mean score, standard deviation and

number of respondents for each item for each group (TI and CAPS). Items on the subscales that have been reverse-scored are marked, and the reversed score is presented. Higher scores indicate greater agreement that the item reflects the respondents' behavior (e.g., 1 = never, 5 = always).

The demographic data section summary statistics are presented as the number and percentage of respondents who chose that response option, for each group (TI and CAPS).[3] Where possible (e.g., Question 5: I have been presenting for audiences for ## years), the mean score, standard deviation and number of respondents for each item for each group are presented. The responses to one question (Question 7a, which asked the respondent which TI club they attended most frequently) are not reported. The reason these data are not reported is that they serve only to identify the TI club to which the respondent belongs.

Internal consistency estimates were calculated for each of the subscales of the SSRS, using all of the participant data (combining both TI and CAPS responses). The results of this analysis, broken down by subscale of the SSRS, are presented in Table 3.4. As can be seen in Table 3.4, both the GSE and PSE subscales produced test statistics greater than 0.8, while the SSE (with substantially fewer items than the other two subscales) produced a Cronbach's α greater than 0.7. These reliability levels were judged to be acceptable for the purposes of the present study (Crocker & Algina, 1986).

The inter-scale correlations for the SSRS subscales are presented in Table 3.5. The correlations between the subscales are fairly low (none accounting for more than 16% of the shared variance), and thus do not raise concerns of dependence between the underlying constructs. The significance of the correlations seems to be mainly due to the size of the sample

[3] It should be noted that some respondents to the survey were members of both CAPS and TI. Most (but not all) of the CAPS members responded as CAPS members, but at least two apparently filled out the TI form instead. These two were counted as TI members for the purposes of analysis.

(roughly 300 participants), rather than being indicative of strong and potentially worrisome relationships between the constructs.

Table 3.4

Internal Consistency Estimates (Cronbach's ∝) and Sample Sizes for the SSRS Subscales

SSRS Subscale	Cronbach's ∝	N
GSE (17 items)	0.8656	349
SSE (6 items)	0.7158	352
PSE (16 items)	0.8116	347

The scale scores for the combined sample and each of the groups (TI and CAPS) are presented in Table 3.6. T-tests conducted on the subscale scores for each group of respondents (CAPS and TI) indicated that members of the CAPS group scored significantly higher than members of the TI group on each of the subscales, even after the difference in sample sizes was accounted for. The results of these analyses are presented in Table 3.7.

Table 3.5

Inter-scale Correlations and Sample Sizes for SSRS Subscales

Subscale	GSE	SSE	PSE
GSE	1.000 n = 349		
SSE	0.441* n = 352	1.000 n = 347	
PSE	0.352* n = 347	0.401* n = 299	1.000 n = 352

* - Correlation is significant at the 0.01 level (2-tailed)

Table 3.6

Mean Scale Scores (SD in Brackets) and Sample Sizes for SSRS Subscales

Subscale	Combined	TI	CAPS
GSE (17 items)	3.85 (0.42) n =349	3.83 (0.43) n = 300	3.97 (0.38) n = 49
SSE (6 items)	3.46 (0.58) n = 352	3.42 (0.58) n = 302	3.71 (0.50) n = 50
PSE (16 items)	3.73 (0.49) n = 347	3.65 (0.47) n = 299	4.20 (0.35) n = 48

Note - All scores have been adjusted to a 5-point scale for comparative purposes. The number in brackets after each subscale name indicates the number of questions upon which the score is based.

Table 3.7

T-tests on SSRS Subscale Scores by Group (TI or CAPS)

Subscale	df	t	Significance
GSE	347	-3.35	.001
SSE	350	-2.11	.036
PSE	345	-7.75	.000

*** The results displayed assume equal variances for the underlying distributions. The results for the unequal variance analysis were virtually identical.**

3.2.3 Summary of the Pre-study

The response rate of 34% overall, and 34% for each of the subgroups, was quite good considering the "cold-call" nature of the survey, and the distribution problems encountered. As might be expected, the professional speakers (CAPS members) scored higher than the amateur

speakers (TI members) on the PSE subscale. Although this indicates that the CAPS members reported being more confident in their ability to perform (in terms of presentation skills), this does not reflect that they are necessarily better performers than the TI members.

As no research could be found that compared these two groups in terms of general or social self-efficacy, it is difficult to determine the meaning of the significant difference in scores on the GSE and SSE, with CAPS members again scoring significantly higher than TI members on both subscales. A plausible albeit speculative interpretation might be that professional speakers are more efficacious in general than non-professional speakers, but again, there is little evidence here to support this stance.

Internal consistency estimates (typically Cronbach's α) are a measure of how consistently the items of a scale measure what they purport to measure. The internal consistency estimates presented in Table 3.4 are sufficiently high (greater than 0.7) to indicate that an acceptable level of consistency has been observed.

Given the internal consistency statistics and the large sample size, combined with the pilot study results, the SSRS appears to be an acceptably reliable and valid tool for use as the main data-collection instrument in the main study. These findings also support the use of the instrument for other research and practical purposes.

Chapter 4 – Main Study

4.0 Purpose of the Main Study

The purpose of the main study was to explore in an experimental, comparative group setting the impact of descriptive versus evaluative forms of constructive feedback on the participants' confidence in their ability to perform as public speakers, or their performance self-efficacy. It is hypothesized that greater effects will be attributed to descriptive as opposed to evaluative feedback since descriptive feedback would be more likely to stimulate self-reflection and critique.

The SSRS validated in the pre-study phase of this research program served as the dependent measure for the main study. In addition, qualitative data were collected to enhance understanding of the constructs under study and to monitor the extent to which treatments differed.

4.1 Method

4.1.1 Participants

The participants in the main study were 35 Ottawa area TI members. All participants were recruited by advertisement at their regular meetings through in-person marketing by either the researcher or local club presidents briefed by the researcher. The advertisement brochure distributed at the meetings consisted of an 8½ by 11 inch page, folded in half, producing a front cover, two middle pages and a back cover (samples presented in Appendices B1, B2, B3 and B4 respectively). In addition, a website containing the same information as the brochure was used as an advertisement tool for TI clubs with e-mail lists. The website presented the recruitment information in the same format as the hard-copy brochure.

All participants were volunteers, but they were required to attend all six sessions of the program. As an incentive to participate, they were promised a videotape that included all six of their own presentations made during the study, a "best-of" compilation of other participants' performances, and the opportunity to hear a well-known professional speaker talk on the use of humor in presentations. They were required to attend all sessions and respond to the post-program survey in order to be eligible to receive the video.

Of the 35 participants who originally signed up for the program, 33 showed up on the first day for the training sessions. One of the two who did not show up later informed the researcher that he / she had another engagement on the same day, and the other was a parent of this person and declined to attend alone. Of the 33 who showed up for the training sessions, 32 completed the entire program. One person failed to show up for any session after the initial training session, and failed to respond in any way to numerous attempts at contact. It was later explained by people that knew this person that he or she was well known for starting programs and dropping out soon after the onset. There is no way to determine the real reason for this person's withdrawal, as he / she does not respond to messages.

4.1.2 Instruments

A copy of a customized consent form for participation in a University of Ottawa research project is presented in Appendix B5. The main data collection instrument used was the SSRS developed and validated in the pilot study and pre-study phases. The questions on the SSRS were presented earlier in Appendices A5, A6 and A7.

Other instruments were used as well. The Specific Area of Interest Scale (SAIS) consisted of a list of areas of presentation skills performance, and a 10-point Likert-scale anchored with 'poor' at 1 and 'excellent' at 10. Participants were to select a specific area of

interest and indicate their level of ability in that area. The items on the SAIS were culled from various sources that purport to assess public speaking capability, such as the work of Morreale (1990), Morreale et al. (1991) and Morreale and Hackman (1994). This measure served as a pre- and posttest of their self-rated ability level. A sample of the SAIS is presented in Appendix B6.

The Feedback Expectation Sheet (FES) was developed from the literature review on feedback and was used during the program as a reminder for the participants as to how to conduct the feedback sessions. A sample of the FES is presented in Appendix B7.

The Elements of Effective Constructive Feedback Sheet (EECFS) was developed by the researcher from the literature review on feedback and was also used during the program as a reminder for the participants as to the characteristics of constructive feedback. A sample of the EECFS is presented in Appendix B8. Sample probing questions for use in the feedback sessions were also included on the EECFS.

The Feedback Session Evaluation Sheet (FSES) was used to monitor the supervisor and appraisee perceptions of the feedback sessions, as well as a backup data-collection device as to what was discussed in the feedback session. Each appraiser completed the FSES after each session. A sample of the FSES is presented in Appendix B9.

Cardboard cards approximately 18" by 36" in size and colored green, yellow, and red were used as timing cues for the presentations. A similar timing method is used in regular TI meetings. A videotape of each participant's performance was used as a stimulated recall device in the feedback sessions. Audiotapes of the feedback sessions were used to collect qualitative data on the participants' experience in the program.

At the conclusion of the program, a post-program survey of the participants was conducted, in order to augment the data collected in the program. The open-ended questions for

the survey are presented in Appendix B10. The first six questions of the survey were distributed by email two days after the conclusion of the program. Most participants (20) responded within one week. All participants (except the one who became ill) responded within three weeks. Approximately one month after the conclusion of the program, when the results of the main study were known and the issue of contamination arose, the seventh question was developed and distributed by email. The participants responded to this email within two weeks.[4]

4.1.3 Procedure

4.1.3.1 Assignment to treatment conditions

Participants were randomly assigned to the two treatment groups, evaluative feedback (EF) and descriptive feedback (DF) as they signed up for the program. In order to ensure equal numbers in each group, the first person to sign up was assigned to one or the other group by random draw; the next person to sign up was assigned to the other group in order to equalize numbers. The next person to sign up was assigned to one group or the other by random draw, again. This process was repeated with each successive registrant. The participants attended an initial training session for their respective group. They then completed five additional sessions, making a total of six sessions. The first and last sessions were offered on Saturdays. The four intervening sessions were offered on Monday through Thursday evenings every week for four consecutive weeks. Participants had indicated their availability for the weeknight sessions, and every effort was made to accommodate their schedules. An overview of the program is presented in Table 4.1.

---

[4] An adapted version of the performance subscale of the SSRS, the Peer Performance Rating scale (PPRS), was also developed for use as a pre- and posttest measure of performance as assessed by peers. A sample of the PPRS is provided in Appendix B11. The PPRS was developed for use as a triangulation tool to supplement the SSRS. The participants were also to assess each performance they observed throughout the study. Unfortunately, due to time constraints and concerns about the reliability of the instrument, the PPRS was not used in the study.

Table 4.1

Overview of Main Study

Session #	Purpose	Evaluative FB Group	Descriptive FB Group
1	Pretest & Training	Saturday AM	Saturday PM
2 – 5	Data Collection	1x per week, Monday through Thursday evenings (joint sessions, fit to participant schedules)	
6	Posttest & Debriefing	Joint Session – Saturday Afternoon	

4.1.3.2 Initial training sessions

Participants in the two training sessions underwent the exact same training (one Saturday a.m., one Saturday p.m.), except for the specific feedback delivered in the sessions. The differences between the EF and DF training are detailed later. A general outline of the training sessions follows. Note that a random draw held the night before the training sessions determined which session (morning or afternoon) received which form of feedback training (evaluative or descriptive).

Upon arrival at the training facility at the University of Ottawa, the participants were greeted by the researcher and assistants and were directed to the training room. After a brief explanation of the session agenda and program objectives, the participants were presented with the SSRS and asked to complete the instrument. After the participants' completed the SSRS, the SSRS was collected and the participants were presented with the SAIS. They were asked to select the area of performance in which they were most interested in improving (with a space provided to identify areas not listed on the SAIS) and to rate their current level of ability. After

they completed the SAIS and it was collected, the participants were randomly divided into two groups to present their 5-7 minute speech for the first time.[5]

Two rooms were used for the presentations, in order to expedite this phase of the study. The presentation order was randomly determined. Each participant presented his/her speech to an audience consisting of the other members of their group, as well as the researcher or assistant who videotaped the performance. Visual cues (i.e., the green, yellow and red cardboard cards) were used to indicate when 5, 6 and 7 minutes, respectively, had elapsed. These visual cues, which are very familiar to the TI participants from their weekly meetings, were used to help the participants keep within the proscribed time frame of five to seven minute for their presentations. After all the presentations were completed, the participants reconvened in the large training room.

The researcher outlined the general format of the four weekly sessions, which was exactly the same for both groups: participant arrival, presentations, feedback sessions, and departure. The only difference between the EF and DF treatment conditions was in the actual process and content of the feedback sessions themselves. This was not disclosed to the participants. The researcher distributed to the participants a sheet outlining the Elements of Effective Constructive Feedback (sample presented in Appendix B8) and explained the meaning of each component. On the same sheet were examples of probing questions that could be asked during the feedback session. Participants were encouraged to ask questions if they did not fully understand the concept of constructive feedback as explained or the probing questions.

---

[5] At this time, because of a concern over the amount of time required to complete the PPRS, and the researcher's uncertainty as to the whether the participants could properly utilize the PPRS, the decision was made to omit the instrument from the study. The possibility of introducing it later was entertained, but the opportunity (i.e., available time) never presented itself.

The researcher and assistant next presented a sample feedback session appropriate to the feedback condition of the participants, detailing the exact procedure to be followed. After the sample session, participants were given the opportunity to ask questions and to practice giving each other appropriate feedback, using the video of their own performance from earlier in the session and a partner they themselves selected. The researcher provided guidance to the participants on their feedback as they practiced. Once the participants were satisfied that they understood how to give feedback in accordance with the instructions for their condition, they were free to leave. Participants in each of the two training sessions were aware that another training session with different participants had occurred / would occur. They did not know that participants in the other session received different instructions for the delivery of the feedback in the feedback sessions.

4.1.3.3 Generic schedule for weekly sessions

It is important to note that participants from the group who received evaluative feedback (EF) training participated in the weekly sessions along with participants who received descriptive feedback (DF) training, although none of the participants were informed that they might be giving and receiving a different form of feedback than other pairings. Participants were always paired with individuals from the same feedback group for the feedback pairings. The generic schedule for the weekly sessions is presented next; details particular to each treatment condition (evaluative or descriptive feedback) follow.

The participants for each evening were asked to arrive, ready to present, at 19:30. Once all the participants who were scheduled for a given evening had arrived (with a 5 minute grace period, if necessary), each participant gave a presentation before the audience of peers. The same green, yellow and red cards as used in the first meeting were used as time-cues throughout the

four weeks of feedback sessions. The presentation order was pre-determined by the researcher, and varied for each weekly session so that no participant was always first or last to present. If there were more than six participants at a given session, the participants were split into two groups and the presentations took place in two separate rooms (with the researcher or an assistant running the video camera in each room). This was done in order to reduce the length of time for the presentations so that the participants were not required to commit more than 90 minutes per weekly session. The feedback pairings were pre-determined by the researcher, to ensure that a participant in the EF condition was never paired with a participant in the DF condition, and vice-versa. If there were an odd number of people on an evening (due to cancellations), the researcher or his assistant assumed the role of the missing person for the feedback session. As much as possible, participants were partnered with different individuals each week in order to minimize the impact of strong or weak evaluators and possible personality conflicts on the results.

Each presentation was videotaped by the researchers. Once all presentations were complete, the participants were given the identity of their partners for the feedback sessions, and they proceeded to the appropriate room with their partner and the videotape cassette containing both of their performances. The feedback room contained a VCR and monitor, an audiotape recorder with a labeled cassette, and two copies of each of the FES, the EECFS, and the FSES (samples provided in Appendices B7, B8 and B9 respectively). Participants were asked to audio-record their feedback sessions. The partners were free to conduct the feedback sessions in whichever order they preferred; they tended to opt for the order in which they presented.

While the participants were conducting the feedback sessions, the researcher and assistant monitored the form of the feedback being given by wandering between the feedback rooms and listening to the discussion. The researcher and assistant also ensured that the participants had all

required materials and dealt with any equipment problems that arose. If any questions were raised as to the form of feedback to be delivered, the researcher was called to assist the participants.

4.1.3.4 Feedback conditions

In this section the differences between the two treatment conditions are described. Other than the specific differences noted below, the participants received highly similar conditions in order to maximize the effects of the treatment.

4.1.3.4.1 Evaluative feedback condition

Participants in the morning training session received training in the delivery of evaluative feedback (EF) for use in the four weekly feedback sessions. An EF session was defined as consisting of the evaluator (the person *not* performing on the video being reviewed) and the performer (or appraisee). The role of the evaluator was to control the replay of the video and to provide verbal constructive feedback to the performer on the performance they both reviewed. The evaluator was instructed to provide as much verbal feedback as possible and to cover as many aspects of the performance as possible. The length of the feedback session was to be determined by the evaluator.

The role of the performer was to listen to the feedback, to request clarification if necessary, but not to provide evaluative commentary on the performance being reviewed. When the evaluator had completed her / his comments on the performance, the evaluator and the performer would complete the Feedback Session Evaluation sheet (sample provided in Appendix B9).

On the FSES the evaluator was required to list three things that the performer did well, and at least two things that the performer should strive to improve. The performer was to list the

same things on her / his sheet as the evaluator did. The evaluator and performer then switched roles, with the performer from the first session assuming the role of the evaluator, and vice-versa. After the second performance review was completed, the participants would both complete the FSES for the second performance review. The participants were then free to leave after informing the researcher that they were finished.

4.1.3.4.2 Descriptive feedback condition

Participants in the afternoon training session received training in the delivery of descriptive feedback (DF) for use in the four weekly feedback sessions. A DF session was defined as consisting of the performer (the person performing on the video being reviewed) and the partner. The role of the performer was to control the replay of the video, and to provide verbal constructive feedback to her / himself on the performance being reviewed. The performer was instructed to provide as much verbal feedback as possible, and to cover as many aspects of the performance as possible. The length of the feedback session was to be determined by the performer.

The role of the partner was to try to get the performer to verbalize their self-reflective comments as much as possible, and to ask probing questions of the performer when appropriate. The partner was instructed *not* to provide evaluative comments to the performer. Examples of acceptable probing questions for use in the feedback sessions were provided on the EECF sheet (sample provided in Appendix B8). When the performer had completed her/his comments on the performance, the performer and the partner would complete the Feedback Session Evaluation sheet (sample provided in Appendix B9).

On the FSES the performer was required to list at least three things that he or she did well in the performance, and at least two areas that he or she should strive to improve. The partner

was to list the same things as the performer indicated. The performer and partner then switched roles, with the performer from the first session assuming the role of the partner, and vice-versa. After the second performance review was completed, the participants would both complete the FSES for the second performance review. The participants were then free to leave after informing the researcher that they were finished.

4.1.3.5 Final session

The final data collection session was scheduled for the Saturday after the four weekly feedback sessions were completed. The format for the final session was similar to the first Saturday session, except that all 32 participants met at the same time. Upon arrival at the training facility, the participants were greeted by the researcher and assistants and directed to the main room. After a brief explanation of the agenda, the participants were presented with the SSRS and asked to complete the instrument (posttest). After the participants completed the SSRS, it was collected and they were presented with the SAIS. The area of performance they had previously indicated was highlighted, and their previous rating was circled. The participants were asked to rate their current level of ability. After they completed the SAIS and it was collected, the participants were divided into three groups for presenting their 5-7 minute speech for the last time.

Three rooms were used in order to reduce the amount of time required for the presentations. The presentation order was randomly determined. Each participant presented a speech to an audience consisting of the other members of their group, as well as the researcher or assistant who videotaped the performance. The time cards (green, yellow and red) were again used to help performers stay within the time limit.

Each participant was also asked to list the three presentations that they would most like to have included on their compilation videotape. These presentations could be presentations they viewed that day or any previous presentation they had seen. These lists were handed to the researcher after the presentations were completed.

After all the presentations were completed, the participants reconvened in the main room. The finale of the program was a 90 minute presentation by a professional speaker and member of CAPS entitled "How to Use Humor in Your Presentations." After this presentation, the participants were given certificates commemorating their completion of the presentation skills program and the research study. At this point, the participants were free to leave, with the understanding they would receive the customized video compilation of their performances after completion of the post-program survey.

The post-program survey questions were developed based upon the expected results of the program, in order to enhance understanding of the results. The post-program survey questions were distributed through e-mail to the participants (all participants had internet access). This method was confirmed as acceptable with the participants before the end of the program. Based upon the actual results of the study a seventh question was added. The participants understood that they would not receive the videotape of their performances until they had responded to the interview questions.

After the program was finished, the researcher determined those presentations most frequently selected by participants for inclusion on their compilation videotape. The top 10 selected presentations were included on the compilation videotape. After the participants responded to the seven post-program survey questions, the researcher delivered the videotape to

the participant. Each customized video included all six of the participant's own performances, the 10 selected presentations, and the 90 minute presentation from the final session.[6]

4.2 Results

The mean pre- and posttest scores for each treatment condition for each of the subscales of the SSRS are presented in Table 4.2. The mean scores reveal that, for the combined groups, there was an increase from pretest to posttest on each of the SSRS subscales and the SAIS. This indicates that there was an overall increase in the participants' self-efficacy and self-rated performance as measured by these self-report instruments. An exception to this pattern was a mean decrease from pre- to posttest on GSE scores for the EF group. When compared to the norms for the instrument obtained in the pre-study, the main study participants' mean pretest scores did not appear to be different than the survey participants' scores on any of the three subscales (GSE, SSE, or PSE).

The inter-correlations for the pre- and posttest scores on each of the subscales are presented in Table 4.3. The significant and fairly large correlations (ranging from 0.68 to 0.91) between the pre- and posttests for each of the SSRS subscales and the SAIS rating indicate that the pretests will serve as good covariates for the posttest variables in the main analysis. Significant variation in posttest scores is explained by pretest performance. The other correlations are not large enough to indicate any serious multi-collinearity problems for the main analysis.

The main analysis was a test of the effects of the feedback condition and consisted of a multivariate analysis of covariance (MANCOVA) of the posttest scores on the three SSRS subscales (GSE, SSE, and PSE) and the SAIS ratings, using the pretest scores on the subscales

[6] Permission was obtained from the professional speaker to include his presentation on the compilation video. All participants voluntarily signed a waiver allowing their performances to be included on the compilation video if they were selected.

and the SAIS as covariates. The purpose of this analysis was to test the main hypothesis for the thesis, that is, to determine if participants in the DF condition outperformed those in the EF group, after any initial difference was removed. The results of this analysis revealed no significant difference between the participants on the three SSRS subscales or on the SAIS ratings after pretest variability was accounted for using the covariates. The results of this analysis are presented in Table 4.4.

Table 4.2

Mean Pre- and Posttest Scale Scores (SD in brackets) and Sample Sizes for SSRS Subscales and SAIS for Main Study Participants

	Combined Groups		EF Group		DF Group	
Subscale	**Pretest**	**Posttest**	**Pretest**	**Posttest**	**Pretest**	**Posttest**
GSE (17 items)	3.92 (0.48) n = 32	3.95 (0.54) n = 32	3.81 (0.57) n = 15	3.78 (0.61) n = 15	4.01 (0.38) n = 17	4.10 (0.44) n = 17
SSE (6 items)	3.40 (0.66) n = 32	3.46 (0.59) n = 32	3.18 (0.87) n = 15	3.20 (0.70) n = 15	3.60 (0.32) n = 17	3.69 (0.36) n = 17
PSE (16 items)	3.44 (0.46) n = 32	3.54 (0.44) n = 32	3.23 (0.32) n = 15	3.34 (0.40) n = 15	3.61 (0.50) n = 17	3.72 (0.40) n = 17
SAIS (1 item)	2.23 (0.78) n = 32	3.36 (0.83) n = 32	2.20 (0.75) n = 15	3.20 (0.82) n = 15	2.26 (0.83) n = 17	3.50 (0.83) n = 17

Scores: 1 = never, 3 = sometimes, 5 = always

In keeping with Huberty and Morris's (1989) recommendations vis-à-vis multivariate analysis and multiple univariate analyses, a second analysis was performed to ensure that individual differences on the subscales were not masked by the relatively small sample size and the complexity of the overall analysis. The results of these secondary analyses (individual

ANCOVA's using the posttest subscale scores and the SAIS rating as the dependent variables and the corresponding pretest scores as covariates) were consistent with the overall results: No significant differences existed between the participants on the posttest scores once initial differences were accounted for by the pretest scores. The results of this analysis are presented in Table 4.5.

Table 4.3

Inter-scale Correlations for Pre- and Posttest SSRS and SAIS (n = 32)

		Pretest				Posttest			
		GSE	**SSE**	**PSE**	**SAIS**	**GSE**	**SSE**	**PSE**	**SAIS**
	GSE								
Pretest	**SSE**	.50**							
	PSE	.49**	.42*						
	SAIS	.30	.39*	.33					
	GSE	.91**	.45**	.52**	.38*				
Posttest	**SSE**	.54**	.85**	.51**	.35	.55**			
	PSE	.46*	.48**	.86**	.38*	.49**	.54**		
	SAIS	.40*	.46**	.54**	.68**	.47**	.52**	.62**	

*** - Correlation is significant at the 0.05 level (2-tailed)**
**** - Correlation is significant at the 0.01 level (2-tailed)**

Additional analyses were performed in order to assess the sensitivity of the SSRS scores and the SAIS rating to detect differences in self-efficacy from pretest to posttest. If the instruments were accurately assessing the self-efficacy of the participants and the program had some impact on their performance self-efficacy, we would expect that the scores on the GSE and SSE subscales would remain constant from pretest to posttest, while the PSE scores and SAIS

ratings should increase from pretest to posttest interval. The rationale for this expectation is that the intent of the intervention was to improve the performance self-efficacy of the participants and their actual presentation skills, not their general or social self-efficacy. This analysis took the form of one-sample related t-tests on the difference scores (i.e., posttest less pretest) for each of the scale scores and the SAIS ratings. The results of this analysis revealed that, for the combined treatment groups, the difference scores on the PSE and the SAIS were significantly different than zero. The test statistic for the PSE score was $t(31) = 2.46$, $p < .05$ and for the SAIS scores $t(31) = 9.83$, $p < .05$. No significant differences were observed for GSE or SSE scores.

Table 4.4

Results of Multivariate Analysis of Covariance using Hotelling's Trace

Effect	F	df	Error df	Sig
Covariates				
Pretest GSE	20.237	4	23	.000
Pretest SSE	9.594	4	23	.000
Pretest PSE	8.383	4	23	.000
Pretest SAIS	3.546	4	23	.021
Between Groups Factor				
Feedback Condition	0.904	4	23	.478

A further analysis of the difference scores, one-sample related t-tests within treatment conditions, was conducted to determine whether these results would be reflected within the treatment conditions. The results of these analyses revealed that, within both treatment conditions, the participants' scores on the SAIS were significantly different than zero. The test

statistic for the DF group was $t(16) = 6.270$, $p < .05$, and for the EF group $t(14) = 10.25$, $p < .05$. There were no significant differences observed for the other scores.

Table 4.5

Results of Multiple Univariate Analyses of Covariance

Source	df	F	Probability
GENERAL SELF-EFFICACY			
Pretest GSE	1	133.656	.000
Feedback Condition	1	2.251	.144
Error	29		
Total	31		
SOCIAL SELF-EFFICACY			
Pretest SSE	1	65.226	.000
Feedback Condition	1	2.644	.349
Error	29		
Total	31		
PERFORMANCE SELF-EFFICACY			
Pretest PSE	1	61.017	.000
Feedback Condition	1	0.796	.796
Error	29		
Total	31		
SPECIFIC AREA OF INTEREST			
Pretest SAIS	1	25.131	.000
Feedback Condition	1	1.365	.252
Error	29		
Total	31		

4.3 Post-hoc Analyses

Due to the failure of the experimental design to confirm the hypothesis, the qualitative data on the feedback sessions were examined to determine whether the treatment conditions were in fact different throughout the course of the study. It is necessary to establish that treatments differed prior to applying a theoretical interpretation to the observed non-differences. A list of all

the session audiotapes was made, according to session and treatment condition. One session from each feedback condition from each week was randomly selected for transcription, for a total of 8 sessions. The tapes of these feedback sessions were transcribed into a word-processing program and subsequently printed and analyzed.

The purpose of this analysis was to assess the extent to which the participants in the two treatment conditions followed the guidelines for the feedback condition to which they were assigned. Specifically, we are interested in the extent to which participants in the DF condition received descriptive feedback (i.e., non-evaluative), and the extent to which participants in the EF condition received evaluative feedback.

The key criteria for the provision of feedback in the EF treatment condition were that the partner (the person providing feedback to the performer) was to control the video replay and to lead the feedback discussion. Participants in the DF treatment condition were required to control the video replay and provide the feedback commentary on their own performance, with the partner simply coaxing the performer to verbalize their thoughts. The participants were also encouraged to give as much feedback on the performance as possible. The other guidelines for feedback sessions in both treatment conditions were outlined in the EECF sheet made available to the participants at each feedback session (sample provided in Appendix B8): to be as positive in the commentary as possible, to be as specific as possible, to draw examples from the performance on video, to provide a future focus to the commentary, and to provide feedback on as many areas of the performance as possible.

4.3.1 Evaluative Feedback Condition Results

Participants in the EF condition tended to follow the guidelines provided for the feedback sessions fairly well. The partner (the person providing feedback to the performer) controlled the

video replay. As an example, one partner stated: "Maybe we can review a little bit to see what I mean here. Okay, when you say ..." (session 1TU3-1). In this case, the partner rewound the videotape to the specific area of interest to make her point. Another example of the partner controlling the video review comes from session 2M4-1: "So I'm going to go back to the beginning because there were a couple of places near the beginning that show good examples."

Partners in the EF condition also led the feedback discussions, which was the other main criterion for feedback sessions in the EF treatment condition. This result is not surprising, given that this is similar to the traditional feedback model followed in their Toastmasters clubs, although normally without the benefit of the video as a guide. The performers did occasionally ask questions of the partner providing the feedback. These questions tended to be requests for clarification of commentary, but sometimes the performer would ask for feedback on a specific area of their performance. This interaction did not seem to interfere with the evaluative nature of the feedback session, as the partner was still evaluating the performance and providing the evaluative information.

In terms of the quantity of feedback provided, there was not as much feedback given on the performances as expected by the researcher. Participants tended to drift towards commentary on the content of the performance, which is the main focus of their evaluations in TI. While not unexpected, some participants seemed to be more comfortable providing content feedback than performance feedback.

The seven EF treatment condition feedback sessions transcribed (two paired sessions from each of three nights and one single session from the fourth night) were about as long as expected, averaging 18:00 minutes in length. The researcher was expecting the feedback sessions to be roughly 15-20 minutes in length.

As for the positive nature of the feedback provided, the participants were used to providing positive feedback to others during their TI evaluations. In general, the feedback provided was positive. There were no blatantly negative comments made. Feedback tended to be encouraging and respectful. For example, one participant commented, "You look really relaxed, but you're behind the table. And if you went to the side of the table or out in front of the table, maybe you could connect in a different way" (session 2M4). This feedback addressed the issue of movement about the presentation area and reinforced that the performers appeared relaxed and connected with the audience even if they didn't move about. Another participant commented, "I see you as a very good actor, and as very gifted at performance. It's very natural, just like you're telling stories to kids. It comes so natural to you" (session 3M5).

The feedback provided in the EF treatment condition feedback sessions ranged from being quite specific to being categorically vague. This, again, was not unexpected given the background of the participants. In their TI evaluations, participants tend to be more global than specific in their feedback.

As far as drawing examples from the performance on video, there was again a broad range. It is difficult to assess the extent to which the feedback was directly relating to the video, because we only have the audio recording of the feedback sessions. Sometimes the partner would expressly say, "Look, here ..." and explain the action being viewed and what they thought about it. Other times, they would be talking while the video was not on and directly refer to certain actions from the videos. On the other hand, some feedback sessions were less clear; no direct links to the video or the performance were made.

Not very many instances of the partner providing future-oriented feedback were found in the review of the audiotape transcripts. One situation where it did arise was in session 1TU3-2,

where the partner was talking about a particular behavior and that the performer should try to do it differently the next time. "Throw it out. And then if it doesn't work, next week you can put it back in. You've got nothing to lose." This type of commentary was not very frequent.

As for the number of areas covered in the feedback session, there was consistency between sessions. A count of the number of different behaviors referenced in the seven feedback session transcripts was conducted, using the described areas from the Specific Area of Interest Scale (sample provided in Appendix B6). The most specific behaviors commented on was 8, while the least was 3. The average number of areas to which comments applied was six. The behaviors most frequently discussed (in 5 of the 7 sessions transcribed) were movement about the presentation area, gestures and eye contact with the audience. The areas least frequently addressed (in only 1 of the 7 sessions transcribed) were breathing / support, articulation, and pronunciation.

In summary, the participants in the EF treatment condition seemed to follow the guidelines for the feedback sessions fairly well. The partner (the person providing the feedback) controlled the video replay and led the feedback session, although the performers did ask questions for clarification or more information. As for the length of the feedback sessions, they were about as long as the researcher expected. The length was also considerably longer than the fairly short evaluative-type feedback delivered in the traditional TI format.

4.3.2 Descriptive Feedback Condition Results

Participants in the DF condition, unfortunately, were not as effective as the EF participants at following the guidelines for the feedback sessions. The main areas of concern were control of the video playback and the issue of who provided the evaluation of the

performance. The review of the transcripts revealed that some of the participants in the DF treatment condition understood and followed the guidelines, but some participants did not.

For the most part, the performer controlled the video replay, at least until the initial review was complete. At this point, in the ideal case, the performer would provide commentary on his/her performance, and the partner would ask questions geared towards prompting the performer to verbalize his/her thoughts, without providing evaluative commentary. This was the case for most of the sessions; however, there were some sessions where the "silent" partner actually hijacked the session and seized control from the performer.

An example was session 3TU1-2. The partner opens the session by saying "I'd just like to make a comment. I've heard that part three times, okay ..." The partner proceeded to lead the performer through a discussion of the opening segment of her presentation, which in itself was not too evaluative. Unfortunately, the performer was not in control of the session, which was further evidenced when the partner said, "Do you want to just back that up for one second." The partner then made more comments on the presentation.

A second case was session 4TU2-5: "Okay, let's start, and what we'll do is we'll back up a little bit later on, just to go over some stuff. Let's start looking at it not on a frame by frame, but here." Again, the partner took over control of the video playback, and the feedback session itself.

Of the seven sessions transcribed (again, two paired sessions from each of three nights and one single session from the fourth night), five of them seemed to follow the guidelines well. However, after listening carefully to the feedback sessions that were not transcribed, more evaluative feedback appeared to be offered by the partners. This may have been a result of the constantly changing partnering used from week to week, allowing the "non-conformists" to

adversely affect those who were acting properly. It may also have been a result of the performers wanting feedback from the partners, as they are used to receiving in their TI meetings.

The same problems seen in the control of the video replay can be seen in the control of the feedback session. In those cases where the partner took control of the video replay, they also took control of the feedback session.

In terms of the quantity of feedback provided, more feedback was provided than was expected by the researcher. The seven DF treatment condition feedback sessions transcribed averaged 28:00 minutes in length, a full ten minutes longer than the EF sessions. Participants did not seem to drift towards commentary on the content of the performance, as those in the EF condition did. This was interesting because there was concern among the participants at the outset about filling up time with commentary. The researcher was also concerned that there would not be enough commentary from the DF participants. This was not the case.

As for the positive nature of the feedback provided, the participants were used to giving and receiving positive feedback to performers during their TI evaluations. In general, the feedback they provided to themselves was positive. There were only a few exaggerated negative comments made, which were typically made in an attempt to bring some humor to the session and seemed to be treated in that way.

The feedback provided in the DF treatment condition feedback sessions ranged from being quite specific to being categorically vague. This, again, was not unexpected given the background of the participants. In their TI evaluations, they tend to be more global than specific in their feedback.

As far as drawing examples from the performance on video, again it is difficult to assess. Because the participants were viewing the video as they were verbalizing their feedback, it

seems safe to assume that to at least some extent they were commenting on what they were watching. The true extent to which the feedback was directly related to the video is difficult to assess, again because we only have audio recordings of the feedback sessions.

Once again, not very many instances of the performer providing feedback that were future-oriented were found in the review of the audiotape transcripts. One example was in session 4TU2-2, where the performer was speaking about a sound that he made during his presentation, and that he wanted to use to more advantage in future presentations: "But I like that ... honking. And I can work that, I can milk that."

As for the number of areas covered in the feedback session, there was consistency between sessions. A count of the number of different behaviors referenced in the seven feedback session transcripts was conducted, using the described areas from the Specific Area of Interest Scale (sample provided in Appendix B6). The largest number of behaviors on which comments were made was eight, while the least was four. The average was six. The behaviors most frequently discussed (in 5 of the 7 sessions transcribed) were inflection / pitch, pausing and movement about the presentation area. The areas least frequently addressed were breathing / support (in none of the 7 sessions transcribed), volume and posture (in only 1of the 7 transcribed sessions).

In summary, some of the participants in the DF treatment condition followed the guidelines for the feedback sessions fairly well, but others did not. It cannot be concluded that the DF treatment was distinct from the EF treatment.

4.3.3 Post-hoc Analyses Summary

In summary, the results of this analysis of the audiotapes revealed that, although participants in the EF condition (evaluative feedback) tended to follow the guidelines fairly well,

participants in the DF condition (descriptive feedback) did not conform to the guidelines very well in some cases. For example, the partner made inappropriate comments or led the feedback discussion. While the partner in the EF feedback sessions may not have fulfilled the role perfectly, the impact on the study would not be as important as in the case of DF participants. In order to interpret the observed lack of difference between treatment groups it is necessary to establish that the feedback conditions were in fact distinct. This criterion appears not to have been met.

4.4 Post-program Survey Results

The fact that all participants had easy access to e-mail made this an ideal method of distributing and completing the open-ended questions. As promised, after each participant's completed questionnaire was received, the researcher delivered the videotape to the participant. A supplemental question (which arose during the analysis phase) was e-mailed to each participant approximately one month after the program was completed.

The method employed produced a response rate of almost 100%. One participant became ill after the program, and did not complete the interview questions (the participant did provide a response and received the video later on; several months had elapsed, however, so these responses are not included in the results). The following summary is based on the responses of the other 31 participants. Note that not every participant had a response to every question, although each participant did respond to most of the questions.

4.4.1. Post-program Survey Question 1

In response to Question 1 (*What did you learn from participating in the presentation skills program? Did you improve in your presentation skills?*) a clear majority of the participants (26) indicated that they felt they had improved in their presentation skills, while 4 felt that they

had not improved. The 4 that felt they had not improved were evenly split between the DF and EF treatment conditions. One person did not actually answer the question, although they did provide a lot of verbiage. A summary of the responses to the question of whether the participants felt they had improved in their presentation skills is presented in Table 4.6.

Table 4.6

Results of Post-program Survey Question 1: *Did you improve in your presentation skills?*

Improved	Did Not Improve
I improved a great deal.	I do not feel I improved my presentation skills from the sessions.
I feel that my presentation skills improved more during the few weeks of your program than in the previous eight months of Toastmasters ...	Well, I must have improved some of the skills. Nevertheless, the improvement is so little that I hardly recognized it.
Yes, in my opinion I improved the presentation skills.	
My presentation skills improved dramatically.	

4.4.2. Post-program Survey Question 2

As for their views on the feedback they received (Question 2), the respondents indicated that in general the feedback was constructive and positive. A summary of the responses to Question 2, separated by treatment condition, is presented in Table 4.7. As can be seen in Table 4.7, not all respondents found the feedback to be useful.

4.4.3 Post-program Survey Question 3

In response to Questions 3 (*What did you like/dislike about the feedback sessions?*), participants in the DF treatment condition seemed to want to receive feedback from their partner on their performance, as they normally would in a Toastmasters feedback session. As one participant put it, she disliked "the inability of getting other people's views on the feedback as well as sharing

Table 4.7

Results of Post-program Survey Question 2: *What were your views on the feedback you received during the program?*

Positive	Negative
Descriptive Feedback Participants	
Given that the feedback on each of my "public speaking" performances was provided by my toughest critic (i.e., myself), I think the program allowed for a very unusual and reliable means of capturing such a "brutally honest" feedback.	The feedback was neutral, just fine for the first 2 or 3 times but you can only get so far without some feedback which you can accept as being appropriate and to the point or discount.
I found the feedback very useful and helpful. I have modified my presentation.	The feedback was inconsistent because it came from a different person each night, and the "rules" of the feedback were so unfamiliar that we spent more time trying to follow the rules than providing constructive feedback.
Self-feedback reinforced the images seen on the video. The video provided 80% of the benefit, the verbal review 20%.	
Evaluative Feedback Participants	
The feedback was useful in helping me see both strengths and weaknesses.	I would give the feedback a C grade; the two messages were: walk from one side of the room to the other and exaggerate/emphasize some words.
The feedback that I received during the program was excellent.	
I found everyone to be direct and honest about the feedback they gave me. The feedback was always delivered in a positive way but included pointers that I needed to hear.	

my views on their performance. Perhaps that opportunity could have been added at the very end of the feedback session in order to allow the buddy/coach to confirm or add to the observations made by the participant (but not debated)." Another participant echoed these thoughts: "I am the kind of individual who likes feedback from others. I like to evaluate comments – take what I

think will help – and use it." One participant found that she was too critical of herself and wanted other peoples' opinions in order to make herself feel better. "I'm very critical about myself and I like to hear other's opinions as they tend to boost me more than my own evaluation."

Other DF participants liked the fact that they were responsible for evaluating their own skills. As one participant explained, she liked the fact that "I was the one evaluating my skills." Another participant had similar thoughts: "I liked it all. Having to provide feedback in a systematic way using the topic areas on the suggestion sheet made sure I didn't miss thinking about anything."

The major problem that DF participants reported was that there was no time limit for the feedback sessions. "I disliked the lack of structure in terms of time limits for each person." This comment was balanced by another participant, who said the opposite: "I liked that we could take our time with the self-evaluations." The negative side may be explained further by the following comment: "I disliked that some people were very rushed once their feedback part was over." Perhaps it was a matter of the partner not being focused or interested once their session was complete rather than a critique of the process itself.

Participants in the EF treatment condition indicated that they liked the ability to view the performance on video and to discuss the performance with the video as a reference. In the words of one participant, "I liked the idea of viewing the speech on video and being able to stop the tape at certain points to discuss issues as we viewed them." Other things that the participants liked included reviewing the video very soon after the performance, receiving feedback from only one person at a time, and getting opinions from more than person (over the course of the program).

4.4.4. Post-program Survey Question 4

In response to Question 4 (*What did you like / dislike about the program in general (location, format, etc.)*) the respondents indicated that they thought the program was well run in terms of the location, timing and flow. The flexibility to participate on different nights if schedule conflicts arose was much appreciated. The main negative comments were in reference to the parking situation at the University of Ottawa ("Parking was an adventure", according to one participant). The only other negative comments came with respect to the variety of the presentations; many participants gave the same speech five or six times, and some participants indicated that they would have preferred more variety in the presentations. Interestingly, most participants thought their *own* presentation was interesting, but they would have preferred more variety in *other* peoples' presentations.

4.4.5. Post-program Survey Question 5

In response to Question 5 (*How would you change the program (esp. the feedback sessions)?*), the respondents indicated that they would like more feedback from different people. This was fairly universal across the treatment conditions. In addition, participants in the DF treatment condition indicated that they would like to have received feedback from their partner during the feedback session.

The only other suggestion for changing the program had to do with the number of people in the room for the presentations. Several participants noted that it was difficult to present to a small audience, and more difficult to present to that same small audience week after week.

Two participants raised the issue of the timing of the feedback, that is, whether it should be delivered immediately before the next performance or after the last performance. They both suggested a format where the performer would present before the audience, receive feedback on

the performance, and then immediately present again before the audience. The performer would then receive feedback on the second performance. In one proposal, the performer would present again! While interesting in the abstract, practical considerations prevent the implementation of these suggestions, at least in a volunteer-based program.

4.4.6. Post-program Survey Question 6

In response to Question 6 (*Did you incorporate the suggestions made in the feedback sessions into your performances? If yes, how, and with what effect?*) the respondents were again very positive: 28 indicated that they had incorporated the suggestions made in the feedback sessions into their performances, with varying degrees of effect. Only one participant indicated that they had not incorporated any suggestions into their performance.

One participant in the EF treatment group stated, "Yes, I've incorporated suggestions into my presentations and the results have been positive. I still have a long way to go ... but at least I'm aware of them and will continue working on them. The key was to have them identified and your program helped me in this area."

A participant in the DF treatment group reported, "Yes. One of the suggestions from the second last presentation was that I should be more expressive both with my voice and my gestures. I made a drastic change which I think was fun for the audience and certainly was more fun with me."

4.4.7. Post-program Survey Question 7

As for Question 7, which asked about whether they had received or given feedback outside the confines of the program, nine participants indicated that they had not given or received any feedback outside the program. Fifteen participants indicated that they had given and/or received feedback outside the program, to varying degrees. The extremes ranged from one

participant who said they talked to someone every evening on the way home (they purposefully drove home someone different each week) to several participants who discussed their own performances with other participants at their regular weekly Toastmasters meetings.

4.4.8 Post-program Survey Summary

The response rate of 97% (31 of 32 participants) is outstanding, but understandable given that the participants agreed that they would not receive the video compilation of their performances until they responded to the survey questions. Most of the participants provided responses to all of the questions. Some of the responses to separate questions were clustered together, but they were not that difficult to separate out.

In general, the participants seemed to enjoy the program and to have received value for their time commitment. Of the 30 participants who provided a response to the first question, 26 (87%) felt that they had improved their presentation skills, while 4 felt that they had not improved. Most of the participants felt that the feedback they received was both constructive and positive.

As for the participants' feelings about the program and feedback sessions, the response was generally positive, with few opposing views. Several practical suggestions were made that may help in the development of future programs. A number of participants indicated that they would have liked to receive more feedback from different people.

4.5 Summary of Results of Main Study

The overall failure of the experimental design to demonstrate significant differences between participants in the two treatment conditions is obviously a great disappointment. Unfortunately, the confound in the experimental design prevents us from drawing any firm conclusions one way or the other, with respect to the main hypothesis of whether the impact on

performance self-efficacy of predominantly descriptive feedback would be more powerful than that of predominantly evaluative feedback.

The only significant finding of the main study was that the SSRS did seem to be sensitive enough to detect a significant increase in the participants' self-rated performance self-efficacy from the pretest to the posttest. This finding is reinforced by the result that there were, as expected, no significant changes in the participants' scores on the GSE and SSE from pretest to posttest.

Chapter 5 – Discussion

5.0 Introduction

In this chapter, I will review the results of the pre-study and main study and discuss their implications. Then the limitations of the current study are examined. Third, the question of whether descriptive or evaluative forms of constructive feedback are more potent in fostering growth in performance self-efficacy is revisited. Fourth, suggestions for dealing with methodological problems are offered with the goal of establishing an agenda for further research. Finally, the general conclusions of the present study are presented.

5.1 Results of the Studies

In this section, the results of each of the two studies are reviewed and presented in chronological order.

5.1.1 Pre-study

5.1.1.1 Pilot study phase of pre-study

The pilot study phase of the pre-study, wherein the face validity of the Speaker Self-efficacy Rating Scale was assessed, was successful. The respondents offered several suggestions for refining the instrument. Several new questions were added, which enhance the scope of performance behaviors assessed by the instrument. Other questions were reworded based on suggestions from the respondents. The reasons for rewording the questions were either to improve the comprehensibility of the question or to make more specific the description of the behavior being assessed.

The revised SSRS is a valid instrument for assessing performance self-efficacy in the domain of public speaking. This validity claim is based on both the results of the pilot study and prior research. The methods employed seem to have been adequate to the task.

5.1.1.2 Pre-study (survey)

The pre-study served as an opportunity to generate normative data and establish the reliability of the revised version of the SSRS developed in the pilot study. The results of this phase of the study provide strong support for the use of the SSRS as the principal data collection instrument in the ensuing main study. The reliability coefficients and item inter-correlations are both in keeping with accepted psychometric parameters, indicating that the instrument is acceptably reliable for use in further studies with public speakers. Further, the items on the subscales of the SSRS (addressing general, social and performance self-efficacy) were answered in a statistically consistent fashion, indicating that they are addressing similar yet distinguishable facets of the same construct.

The data set produced in the survey includes members of both the professional and amateur public speaking organizations in Canada, respectively the Canadian Association of Professional Speakers and Toastmasters International. These data were generated from a broad range of respondents, from beginners to full time professionals, as well as respondents from all parts of Canada except Quebec. The entire population of professional CAPS members were invited to participate, while a large sample (approximately 25%) of the 4000 Canadian members of TI were invited. The norms for these two groups (CAPS and TI) can be used as a starting point for developing standards for testing.

It is important to note that all experiences with TI members were extremely positive. The main reason some declined to participate was due to strict adherence to TI regulations that others were perhaps more willing to bend for the benefit of their members and to help out the research effort. The assistance and participation of all members of the TI organization was exemplary. All in all, the results of the survey were good. Although more respondents would have been

desirable, the achieved response rate and number of respondents was sufficient for the present purposes.

In summary, the results of the pre-study showed that the SSRS instrument was valid and reliable for use in the main study as a measure of performance self-efficacy, the dependent variable for the main study. No refinements to the instrument were made as a result of the pre-study. The obtained scores on the SSRS represent norms for the populations that will participate in the main study.

5.1.2 Main Study

The main study suffered from internal validity problems that compromised the experimental design employed. Unfortunately, this means that the results comparing the descriptive and evaluative feedback conditions cannot be trusted. Conclusions regarding the differential impact of descriptive and evaluative forms of constructive feedback cannot be made with any confidence based on these data.

An encouraging observation of the main study was that, of the 32 participants who actually took part in the program, all 32 completed all of the sessions. This indicated that the participants saw the presentations skills program itself as beneficial. This is heartening because motivation to participate in the program is essential to conducting this sort of research.

Further, the results of the main study revealed several indications that support the validity of the instrument. First, when collapsing the groups together, all of the participants showed a significant increase in their scores from pretest to posttest on the performance self-efficacy subscale of the SSRS and the Specific Area of Interest scale (SAIS). No difference was seen in the scores on the general and social self-efficacy subscales. This indicates that the SSRS and

SAIS were, in fact, sensitive to changes in performance self-efficacy, which was the dependent variable of interest in the present study.

Secondly, when the participants were treated as separate groups, similar differences were seen on the SAIS scores (although not on the PSE). This indicates that the difference in the PSE subscale scores was masked by the smaller sample size, and was only revealed through the increased power of the larger sample (all participants taken as a whole).

In summary, the results of the main study were limited to conclusions with respect to the instruments utilized. As for the implementation of the main study, devastating problems arose. These problems are reviewed in detail below in the section on limitations of the main study.

5.2 Limitations of the Study

In this section, limitations of each phase of the current study are presented and potential remedies are explicated. The phases are addressed in chronological order: pre-study and main study.

5.2.1 Limitations of the Pre-study

5.2.1.1 Pilot study limitations

With respect to the scholarly literature review, the ERIC, PSYCINFO and SPORTDISCUS on-line databases were queried for articles and papers that were matched to the keywords Public Speaking, Feedback, Evaluation, Assessment, Self-efficacy, Toastmasters, Presentation, Speaking, Video and various combinations of these terms. The limited number of articles and papers culled from these databases were reviewed to determine if they had a bearing on the current research. The original dates searched in the databases were from 1970 through 1999. The search was regularly updated to capture more recently generated knowledge. Other data sources were also used, such as personal libraries.

To improve this search method, a broader time frame could be queried. However, this would not necessarily increase the number of relevant articles and papers found, because these earlier pieces of import are likely found in the reference sections of later literature. A second method would be to expand the scope of the keywords searched. Again, there would not likely be a great increase in the number of relevant articles and papers found since the keywords used were quite comprehensive.

Including more participants in the pilot study phase may have provided a broader perspective on the issues, but would not necessarily have resulted in more information than was collected from the sample of 10 (seven of whom replied). The reviewers selected were both knowledgeable and accessible, accessibility being a key factor in all research in the social sciences. The use of members of both Toastmasters International and the Canadian Association of Professional Speakers as reviewers adds further support to the claims of face validity.

In summary, the methodology utilized for the pilot study could have been improved in terms of the literature review and the number and quality of reviewers for the SSRS instrument. The methods employed, however, seem to have been sufficient for the purposes of the present study.

5.2.1.2 Pre-study (survey) limitations

In the pre-study (survey), there were several limitations. The two main areas of concern were the decision not to translate the SSRS into French, and the recruitment of the participants. We will look at each of these areas in turn.

The main reasons the SSRS instrument was not translated into French for use in the Francophone community were cost and time constraints. This represents a weakness in the thoroughness of the data set that could be remedied. In doing so, however, standardization and

validity issues would have to be addressed in order to ensure the equivalency of linguistic forms and to eliminate differential item bias.

To improve the methodology for the pre-study, procedures for recruiting participants would be an obvious focus. The recruitment of the CAPS participants was facilitated through a personal relationship with the president of the national board of CAPS. This significantly decreased the turn-around time from first contact through approval for using their monthly mailing as a contact point with participants. In addition, it was possible to obtain a letter of support from the president, which may have helped encourage participation. Even so, the response rate was not different than that achieved for the TI participants. Perhaps recruiting CAPS participants at the chapter level would produce a greater response rate.

No such access to a national organization was available in the case of TI. Although a large sample with an adequate response rate was achieved, a larger sample is almost always desirable. To this end, budgeting more lead time to contact local club presidents and district governors would have been prudent. As well, future researchers could attempt to access large numbers of TI participants at their annual national, district and zone meetings.

In summary, the methodology utilized for the pre-study could have been improved in terms of improving the recruitment of both CAPS and TI participants. The methods employed, however, again seem to have been sufficient for the purposes of the present study. It should be noted that, despite the different methods used to obtain the two samples, the exact same response rate of 34% was achieved.

5.2.2 Limitations of the Main Study

In this section we will examine the limitations to the main study. The central problem was that the two treatment groups did not differ from one another sufficiently. The classic

Campbell and Stanley framework for internal validity (adapted by Posavac and Carey, 1992) will be reexamined as providing support for the current design choices. The contamination will then be explicated in terms of the supervision of the feedback sessions and the interaction of the participants outside the confines of the program.

5.2.2.1 Potential threats to internal validity

Campbell and Stanley (summarized by Posavac & Carey, 1992) describe seven threats to internal validity. The first two threats to internal validity outlined by Posavac and Carey, *maturation* and *history*, deal with real changes that occur in participants that are not directly attributable to participation in the program. Essentially, changes that the population from which the sample studied was drawn would undergo independent of whether they participated in the program or were even aware of the program's existence. Natural changes that occur in people over the course of time are referred to as *maturation*. In the present study, maturation was not an issue, as the groups would not have had systematically different experiences during treatment.

Community or societal events that occur during the duration of a program may also cause or contribute to behavioral changes in program participants. *History* is the term Posavac and Carey (1992) use to refer to these events. Examples of historical events that may contribute to behavioral changes include such events as political upheaval and economic recession. History becomes a problem when one group's experiences differ from another. Again, there were no significant large-scale societal events that occurred during the course of the program that might affect one group over the other. One possible confounding factor is that the participants in the main study were also attendant at their regular TI meetings throughout the course of the study. That is, some of the participants from each group were members of the same TI clubs, and thus

had the opportunity to interact outside the program. But since participants were randomly allocated to groups the concern is one of contamination, not history.

Although the participants were asked not to discuss the program outside the confines of the planned activities, several participants admitted discussing such issues with other TI members. This was to be expected, given the novelty of the program (none of the participants had been videotaped presenting before participation in the present study). In the future, it might be worth sacrificing the advantage of random assignment to groups to ensure that members who may see each other outside the confines of the program are in the same treatment condition. A quasi-experimental matched group approach might suffice.

The next three threats to validity referenced by Posavac and Carey (1992) deal with the sample under study. *Selection* refers to characteristics of the program participants in one treatment group compared to the other. The random allocation of participants to groups ensured that selection was not operative as a biasing factor.

Mortality refers to subject attrition during the course of the study. In the present study, this was not really an issue. Of the 33 participants who actually arrived at the preliminary meeting, 32 completed the program. The one who did not complete the program was well-known amongst the TI participants for starting and not completing similar programs. Otherwise, although accommodations to suit last-minute problems had to be made, 31 of the 32 participants completed all aspects of the program. One woman became ill after the program was finished but before the post-program survey was completed. Because her answers arrived long after the end of the program and she reported not being able to remember things clearly, her responses were not included in the post-program survey analysis. Her other data, which was completed during

the course of the program, were included in the analyses. Thus mortality did not affect the groups differently.

Regression to the mean refers to the tendency for extreme performance levels to be followed by less extreme performances. In plain language, the average score, or mean, is the expected score for any given administration of a test. Participants who score exceptionally high or exceptionally low on instruments at the pretest interval would be expected to yield scores closer to the mean on subsequent administrations, independent of the intervention. This threat to validity is not an issue in the present study, since there were no extreme scores at pretest for either group and the scores on the more stable general and social self-efficacy scales stayed fairly constant, while the scores for all participants on the performance self-efficacy scale increased significantly.

The final two threats to internal validity are *testing* and *instrumentation*. These two threats are a byproduct of the researchers and the data collection means employed. *Testing* refers to changes in the participants' behaviors attributable to the observation techniques employed. Posavac and Carey report that the pretest / posttest design is particularly susceptible to this form of bias. Traditionally, on unfamiliar performance and ability tests, participants tend to improve their scores on the posttest, if only because of increased familiarity with the instrument.

It is possible, but unlikely, that the participants in the present study improved their scores due to increased familiarity with the instrument. This is unlikely because data revealed improvement on only one of the subscales, and not the other two. If this threat was in operation, one would expect the participants to have improved on all aspects of the instrument, not just one subsection. Nevertheless, testing would not have affected the treatment groups differently given the random allocation of participants to groups.

Instrumentation refers to the objectivity of the instruments employed. The less objective the instruments, the more potential there is for scoring bias to enter into the equation, skewing the results unintentionally. In addition, the examiners may become more skilled at administering or interpreting the tests from pre- to posttest. This type of change, unrelated to the intervention effect desired, could affect either the grading of instruments or even impact on the participants perceptions of the instruments or procedures.

In the present study, there was no apparent instrumentation effect. The instruments were as objective as is reasonably possible. The researcher carefully reviewed the responses after they were submitted to ensure that the participants had properly completed the instruments. There was no pressure on the participants to complete their responses in a given time frame, and the researcher made sure that any questions about completing the instrument were answered to the satisfaction of the participant. There was no interpretation required of the responses, because the responses were to Likert-scale items. Scale scores were simple linear combinations of the items, which further ensured bias-free assessment.

It may be concluded, from the review presented above, that the threats to internal validity outlined by Posavac and Carey (1992) were not the source of the problems encountered in the present study. Unfortunately, the present study did suffer from other problems, including a larger number of participants than anticipated, a short time-frame for completing the program, and cross-group contamination during the implementation of the program.

5.2.2.2 Implementation problems

While the researcher and assistants did their best to ensure that the participants were receiving the proper form of feedback, complete success, in this regard, was not possible under the present circumstances. One reason for this failure was the number of participants. Only 16

were expected, but 35 expressed serious interest and 32 actually completed all aspects of the program. Two people who signed up for the program failed to show up at all, and there was one drop-out, mentioned earlier. This large sample size was somewhat daunting for the researcher and assistants. During recruitment, the prospect of obtaining a large enough sample size loomed as a much larger issue than the more practical issues of how to adequately supervise a larger sample than expected. Future researchers would be well advised to carefully consider the benefits of capping enrollment, especially if resources are limited.[7]

A second, and related, factor was the compressed time frame of the main study. The entire program was completed in four weeks. The training sessions were on a Saturday, the feedback sessions were the ensuing four weeks, and the final session was the Saturday at the end of that week. This short time frame was a big selling feature to the participants, but also posed a significant management challenge for the researcher. Because of heavy demands, only limited time could be devoted to reviewing the feedback session audiotapes. A more intensive regime of reviewing audiotapes may have provided a better basis from which the researcher could intervene in order to ensure that the treatment conditions were distinct.

One might also question the overall design of the study: Was an experimental setting the correct design to use in order to study this phenomenon? The power of the comparative groups design and the added strength of random assignment to groups make the experimental design, although difficult to achieve in non-laboratory conditions, extremely tempting to use since it provides the best support for causal claims. In the present case, however, the implementation of the experimental method proved not to be ideal.

---

[7] It should be noted that there are methodological advantages and disadvantages to both larger and smaller sample sizes. Also not to be discounted, however, is the impact on the statistical power of the design, which is also dependent on the sample size. In essence, statistical power is the probability that a design will correctly detect statistically significant differences when they in fact exist. For more discussion on this topic, see, for example, Cohen (1992).

The need to clearly dichotomize the treatment conditions in an experimental study is of paramount importance, and that was highly difficult to achieve in the current study. There are other designs that can allow for comparisons between the different forms of feedback but which do not have the same stringent requirements of the treatment conditions. For example, more qualitative methods could be employed which might allow for less strictly compartmentalized treatment conditions and yet still provide insight into the differential impacts of descriptive versus evaluative feedback on the participants.

Quasi-experimental approaches where assignment to groups is non-random might also be useful. It would be desirable to completely separate the treatment groups but this would be difficult to do practically. These alternate approaches are explicated later in this document.

Finally, and most importantly, there seems to have been insufficient monitoring of the feedback sessions, especially the DF condition sessions. Although great effort was made to ensure that the participants understood the desired format for the feedback sessions, and the researcher and assistants monitored as closely as possible and intervened when necessary, there were gaps in the system attributable to the larger than expected numbers.

The short time frame did allow for the participants to participate fully without interfering unduly with their schedules. Participants indicated that the short time frame was an important factor in their decision to participate. The week-by-week feedback sessions allowed the participants to see how they were progressing on a regular basis, and probably allowed the feedback sessions to build on one another incrementally over time.

Unfortunately, the short time frame and large number of participants also had drawbacks: limited time for the research team to review the audiotapes from previous feedback sessions while preparing for the next sessions and troubleshooting as events occurred. The benefits to the

participants of the short time frame turned into serious problems from the viewpoint of the implementation of the research project.

Obviously, the supervision of the feedback sessions was not adequate to ensure that the treatment conditions were distinct. Several factors were discussed that explain how the failure occurred. The two main reasons were that there were just too many participants and not enough time to properly supervise them.

5.2.2.3 Cross-group contamination outside the program

The fact that the participants interacted outside the confines of the program came as no surprise. The participants continued to attend their regular TI meetings throughout the course of the program. Several participants were members at the same TI clubs. The researcher attempted to dissuade the participants from seeking extra feedback until the end of the program, a request that was made at the initial training sessions and repeated each week at the feedback sessions.

Unfortunately, some participants could not resist the temptation to receive feedback from more than one source. The extreme case was one participant who intentionally drove home a different participant each week in order to solicit varying commentary on his or her own performance. This was both unexpected and not in keeping with the instructions to the participants. Unfortunately, it was not revealed until the post-program survey responses were received.

There were other incidents of interaction between the participants outside the parameters of the program. This interaction took the form of informal discussions between participants at convenient times, such as waiting for rides or at their TI meetings. There were no other reported systematic attempts to receive feedback from other participants.

In summary, it seems that the participants wanted even more feedback than was being provided in the sessions. They were even willing to "break the rules" to obtain this feedback. This is in keeping with the educational literature that showed teachers want more feedback of any kind than they were currently receiving (e.g., Natriello, 1984). Unfortunately, it also contributed to the failure of the treatment conditions to be sufficiently distinct to allow a valid test of the central hypothesis.

5.2.3 Summary of Limitations

In summary, we have seen that there were several limitations to the present study. Comparatively speaking the pre-study was well implemented and suffered only minor drawbacks. However, the main study was hobbled by cross-group contamination. While there did not seem to be a problem with the internal validity of the main study as outlined by Posavac and Carey (1992), implementation challenges proved to be devastating. These challenges included unwieldiness in implementation due to larger than expected sample size and the tendency for participants to disregard directions with respect to obtaining feedback outside the confines of the program.

5.3 Descriptive Feedback as Theoretically Compelling

The failure of the present study to provide compelling evidence that descriptive forms of constructive feedback are more potent than evaluative forms of constructive feedback in no way diminishes the allure of the theoretical argument on which the hypothesis was based. The result was attributable to methodological inadequacies that surfaced in a domain of inquiry where little empirical research exists.

To recap, the essence of the argument for descriptive as opposed to evaluative forms of constructive feedback is rooted in the theories advanced by Tracy and McNaughton (1989). They

argue that the traditional approaches to formative appraisal favored by neo-traditionalists have not produced the promised results in terms of improved performance. In this form of appraisal, the supervisor typically assumes the role of the expert, interpreting the observation data for the 'novice' appraisee. The supervisor explains how the appraisee should change their behavior in order to improve their performance. The appraisee is not so much involved in the interpretation of the data as he or she is a recipient of interpreted data. This form of appraisal has become the norm in education and other fields and is championed by Hunter (1988a, 1988b), among others.

Tracy and McNaughton (1989) urge the use of more descriptive feedback in growth-oriented appraisal. In this neo-progressive view, the appraisee is intimately involved in the interpretation of the data, with the supervisor assisting where needed. This descriptive feedback consists of vivid description of the performance, without evaluative commentary on the quality of the performance or the outcomes of the performance.

In essentially being forced to make interpretations, appraisees develop their capacity to critique their own performance. The concept of self-evaluation has seen increasing attention of late. Self-evaluation and peer evaluation are more and more frequently used in performance appraisal, at least as supplementary information. Similar trends have been seen in various cognate fields including student assessment and program evaluation.

For example, in student assessment, we have seen a shift towards authentic forms of assessment that engage students as judges of their own performance (Stiggins, 2001). Portfolio assessment, performance assessment and students self-evaluation of their own work are all examples. A similar trend has been seen in the field of counselor training, where portfolio assessment and videotaped feedback sessions are also being used more frequently. The main reason cited for this increase in the use of self-appraisal techniques is that individuals need to

learn how to assess their own work (i.e., develop their capacity for self-critique). This aligns with Tracy and McNaughton's (1989) call for the use of more descriptive forms of feedback in formative appraisal, for similar reasons: the appraisee should serve as the judge of their own performance.

A similar trend has been seen in program evaluation, wherein the inclusion of program stakeholders in decision-making and data-interpretation roles is considered to be important. The rise in popularity of collaborative, participatory and empowerment evaluation approaches is tied to the utility of involving members of the program committee as judges of their own programs (Cousins & Whitmore, 1998). Again, this development parallels recent directions in formative teacher appraisal as conceptualized by Tracy and McNaughton (1989).

Finally, the work on self-reflection developed by Schön (1983, 1987) and others provides a potentially useful framework to motivate and guide further research into the relative merits of descriptive and evaluative forms of constructive feedback. Perhaps a study involving self-reflection, either in place of or in addition to self-efficacy, would provide more useful and detailed information as to how exactly different forms of constructive feedback impact on the participants.

In summary, the theoretical arguments underpinning the research questions detailed in the introductory section of this thesis remain compelling, and unanswered. Essentially, the present study can be thought of as a design implementation failure, rather than an objective test of the hypothesis. No evidence has been offered that either confirms or refutes the hypothesis.

5.4 Possible Solutions to Implementation Problems

The problems of implementation seen in the main study raise the question of whether it is possible to address this hypothesis in the manner proposed. We will look at this in four ways.

First, how could the current study have been done more effectively? Second, are there alternative approaches to studying this research question that could be utilized? Thirdly, are there additional post-hoc analyses that could yield useful results? Finally, what related research questions can and should be asked?

5.4.1 Improving the Present Study

Given that the SSRS was established as valid, reliable and sensitive to changes in the participant's performance self-efficacy, how could the program have been implemented in order to help assure trustworthy results? In order to address this issue, we must first dispense with the realities of financing the project. Within the parameters of available resources, an honest attempt was made to test the hypothesis. Unfortunately, the sample size achieved overwhelmed the available resources, resulting in unsuccessful monitoring of the feedback sessions.

In the ideal case, the researcher would have the resources (e.g., time, money) to train enough research assistants (RAs) so that there would be at least one available per feedback session. Having a dedicated research assistant in each feedback session, rather than the rotating visits utilized in the present study, would have enhanced the process. The RA could monitor each session from start to finish, ensuring that the proper type of feedback was delivered throughout. Alternatively, the RA's could be trained to deliver the feedback appropriate to the treatment condition, eliminating the need for the participants to do so. Such steps would help reduce if not eliminate cross-group contamination.

Another addition that would have been helpful would be the use of video cameras to monitor the feedback sessions. This would allow more in-depth analysis of the feedback sessions in terms of body language, in addition to the content components of the feedback. However, this

addition would be predicated on increased time available for reviewing the videos and taking necessary steps to intervene.

The design could also have been improved by increasing the duration of the program. More time at the beginning, for training, would have been most useful. The participants would have more time to practice, and to show that they understood the expectations of the program. More feedback sessions would also have allowed more time for the participants to increase their performance self-efficacy. Optimally, 10 or more feedback sessions spread over several months would have been desired. This would also allow more time in between performances for implementing the suggestions provided in the feedback sessions.

A longer time period would also have been conducive to the use of the Peer Performance Rating scale (PPRS), which would serve as a tool for triangulating the measurement of participants' performance self-efficacy. This triangulation would increase the confidence in the results of the SSRS results, which indicated that the participants' increased in their performance self-efficacy. The PPRS would provide an independent assessment of the performance.

The longer time frame outlined above, and a more generous supply of research assistants, would also have permitted the different treatment conditions to be implemented at different times. While running the risk of introducing "history" as a threat to internal validity this strategy would reduce the amount of contact among participants in different treatment conditions.

In summary, the present study could have been improved in a number of ways, all of which were not feasible in the circumstances encountered during the research project. The availability of research funding, personnel and time would allow for several, if not all of these suggestions to be implemented. While the present case succumbed to limitations it represents a first attempt to study systematically the impact of descriptive feedback and thereby represents an

important contribution. Now that methodological challenges have been empirically identified they will be easier to handle in ongoing research.

5.4.2 Alternative Approaches

The experimental method utilized in the present study was just one of several ways of exploring the research questions. Other methods could have been used, perhaps more effectively, to address these research questions. For example, one might explore a quasi-experimental approach, where assignment to groups is non-random. A second option would be comparative qualitative methods, which could allow for less strictly compartmentalized treatment conditions and yet still provide insight into the differential impacts of the treatment conditions on the participants. In the next two sections, we will explore how these differing methods could be implemented.

5.4.2.1 Quasi-experimental approach

One alternate approach would be to carry out a quasi-experiment, similar to the present experimental design except that participants are not randomly assigned to treatment groups. Although this non-random assignment weakens the internal validity of the design, it does allow for a lot more flexibility in carrying out the research. Careful procedures could be employed to ensure that treatment groups are matched or that they do not differ in important ways, prior to treatment. The proposed research design would then be to carry out a project similar to the present study, but with the two treatment conditions operating at different times. For example, the evaluative group might run from September through October, while the descriptive group could run from November through December. Again the threat to internal validity of history would need to be addressed. In particular, ongoing participation in TI might lead to differential

skill sets across groups prior to treatment. The use of pre- and posttest measures would provide one way to control statistically for observed differences.

Participants could enroll in whichever group was more convenient to their schedule, as there is no requirement that the groups are equivalent at the outset (the objective of random assignment). The number of participants in each group would ideally be the same, but there are statistical techniques to get around this problem.

The advantage to this type of study is that it would provide flexibility for design implementation while taking steps to ensure the comparability of groups. The validated SSRS could be used as a quantitative data source for comparative purposes. The results of the two different sessions (descriptive and evaluative) could be compared, even though they occurred at different times and with different participant groups. This comparison, as noted above, would be predicated on the assumption that history would not differentially impact the groups.

5.4.2.2 Qualitative approach

Another interesting approach would be to adapt the methodology utilized by Kilbourn (1990). A qualitative study of this sort could be conceptualized as an exploration of the differences between one pair of participants carrying out descriptive feedback on performance over an extended period of time, and another pair of participants carrying out evaluative feedback within a similar time frame. The researcher would assume the role of observer, sometimes present at the feedback sessions. Detailed data collection methods (e.g., video taping of performances and feedback sessions, reflective journals, etc.) would need to be developed for use in the study to capture the details.

The amount of time and energy required to carry out this form of investigation would be extraordinary, as acknowledged by Kilbourn (1990). Yet the potential benefits in terms of

understanding the complexities and dynamics of sustained descriptive feedback would be sizable. It would be desirable but not essential to include participants with similar characteristics (e.g., experience, interest, ability) in each group.

The benefit of this form of research would be the rich detail about the process, content and use of these forms of feedback that have not been studied systematically. Comparisons could be made between the two forms of feedback, using qualitative display techniques (e.g., Miles & Huberman, 1994). The main drawback to this type of study is the reduced generalizability of the results compared with more quantitative methods. This drawback seems to be greatly outweighed by the depth of information that could be collected about the nature and use of the feedback. The challenges of developing the instruments and recruiting the participants would be balanced by the valuable information that could be obtained through this in-depth study of the complex process of feedback.

5.4.3 Additional Post-hoc Analyses

Two additional post-hoc analyses[8] may be useful in enhancing our understanding of constructive feedback. The purpose of the first analysis is to gain a deeper understanding of the actual form of the feedback provided by the participants in the main study. The purpose of the second analysis is to determine whether participants who improved the most as rated by expert raters received feedback that differed from the feedback received by participants who did not improve as much. Methods for use in these two post-hoc studies are proposed below.

5.4.3.1 Actual forms of feedback provided in main study

In order to analyze the form of the feedback presented in the feedback sessions, the 70 feedback session tapes would need to be transcribed. Precision would have to be taken in

[8] It should be noted that these post-hoc analyses are presented as recommendations for future studies, not as components of the current thesis.

determining the components of feedback that are to be assessed (e.g., the evaluative content of the feedback). Similarly, the methods for assessing these components would have to be clearly explained. For example, the characteristics of descriptive and evaluative forms of feedback must be clearly operationalized in order for the raters to be able to identify each form in the feedback session transcripts.

Next, the reviewers would need to be trained in rating the feedback session transcripts according to the specifications spelled out in the previous paragraph, and a standard would need to be established that these raters would need to meet. Triangulation methods would need to be employed in order to ensure the reliability of the ratings. An example of a triangulation method that could be employed to ensure inter-rater reliability is for a random sample of feedback session transcripts to be rated by more than one rater. The consistency of the ratings across transcripts could be compared, yielding a measure of consistency.

The transcripts of the feedback sessions would then be content-analyzed. The raters would be blind to the treatment condition the transcripts represented in order to minimize bias. This 'blind' rating would also allow an assessment of whether participants in one treatment condition provided feedback that more closely resembled the desired form of feedback than participants in the other treatment condition.

The results of this analysis could provide more information as to the nature of the feedback provided in the feedback sessions of the main study. The importance of operationalizing the parameters to be assessed in the transcripts cannot be overemphasized. As was shown in the present study, operationalizing these parameters would be difficult, if not impossible.

5.4.3.2 Study of participants who showed the most improvement

In order to determine whether participants who improved the most received feedback that differed from the feedback received by participants who did not improve as much, we must first determine levels of improvement on a scale of improvement. Expert raters would need to be recruited and/or trained in order to assess the initial and final performances of the participants. These expert raters would have to develop a shared understanding of what constitutes improvement in public speaking performance. This shared understanding could be achieved through various methods, including the researcher providing operational definitions of the variables involved, the raters coming to a consensus definition, or even through factor analytic methods. Criteria for deciding which group a participant belongs to would need to be established. A decision as to how many participants to analyze would need to be made. The decision could be to include all participants, or only those who represent extreme cases (i.e., the most and least improved participants as rated by the experts).

Once the participants are categorized, the feedback session tapes would need to be transcribed, and the sessions that each participant was involved in would need to be identified. Essentially the same analysis as described in the previous study would need to be done. The components to be assessed need to be operationalized, and the raters must be trained to properly assess the transcripts. Triangulation methods similar to those outlined earlier would also need to be implemented to ensure the reliability of the ratings.

The results of this content analysis would provide information as to whether those who demonstrated improvement in their public speaking performance as rated by experts differed in the form of feedback they received in the feedback sessions. Differences between the groups in

the form of feedback received may provide direction for future research concerning constructive feedback.

The analysis would be complicated by the fact that, in most cases, each participant had a different partner for each feedback session. This would make it difficult to assess the overall form of feedback provided to a participant, when they may have received markedly different forms of feedback throughout the multiple sessions.

5.4.3.3 Summary of additional post-hoc analyses

The results of these additional post-hoc analyses may be useful in enhancing our understanding of descriptive and evaluative forms of constructive feedback. The most important factor in such analyses would be clearly delineating the components of feedback to be assessed, and the methods of assessing those components. Another important factor would be the training of the raters.

5.4.4 Emergent Research Agenda

Other research questions arise that may be important to examine, even before trying to replicate the current study with proper implementation. One question is whether or not, in fact, people can be trained to deliver descriptive feedback. This is a key assumption of the present study and the alternatives proposed above. Given the natural tendency for people to gravitate to evaluative feedback it will be important to show that such training can be successful and under which circumstances. A second question relates to the self-report nature of the measure of performance self-efficacy. The development of more objective measures for use in tandem with the SSRS seems to be indicated.

5.4.4.1 Training people to deliver descriptive feedback

Training people to give descriptive feedback may be more difficult than it seems. Showers and Joyce (1996) report that participants in their peer coaching programs tend to gravitate towards evaluative feedback from the desired descriptive feedback. To counter this tendency, Showers and Joyce have been forced to remove the verbal feedback component from their programs altogether. Participants now provide strictly written feedback that can be screened for evaluative content before it is delivered to the appraisee. This is just one example of how difficult the provision of descriptive feedback can be.

How could one go about investigating whether people can be trained to deliver descriptive feedback? First, one would have to clearly define exactly what descriptive feedback is, as differentiated from evaluative feedback. This was done in the present study but the methodological challenges that emerged suggest that this strategy is very important. In order to investigate whether the specific skill of delivering descriptive feedback can be taught, a very strict definition of the parameters of that skill is required and some sort of mechanism for checking the understanding of participants would be important.

Next, we would need to decide how to measure the level of skill in delivering descriptive feedback. While qualitative data were collected in the present case implementation could be measured more exactly through the use of some sort of rubric or scoring scheme. The use of a rubric necessitates the development of expert raters as well. A related topic would be what level of skill (as measured by our tool) is required in order to be considered 'capable'. The development of some sort of measurement tool would be beneficial, as would the development and training of expert raters.

Finally, the methods to be used would have to be decided. As an exploratory study, it would make sense to implement a pretest / posttest design, with an intervention (training program) occurring between the pre- and posttest. The pretest and posttest could employ a developed tool as discussed earlier. The intervention, or training program, would consist of multiple training sessions wherein the participants would be taught how to deliver descriptive feedback. The frequency and temporal proximity of training sessions are variables that could be manipulated and their potency in predicting skill development assessed.

5.4.4.2 Objective measurement of performance self-efficacy

The performance measures of the SSRS are strictly self-report by nature. More objective measures of performance might be useful, especially if they can be used as tools for triangulating the results of the SSRS. The Peer Performance Rating Scale (PPRS), developed for use in the present study but not implemented, may prove to be useful for this purpose. The PPRS is an adapted version of the performance self-efficacy (PSE) subscale of the SSRS (sample provided in Appendix B11).

One way to establish the validity of the PPRS would be to check the correlation between a performer's self-reported performance self-efficacy (as measured by the PSE) after a performance with the scores on the PPRS as determined by a group of observers. The results of this study would indicate how closely the observers' assessment corresponds to the self-assessment of the performer. High correlations would indicate that the instruments are measuring similar constructs (i.e., performance self-efficacy). Modifications to the instrument would be made as indicated by the results. These modifications could include the elimination, addition or rewording of items.

A second study could be done to assess the test-retest reliability of the PPRS. One method of assessing this facet of the instrument would be to have a group of observers rate a performance with the PPRS. The ratings would be collected, and then new rating sheets (identical to the first) would be distributed after a given interval (e.g., five minutes). The observers would then be asked to rate the same performance again, and the correlation between these two ratings would yield an indicator of test-retest reliability. This type of study can be done with or without an intervening task (i.e., other performances between the first and second rating of the criterion performance). It would be prudent to conduct a similar exploration of the test-retest reliability of the SSRS, as well.

The benefits of developing the PPRS as an objective measure of public speaking performance would not be limited to providing a new assessment tool. The process would also add weight to the validity of the SSRS.

5.4.5 Summary of Potential Solutions to Implementation Problems

In the preceding sections, we have looked at various methods of remedying the implementation problems observed in the present study. These approaches include methods of improving the experimental design used in the current study, two alternative approaches – a quasi-experimental design and a qualitative design – to studying the research questions, two additional post-hoc analyses, and two more fundamental research questions, one relating to the facility with which appraisers can be trained to deliver descriptive feedback and the other relating to objectivity in the measurement of performance self-efficacy. All suggestions hold promise for moving a research agenda ahead in this domain.

5.5 General Conclusions

In conclusion, the results of the current study are consistent with the findings of Showers and Joyce (1996). They found it necessary to remove the verbal feedback component of their appraisal sessions because the participants could not resist the tendency to gravitate towards evaluative feedback from the desired descriptive feedback. In the current study, the participants in the descriptive feedback treatment condition also tended to drift from descriptive to evaluative feedback to the detriment of the hypothesis test.

It is important to recognize that the present results constitute a failure of design implementation rather than a valid test of the hypothesis. Given the limited amount of empirical research in this domain such methodological challenges were difficult to anticipate. The problems encountered in the main study do not impact the validity or reliability of the pre-study instrument development (SSRS). Unfortunately, the question of whether descriptive constructive feedback is more potent than evaluative constructive feedback in increasing performance self-efficacy remains unresolved, but the importance of the question remains undiminished.

References

Acheson, K.A., & Gall, M.D. (1987). Techniques in the clinical supervision of teachers: Preservice and inservice applications (2nd edition). New York: Longman.

Alschuler, A.S. (1996). The portfolio emperor has no clothes: He should stay naked. Counselor Education and Supervision, 36, 133-137.

Argyris, C., & Schön, D.A. (1975). Theory in practice: Increasing professional effectiveness. San Francisco: Jossey-Bass

Baltimore, M.L., Hickson, J., George, J.D., & Crutchfield, L.B. (1996a). Portfolio assessment: A model for counselor education. Counselor Education and Supervision, 36, 113-121.

Baltimore, M.L., Hickson, J., George, J.D., & Crutchfield, L.B. (1996b). A response to the emperor's new clothes. Counselor Education and Supervision, 36, 138-140.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 84, 191-215.

Bandura, A. (1995). Self-efficacy in changing societies. New York: Cambridge University Press.

Bandura, A. (1997). Self-efficacy: The exercise of control. Thousand Oaks, CA: Sage.

Bandura, A., Adams, N.E., & Beyer, J. (1977). Cognitive process mediating behavioral change. Journal of Personality and Social Psychology, 35, 125-139.

Barber, L.W. (1990). Self-assessment. In J. Millman & L. Darling-Hammond (Eds.), The new handbook of teacher evaluation: Assessing elementary and secondary school teachers (pp. 216-228). Newbury Park, CA: Corwin Press.

Baron, R.A. (1990). Environmentally induced positive affect: Its impact on self-efficacy, task performance, negotiation and conflict. Journal of Applied Social Psychology, 20, 368-384.

Beck, J.J., & Seifert, E.H. (1983). A model of instructional supervision that meets todays needs. NASSP Bulletin, 67, 38-42.

Benshoff, J.M. (1994). Peer consultation as a form of supervision. (ERIC Document Reproduction Service No. ED 372352).

Benshoff, J.M., & Paisley, P.O. (1993). The structured peer consultation model for school counselors. Journal of Counseling and Development, 74, 314-318.

Berg, M.H., & Smith, J.P. (1996). Using videotapes to improve teaching. Music Educators Journal, 82, 31-37.

Biggs, S. J. (1980). The me I see: Acting, participating, observing, and viewing and their implications for video feedback. Human Relations, 33, 575-588.

Blair, B.G. (1991). Does "supervise" mean "slanderize"? Planning for effective supervision. Theory Into Practice, 30, 102-108.

Boyce, B.A., Markos, N.J., Jenkins, D.W., & Loftus, J.R. (1996). How should feedback be delivered? Journal of Physical Education, Recreation and Dance, 67, 18-22.

Bright, B. (1996). Reflecting on 'reflective practice'. Studies in the Education of Adults, 28, 162-184.

Brock, S.C. (1981). Evaluation-based teacher development. In J. Millman (Ed.), Handbook of teacher evaluation (pp. 229-243). Beverly Hills: Sage.

Bullard, B. (1998). Teacher selfevaluation. (ERIC Document Reproduction Service No. ED 428074).

Burke, P.J., & Fessler, R. (1983). A collaborative approach to supervision. The Clearing House, 57, 107-110.

Campbell, D.G. (1995, November). Toward refining the assessment of the basic public speaking course: An experimental study. Paper presented at the Annual Meeting of the Speech Communication Association, San Antonio, TX. (ERIC Document Reproduction Service No. ED 395349).

Campbell, D.T., & Lee, C. (1988). Self-appraisal in performance evaluation: Development versus evaluation. Academy of Management Review, 13, 302-314.

CAPS (1999). Membership and meeting planner guide of the Canadian association of professional speakers. Toronto: CAPS.

Carlson, R.E., & Smith-Howell, D. (1995). Classroom public speaking assessment: Reliability and validity of selected evaluation instruments. Communication Education, 44, 87-97.

Carney, J.S., Cobia, D.C., & Shannon, D.M. (1996a). The use of portfolios in the clinical and comprehensive evaluation of counsellors-in-training. Counselor Education and Supervision, 36, 122-132.

Carney, J.S., Cobia, D.C., & Shannon, D.M. (1996b). The use of portfolios in evaluation – revisited. Counselor Education and Supervision, 36, 141-143.

Chhokar, J.S., & Wallin, J.A. (1984). A field study of the effect of feedback frequency on performance. Journal of Applied Psychology, 69, 524-530.

Cogan, M.L. (1973). Clinical supervision. Boston: Houghton Mifflin.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155-159.

Cousins, J.B. (1988). Implications for performance appraisal practice from research on evaluation utilization. In E.S. Hickcox, S.B. Lawton, K.A. Leithwood, & D.F. Musella (Eds.), Making a difference through performance appraisal (pp. 158-174). Toronto: OISE Press.

Cousins, J.B. (1995). Using collaborative performance appraisal to enhance teacher's professional growth: A review and test of what we know. Journal of Personnel Evaluation in Education, 9, 199-222.

Cousins, J.B. (in press). Understanding participatory evaluation for knowledge utilization: A review and synthesis of current research-based knowledge. To appear in: D. Nevo & D. Stufflebeam (Eds.), International handbook of educational evaluation. Klewer Press.

Cousins, J.B., Donohue, J.J., & Bloom, G.A. (1996). Collaborative evaluation in North America: Evaluator's self-reported opinions, practices and consequences. Evaluation Practice, 17, 207-226.

Cousins, J.B., & Earl, L.M. (1992). The case for participatory evaluation. Educational Evaluation and Policy Analysis, 14, 397-418.

Cousins, J.B., & Simon, M. (1996). The nature and impact of policy-induced partnerships between research and practice communities. Educational Evaluation and Policy Analysis, 18, 199-218.

Crocker, L.M., & Algina, J. (1986). Introduction to classical & modern test theory. New York: Harcourt Brace Jovanovich College Publishers.

Darling-Hammond, L., Wise, A.E., & Pease, S.R. (1983). Teacher evaluation in the organizational context: A review of the literature. Review of Educational Research, 53, 285-328.

DeGregorio, M., & Fisher, C.D. (1988). Providing performance feedback: Reactions to alternate methods. Journal of Management, 14, 605-616.

Duke, D.L. (1990a). Developing teacher evaluation systems that promote professional growth. Journal of Personnel Evaluation in Education, 4, 131-144.

Duke, D.L. (1990b) Setting goals for professional development. Educational Leadership, 47, 71-75.

Duke, D.L., & Stiggins, R.J. (1986). Teacher evaluation: Five keys to growth. Washington, DC: National Education Association.

Duke, D.L., & Stiggins, R.J. (1990). Beyond minimum competence: Evaluation for professional development. In J. Millman & L. Darling-Hammond (Eds.), The new handbook of teacher evaluation: Assessing elementary and secondary school teachers (pp. 116-132). Newbury Park, CA: Corwin Press.

Earl, L.A., & Cousins, J.B. (1995). Classroom assessment: Changing the face, facing the change. Toronto: Ontario Public Schools Teachers' Federation.

Ellis, K. (1995). Apprehension, self-perceived competency, and teacher immediacy in the laboratory-supported public speaking course: Trends and relationships. Communication Education, 44, 64-78.

Ferris, G.R., Judge, T.A., Rowland, K.M., & Fitzgibbons, D.E. (1994). Subordinate influence and the performance evaluation process: Test of a model. Organizational Behavior and Human Decision Processes, 58, 101-135.

Fetterman, D.M. (1994). Empowerment evaluation. Evaluation Practice, 15, 1-15.

Fletcher, C. (1986). The effects of performance review in appraisal: Evidence and implications. Journal of Management Development, 5, 3-12.

Geis, G.L. (1986). Formative feedback: The receiving side. Performance and Instruction Journal, 25, 3-6.

Gibson, S., & Dembo, M.H. (1984). Teacher efficacy: A construct validation. Journal of Educational Psychology, 78, 569-582.

Goldhammer, R. (1969). Clinical supervision: Special methods for the supervision of teachers. New York: Holt, Rinehart and Winston.

Gordon, B.G., Meadows, R.B., & Dyal, A.B. (1996). School principals' perceptions: The use of formal observation of classroom teaching to improve instruction. Education, 116, 9-15.

Grimmett, P.P. (1990). A commentary on Schön's view of reflection. Journal of Curriculum and Supervision, 5, 19-28.

Grimmett, P.P., & Crehan, E.P. (1990). Barry: A case study of teacher reflection in clinical supervision. Journal of Curriculum and Supervision, 5, 214-235.

Haefele, D.L. (1993). Evaluating teachers: A call for change. Journal of Personnel Evaluation in Education, 7, 21-31.

Harris, I.B. (1989). A critique of Schön's views on teacher education: Contributions and issues. Journal of Curriculum and Supervision, 5, 13-18.

Hart, G.M. (1994). Strategies and methods of effective supervision. (ERIC Document Reproduction Service No. ED 372341).

Hosford, R.E., & Johnson, M.E. (1983). A comparison of self-observation, self-modeling, and practice without video feedback for improving counselor interviewing behaviors. Counselor Education and Training, 23, 62-70.

Huberman, M. (1989). Predicting conceptual effects in research utilization: Looking with both eyes open. Knowledge in Society, 2, 6-24.

Huberman, M. (1994). Research utilization: The state of the art. Knowledge and Policy, 7, 13-33.

Huberty, C.J., & Morris, J.D. (1989). Multivariate analysis versus multiple univariate analyses. Psychological Bulletin, 105, 302-308.

Hugenberg, L.W., & Yoder, D.D. (1995, April). Communication competence: A reaction to the 'Competent Speaker Speech Evaluation Form'. Paper presented at the Annual Meeting of the Eastern Communication Association, Washington, DC. (ERIC Document Reproduction Service No. ED 372432).

Hunter, M. (1988a). Create rather than await your fate in teacher evaluation. In S.J. Stanley & W.J. Popham (Eds.), Teacher evaluation: Six prescriptions for success (pp. 32-54). United States: ASCD.

Hunter, M. (1988b). Effecting a reconciliation between supervision and evaluation – a reply to Popham. Journal of Personnel Evaluation in Education, 1, 275-279.

Ilgen, D.R., & Moore, C.F. (1987). Types and choice of performance feedback. Journal of Applied Psychology, 72, 401-406.

Ilgen, D.R., Fisher, C.D., & Taylor, M.S. (1979). Consequences of individual feedback on behavior in organizations. Journal of Applied Psychology, 64, 349-371.

Joyce, B., & Showers, B. (1980). Improving inservice training: The messages of research. Educational Leadership, 37, 379-385.

Joyce, B., & Showers, B. (1982). The coaching of teaching. Educational Leadership, 40, 4-10.

Karl, K.A., & Kopf, J.M. (1994). Will individuals who need to improve their performance the most, volunteer to receive videotaped feedback? Journal of Business Communication, 31, 213-223.

Kilbourn, B. (1990). Constructive feedback: Learning the art. Toronto: OISE Press.

Kruger, M.J. (1985). Improving performance appraisals: Confronting subjective factors that influence ratings. Performance and Instruction Journal, 24, 2-5.

Larson, J.R. (1989). The dynamic interplay between employees' feedback-seeking strategies and supervisors' delivery of performance feedback. Academy of Management Review, 14, 408-422.

Latting, J.K. (1992). Giving corrective feedback: A decisional analysis. Social Work, 37, 424-430.

Lawton, S.B., Hickcox, E.S., Leithwood, K.A., & Mussella, D.F. (1986). Development and use of performance appraisal of certificated education staff in Ontario school boards (Volume 1 – Technical Report). Toronto: Ministry of Education.

Louis, K.S. (1995, May). Two steps forward, one step back: Do we need a new theory of dissemination and knowledge utilization? Paper presented at the Seminar for Research Dissemination and Utilization, Oslo, Norway.

Marita, P., Leena, L., & Tarja, K. (1999). Nurses' self-reflection via videotaping to improve communication skills in health counseling. Patient Education and Counseling, 36, 3-11.

McCroskey, J.C. (1970). Measure of communication-bound anxiety. Speech Monographs, 37, 269-277.

McGreal, T.L. (1988). Evaluating for enhancing instruction: Linking teacher evaluation and staff development. In S.J. Stanley & W.J. Popham (Eds.), Teacher evaluation: Six prescriptions for success (pp. 1-29). United States: ASCD.

McLaughlin, M.W. (1990). Embracing contraries: Implementing and sustaining teacher evaluation. In J. Millman & L. Darling-Hammond (Eds.), The new handbook of teacher

evaluation: Assessing elementary and secondary school teachers (pp. 403-415). Newbury Park, CA: Corwin Press.

McLaughlin, M.W., & Pfeifer, R.S. (1988). Teacher evaluation: Organizational change, accountability and improvement. In E.S. Hickcox, S.B. Lawton, K.A. Leithwood, & D.F. Musella (Eds.), Making a difference through performance appraisal (pp. 113-140). Toronto: OISE Press.

McMillan, J.H. (1997). Classroom assessment: Principles and practice for effective instruction. Toronto: Allyn and Bacon.

McQuarrie, F.O., & Wood, F.H. (1991). Supervision, staff development, and evaluation connections. Theory Into Practice, 30, 91-96.

Menzel, K.E., and Carrell, L.J. (1994). The relationship between preparation and performance in public speaking. Communication Education, 43, 17-26.

Mezirow, J. (1981). A critical theory of adult learning and education. Adult Education, 2, 3-24.

Molloy, J.T. (1971). Dress for success. New York: Warner.

Molloy, J.T. (1977). The women's dress for success book. Chicago: Follet Publishing.

Morreale, S.P. (1990, November). "The competent speaker": Development of a communication-competency based speech evaluation form and manual. Paper presented at the Annual Meeting of the Speech Communication Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 325901).

Morreale, S.P., Awtrey, C.C., Hurlbert-Johnson, R., Morley, D.D., Moore, M.R., Tatum, D.S., & Taylor, K.P. (1991, October). The competent speaker: A short course on assessing public

speaking competence. Paper presented at the Annual Meeting of the Speech Communication Association, Atlanta, GA. (ERIC Document Reproduction Service No. ED 344253).

Morreale, S.P., & Hackman, M.Z. (1994,). A communication competency approach to public speaking instruction. Journal of Instructional Psychology, 21, 250-257.

Natriello, G. (1984). Teachers' perceptions of the frequency of evaluation and assessments of their effort and effectiveness. American Educational Research Journal, 21, 579-595.

Newburger, C., Brannon, L., & Daniel, A. (1994). Self-confrontation and public speaking apprehension: To videotape or not videotape student speakers? (ERIC Document Reproduction Service No. ED 371419).

Ovando, M.N., & Harris, B.M. (1993). Teachers' perceptions of the post-observation conference: Implications for formative evaluation. Journal of Personnel Evaluation in Education, 7, 301-310.

Patton, M.Q. (1994). Developmental evaluation. Evaluation Practice, 15, 311-319.

Patton, M.Q. (1997). Utilization-focused evaluation: The new century text (3rd edition). Thousand Oaks, CA: Sage Publications.

Pavan, B. (1986). A thank you and some questions for Madeline Hunter. Educational Leadership, 43, 67-68.

Peters, J.M. (1991). Strategies for reflective practice. New Directions for Adult and Continuing Education, 51, 89-96.

Poole, W.L. (1994). Removing the 'super' from supervision. Journal of Curriculum and Supervision, 9, 90-94.

Popham, W.J. (1988). Judgment-based teacher evaluation. In S.J. Stanley & W.J. Popham (Eds.), Teacher evaluation: Six prescriptions for success (pp. 56-77). United States: ASCD.

Popham, W.J. (2000). Modern educational measurement: Practical guidelines for educational leaders (3rd edition). Toronto: Allyn and Bacon.

Posavac, E.J., & Carey, R.G. (1992). Program evaluation: Methods and case studies (4th edition). Englewood Cliffs, NJ: Prentice-Hall.

Quigley, B.L., & Nyquist, J.D. (1992). Using video technology to provide feedback to students in performance courses. Communication Education, 41, 324-334.

Ross, J.A., Cousins, J.B., & Gadalla, T. (1996). Within-teacher predictors of teacher efficacy. Teaching and Teacher Education, 12, 385-400.

Schay, B.W. (1993). In search of the holy grail: Lessons in performance management. Public Personnel Management, 22, 649-688.

Schneier, C.E., Beatty, R.W., & Baird, L.S. (1986). Creating a performance management system. Training and Development Journal, 40, 74-79.

Schön, D.A. (1983). The reflective practitioner: How professionals think in action. New York: Basic Books.

Schön, D.A. (1987). Educating the reflective practitioner: Toward a new design for teaching and learning in the professions. San Francisco: Jossey-Bass.

Scriven, M. (1987). Validity in personnel evaluation. Journal of Personnel Evaluation in Education, 1, 9-23.

Scriven, M. (1988). Evaluating teachers as professionals: The duties-based approach. In S.J. Stanley & W.J. Popham (Eds.), Teacher evaluation: Six prescriptions for success (pp. 110-142). United States: ASCD.

Sherer, M., & Adams, C.H. (1983). Construct validation of the self-efficacy scale. Psychological Reports, 53, 899-902.

Sherer, M., Maddux, J.E., Mercadante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R.W. (1982). The self-efficacy scale: Construction and validation. Psychological Reports, 51, 663-671.

Showers, B., & Joyce, B. (1996). The evolution of peer coaching. Educational Leadership, 53, 12-16.

Showers, B., Joyce, B., & Bennett, B. (1987). Synthesis of research on staff development: A framework for future study and a state-of-the-art analysis. Educational Leadership, 45, 77-87.

Smith, B. (1988). Using video for feedback. Journal of European Industrial Training, 12, 22-27.

Soodak, L.C., & Podell, D.M. (1996). Teacher efficacy: Toward the understanding of a multi-faceted construct. Teaching and Teacher Education, 12, 401-411.

Stake, R.E. (1976). To evaluate an arts program. In R.E. Stake (Ed.), Evaluating arts education: A responsive approach. Columbus, OH: Merrill.

Stephan, W.G., & Dorfman, P.W. (1989). Administrative and developmental functions in performance appraisals: Conflict or synergy? Basic and Applied Social Psychology, 10, 27-41.

Stiggins, R.J. (1988). The case for changing teacher evaluation to promote school improvement. In E.S. Hickcox, S.B. Lawton, K.A. Leithwood, & D.F. Musella (Eds.), Making a difference through performance appraisal (pp. 141-157). Toronto: OISE Press.

Stiggins, R.J. (2001). Student-involved classroom assessment (3rd edition). New Jersey: Prentice-Hall.

Stodolsky, S.S. (1990). Classroom observation. In J. Millman & L. Darling-Hammond (Eds.), The new handbook of teacher evaluation: Assessing elementary and secondary school teachers (pp. 175-190). Newbury Park, CA: Corwin Press.

Sullivan, R.L., & Wircenski, J.L. (1988). Clinical supervision: The role of the principal. NASSP Bulletin, 72, 34-37.

Tracy, S.J., & McNaughton, R.H. (1989). Clinical supervision and the emerging conflict between the neo-traditionalists and the neo-progressives. Journal of Curriculum and Supervision, 4, 246-256.

Tziner, A., & Latham, G.P. (1989). The effects of appraisal instrument, feedback and goal-setting on worker satisfaction and commitment. Journal of Organizational Behavior, 10, 145-153.

Vispoel, W.P., & Chen, P. (1990). Measuring self-efficacy: The state of the art. (ERIC Document Reproduction Service No. ED 338712).

Watt, W.M. (1995). Assessing student learning outcomes in teaching business and professional speaking. (ERIC Document Reproduction Service No. ED 394157).

Wildman, T.M. & Niles, J.A. (1987). Essentials of professional growth. Educational Leadership, 45, 4-10.

Willmington, S.C., Benson, J.A., Neal, K.E., & Steinbrecher, M.M. (1993). Meeting communication proficiency requirements for teacher certification through the basic course. (ERIC Document Reproduction Service No. ED 367009).

Wilson, R.J. (1996). Assessing students in classrooms and schools. Scarborough, ON: Allyn & Bacon Canada.

Woolfolk, A.E., & Hoy, W.K. (1990). Prospective teachers' sense of efficacy and beliefs about control. Journal of Educational Psychology, 82, 81-91.

Appendix A

Appendix A1

Letter of Informed Consent – Pilot Test

Principal Investigator: **John J. Donohue**
Affiliation: **Ph.D. Candidate, Faculty of Education, University of Ottawa**
Telephone No.: **(613) 592-3102**

Whenever a research project is undertaken with human participants, the written consent of the participants must be obtained. This does not imply, of course, that the project in question necessarily involves a risk. In view of the respect owed the participants, the University and the research funding agencies have made this type of agreement mandatory.

The purpose of this phase of the study is to determine whether the survey instrument to be used in subsequent phases adequately covers the domain of public speaking performance.

If I agree to participate, my participation will consist essentially of reading and critiquing the draft survey instrument on my own time. I understand that the contents will be used only for the purpose of refining the survey instrument, if necessary, and that my confidentiality will be respected.

I am free to withdraw from the project at any time, before or during an interview, refuse to participate and refuse to answer questions without penalty.

I have received assurances from the researchers that the information I will share will remain strictly **confidential**.

Any information requests or complaints about the ethical conduct of the project may be addressed to the Secretariat of the Ethics Committee (562-5800 x4057). **If I have any questions**, I may contact Professor Brad Cousins of the Faculty of Education, 562-5800 x4075. **There are two copies of the consent form, one of which I may keep.**

Participant's Signature	Date
Researcher's Signature	Date

I, _____, am interested in collaborating in the study **The Impact of Constructive Feedback on Perceived Self-efficacy in Growth-oriented Appraisal of Public Speakers** conducted by John J. Donohue of the Faculty of Education of the University of Ottawa.

Optional: I wish to receive a summary of the findings of this study which will be available in the Summer of 1999 (approximate date) at the following address:

Appendix A2

Demographic Data Section of the SSRS

---

1. Gender (circle one): **Female** **Male**

2. Age in years (circle one): **< 25** **25-34** **35-44** **45-54** **55 +**

3. Presenting before groups is part of my regular job (circle one): **Yes** **No**

4. The number of times I present to audiences per year is (circle one):
   < 10 **11 – 20** **21 - 30** **31 - 40** **41 – 50**

5. I have been presenting before audiences for _____ years (please insert number of years).

6. I am a member of (circle all that apply):
   Toastmasters International (TI) **Canadian Association of Professional Speakers (C.A.P.S.)**

Complete Section 7 OR 8 below (BOTH if member of both TI and C.A.P.S.)

7. a) Toastmasters Branch I attend most - Branch Number:

   b) Toastmasters designations achieved (check all that apply):

   _____ Competent Toastmaster
   _____ Advanced Toastmaster Bronze
   _____ Advanced Toastmaster Silver
   _____ Advanced Toastmaster Gold
   _____ Competent Leader
   Advanced Leader Bronze
   _____ Advanced Leader Silver
   _____ Advanced Leader Gold
   _____ Distinguished Toastmaster
   _____ Other (provide details):
   _____ Other (provide details):
   Other (provide details):

8. a) C.A.P.S. Chapter I attend the most (circle one):
   Vancouver **Calgary** **Hamilton** **Toronto**
   Ottawa **Montreal** **Halifax**

   b) C.A.P.S. designations received (circle all that apply):
   Certified Speaking Professional (C.S.P.) **Council of Peers Award of Excellence (C.P.A.E.)**

   c) I am a member of C.A.P.S. – National (circle one): **Professional Member** **Associate Member** **Not a Member**

Appendix A3

Cover Letter from the President of CAPS

Message from the President

Compiling industry information that can give you insights, benchmarks and comparatives is one of the things a national association can do for you. CAPS is pleased to support member John Joe Donohue in his doctoral research, and encourage everyone to do the same

Please take the few minutes required to fill out this survey. We're all looking forward to seeing the results in a future edition of our newsletter.

Warren Evans, CSP
1998/99 President - Canadian Association of Professional Speakers

Survey on Speaker Self-Efficacy

Dear Colleague,

As part of my Doctoral thesis at the University of Ottawa, I am asking for your voluntary participation in a survey concerning self-efficacy and public speaking performance. The purpose of this research project is to learn more about the nature of public speakers and their perceived self-efficacy vis-à-vis their performance.

Your name was obtained from the current membership list of the **Canadian Association of Professional Speakers (C.A.P.S.)**. Administrative officials of C.A.P.S. have given us permission to use the membership list for this purpose. You may be interested to know that members of other public speaking organizations in Canada are also participating in the survey.

If you agree to participate, please take 20 minutes to fill out the attached questionnaire and return it in the self-addressed envelope (stamp provided) within two weeks of receipt. There is no need to put your name on the questionnaire.

If you have inquiries about the project or wish to receive a two-page summary of the results please contact me at:

John J. Donohue, Ph.D. Candidate
Faculty of Education, University of Ottawa
145 Jean-Jacques Lussier
Ottawa, ON, Canada, K1N 6N5

Phone: (613) 592-3102
Fax: (613) 591-1983
E-mail: Donohue_John@msn.com

Your participation is greatly appreciated.

John J Donohue

Appendix A4

Cover Letter for TI Participants

Survey on Speaker Confidence

Dear Colleague,

As part of my Doctoral thesis at the University of Ottawa, I am asking for your voluntary participation in a survey concerning confidence and public speaking performance. The purpose of this research project is to learn more about the nature of public speakers and their perceived confidence vis-à-vis their performance.

You may be interested to know that members of other public speaking organizations in Canada are also participating in the survey.

If you agree to participate, please take 20 minutes to fill out the attached questionnaire and return it in the self-addressed envelope (stamp provided) within two weeks of receipt. There is no need to put your name on the questionnaire.

If you have inquiries about the project or wish to receive a two-page summary of the results when they become available, please contact me at:

John J. Donohue, Ph.D. Candidate
Faculty of Education, University of Ottawa
145 Jean-Jacques Lussier
Ottawa, ON, Canada, K1N 6N5

Phone: (613) 592-3102
Fax: (613) 591-1983
E-mail: leaders@magi.com

Thesis Supervisor:

Professor Brad Cousins
Faculty of Education, University of Ottawa
145 Jean-Jacques Lussier
Ottawa, ON, Canada, K1N 6N5

Phone: (613) 562-5800 x 4075
E-mail: bcousins@uottawa.ca

Your participation is greatly appreciated.

John J Donohue

Appendix A5

Combined GSE and SSE Subscales of SSRS

Part A: General Views about Self

Please indicate the extent to which each of the following statements reflects your personal behavior by circling ONE of the response codes in the right hand column. The response codes are defined as follows:

N = Never **R** = Rarely **S** = Sometimes **F** = Frequently **A** = Always **NA** = not applicable
Try to use NA as infrequently as possible.

1.	When I decide to do something I go right to work on it.	N	R	S	F	A	NA
2.	If I can't do a job the first time I keep trying until I can.	N	R	S	F	A	NA
3.	I feel insecure about my ability to do things.	N	R	S	F	A	NA
4.	When I'm trying to become friends with someone who seems uninterested at first, I don't give up easily.	N	R	S	F	A	NA
5.	If I meet someone interesting who is hard to make friends with I'll soon stop trying to make friends with that person.	N	R	S	F	A	NA
6.	I do not handle myself well in social gatherings.	N	R	S	F	A	NA
7.	When I make plans I am certain I can make them work.	N	R	S	F	A	NA
8.	When trying to learn something new I soon give up if I am not initially successful.	N	R	S	F	A	NA
9.	When I set important goals for myself I rarely achieve them.	N	R	S	F	A	NA
10.	I avoid facing difficulties.	N	R	S	F	A	NA
11.	One of my problems is that I cannot get down to work when I should.	N	R	S	F	A	NA
12.	I give up on things before completing them.	N	R	S	F	A	NA
13.	If something looks too complicated I will not even bother to try it.	N	R	S	F	A	NA
14.	Failure just makes me try harder.	N	R	S	F	A	NA
15.	When unexpected problems occur I don't handle them well.	N	R	S	F	A	NA
16.	When I have something unpleasant to do I stick to it until I finish it.	N	R	S	F	A	NA
17.	If I see someone I would like to meet I go to that person instead of waiting for him or her to come to me.	N	R	S	F	A	NA
18.	I have acquired my friends through my personal abilities at making friends.	N	R	S	F	A	NA
19.	I do not seem capable of dealing with most problems that come up in life.	N	R	S	F	A	NA
20.	It is difficult for me to make new friends.	N	R	S	F	A	NA
21.	I avoid trying to learn new things when they look too difficult for me.	N	R	S	F	A	NA
22.	I give up easily.	N	R	S	F	A	NA
23.	I am a self-reliant person.	N	R	S	F	A	NA

Appendix A6

PSE Subscale of SSRS

Part B: Views about Performance as a Speaker

Please indicate the extent to which each of the following statements reflects your personal behavior by circling ONE of the response codes in the right hand column.

The response codes are defined as follows:

N = Never R = Rarely S = Sometimes F = Frequently A = Always NA = not applicable
Try to use NA as infrequently as possible.

1.	I increase volume and slow down on key words to emphasize their meaning in a sentence.	N	R	S	F	A	NA
2.	I have difficulty articulating words.	N	R	S	F	A	NA
3.	I have difficulty maintaining eye contact.	N	R	S	F	A	NA
4.	I use dynamic, varied and natural facial expressions.	N	R	S	F	A	NA
5.	I vary my vocal pitch to suit the material and maintain audience interest.	N	R	S	F	A	NA
6.	I have correct posture when giving a presentation.	N	R	S	F	A	NA
7.	I have difficulty using open and varied gestures when giving a presentation.	N	R	S	F	A	NA
8.	I use the entire stage or podium during my presentations.	N	R	S	F	A	NA
9.	I make very few, if any, pronunciation errors.	N	R	S	F	A	NA
10.	Some audience members have difficulty hearing me..	N	R	S	F	A	NA
11.	I move smoothly from idea to idea within my presentation.	N	R	S	F	A	NA
12.	I vary my rate of speech to suit the material and maintain audience interest..	N	R	S	F	A	NA
13.	I maintain a resonating, dynamic vocal quality, fully supported by breath during my presentations.	N	R	S	F	A	NA
14.	I speak from behind a lectern.	N	R	S	F	A	NA
15.	I find it difficult to avoid using filler words.	N	R	S	F	A	NA
16.	I have difficulty pausing during my presentations.	N	R	S	F	A	NA

NOTE: These questions are NOT meant to reflect ALL aspects of speaker behavior, nor are they necessarily the most important behaviors. They are meant to address the types of behavior of interest to this study.

Appendix A7

Demographic Data Section of SSRS

Part C: Background Information

Please circle or write in the appropriate response:

1. Gender (circle one): **Female** **Male**

2. Age in years (circle one): **< 25** **25-34** **35-44** **45-54** **55 +**

3. Presenting before groups is part of my regular job (circle one): **Yes** **No**

4. The number of times I present to audiences per year is (circle one):

< 10 **11 – 20** **21 - 30** **31 - 40** **41 – 50** **51 - 100** **100+**

5. I have been presenting before audiences for _____ years (please insert number of years).

6. I am a member of (circle all that apply):

Toastmasters International (TI) **Canadian Association of Professional Speakers (C.A.P.S.)**

Complete Section 7 OR 8 below (BOTH if member of both TI and C.A.P.S.)

7. a) Toastmasters Branch I attend most - Branch Number: _____

b) Toastmasters designations achieved (check all that apply):

_____ Competent Toastmaster
_____ Advanced Toastmaster Bronze
_____ Advanced Toastmaster Silver
_____ Advanced Toastmaster Gold
_____ Competent Leader
_____ Advanced Leader Bronze
_____ Advanced Leader Silver
_____ Advanced Leader Gold
_____ Distinguished Toastmaster
_____ Other (provide details):
_____ Other (provide details):
_____ Other (provide details):

8. a) C.A.P.S. Chapter I attend the most (circle one):

Vancouver **Calgary** **Hamilton** **Toronto**

Ottawa **Montreal** **Halifax**

b) C.A.P.S. designations received (circle all that apply):

Certified Speaking Professional (C.S.P.) **Council of Peers Award of Excellence (C.P.A.E.)**

c) I am a member of C.A.P.S. – National (circle one): **Professional Member** **Associate Member** **Not a Member**

THE END

Thank you for taking the time to complete this survey.
Please return the completed survey in the pre-addressed stamped envelope supplied.

Appendix A8

GSE Subscale Summary Statistics

		CAPS			TI		
		x	SD	N	x	SD	N
1.	When I make plans I am certain I can make them work.	4.00	0.633	51	3.83	0.702	321
2.	One of my problems is that I cannot get down to work when I should. **(R)**	3.96	0.692	51	3.97	0.693	321
3.	If I can't do a job the first time I keep trying until I can.	3.76	0.737	51	3.40	0.817	320
4.	When I set important goals for myself I rarely achieve them. **(R)**	2.96	0.781	50	2.96	0.915	311
5.	I give up on things before completing them. **(R)**	2.82	0.850	50	2.80	0.938	318
6.	I avoid facing difficulties. **(R)**	3.98	0.990	51	3.65	0.902	319
7.	If something looks too complicated I will not even bother to try it. **(R)**	4.34	0.658	50	4.10	0.659	321
8.	When I have something unpleasant to do I stick to it until I finish it.	3.78	0.673	51	3.82	0.753	320
9.	When I decide to do something I go right to work on it.	3.98	0.616	51	3.81	0.788	313
10.	When trying to learn something new I soon give up if I am not initially successful. **(R)**	3.73	0.777	51	3.59	0.900	319
11.	When unexpected problems occur I don't handle them well. **(R)**	3.45	0.856	51	3.37	0.853	320
12.	I avoid trying things to learn new things when they look too difficult for me. **(R)**	3.81	0.633	51	3.79	0.721	321
13.	Failure just makes me try harder.	3.82	0.654	51	3.73	0.765	320
14.	I feel insecure about my ability to do things. **(R)**	3.92	0.796	51	3.61	0.841	321
15.	I am a self-reliant person.	3.94	0.645	51	3.67	0.758	319
16.	I give up easily. **(R)**	3.88	0.653	51	3.87	0.780	319
17.	I do not seem capable of dealing with most problems that come up in life. **(R)**	4.22	0.730	51	3.42	0.958	319

(R) – Item score reversed

Appendix A9

SSE Subscale Summary Statistics

	CAPS			TI		
	x	SD	N	x	SD	N
1. When I'm trying to become friends with someone who seems uninterested at first, I don't give up easily.	3.88	0.773	50	3.91	0.782	319
2. If I meet someone interesting who is hard to make friends with I'll soon stop trying to make friends with that person. (R)	4.14	0.775	51	3.77	0.909	320
3. I do not handle myself well in social gatherings. (R)	4.39	0.603	51	4.14	0.715	321
4. If I see someone I would like to meet I go to that person instead of waiting for him or her to come to me.	4.24	0.513	51	4.15	0.731	319
5. I have acquired my friends through my personal abilities at making friends.	4.63	0.528	51	4.40	0.753	320
6. It is difficult for me to make new friends. (R)	4.10	0.755	51	3.81	0.841	315

(R) – Item score reversed

Appendix A10

PSE Subscale Summary Statistics

		CAPS			TI		
		x	SD	N	x	SD	N
1.	I increase volume and slow down on key words to emphasize their meaning in a sentence.	4.04	0.564	51	3.63	0.813	316
2.	I have difficulty articulating words. **(R)**	4.06	0.759	51	3.68	0.826	323
3.	I have difficulty maintaining eye contact. **(R)**	4.47	0.644	51	4.01	0.810	321
4.	I use dynamic, varied and natural facial expressions.	4.43	0.608	51	3.82	0.887	323
5.	I vary my vocal pitch to suit the material and maintain audience interest.	4.53	0.542	51	3.90	0.824	322
6.	I have correct posture when giving a presentation.	4.14	0.756	50	4.09	0.745	321
7.	I have difficulty using open and varied gestures when giving a presentation. **(R)**	4.41	0.698	51	3.68	0.916	320
8.	I use the entire stage or podium during my presentations.	4.38	0.780	50	3.26	0.987	319
9.	I make very few, if any, pronunciation errors.	3.14	1.080	49	3.22	0.982	319
10.	Some audience members have difficulty hearing me. **(R)**	4.39	0.493	51	4.15	0.796	322
11.	I move smoothly from idea to idea within my presentations.	4.25	0.523	51	3.88	0.657	321
12.	I vary my rate of speech to suit the material and maintain audience interest.	4.41	0.638	51	3.80	0.827	322
13.	I maintain a resonating, dynamic vocal quality, fully supported by breath during my presentations.	4.16	0.766	50	3.62	0.899	319
14.	I speak from behind a lectern. **(R)**	4.45	0.673	51	2.76	0.940	321
15.	I find it difficult to avoid using filler words. **(R)**	4.10	0.755	51	3.43	0.878	321
16.	I have difficulty pausing during my presentations. **(R)**	4.02	0.787	51	3.62	0.842	322

(R) – Item score reversed

Appendix A11

Demographic Data Section Summary Statistics

		CAPS N	CAPS %	TI N	TI %
1. Gender (circle one):	Female	22	46.8	160	49.8
	Male	25	53.2	161	50.2
2. Age in years (circle one):	<25	0	0	7	2.2
	25-34	4	7.8	52	16.1
	35-44	20	39.2	89	27.6
	45-54	19	37.3	109	33.7
	55+	8	15.7	66	20.4
3. Presenting before groups is part of my regular job (circle one):	Yes	45	88.2	127	39.8
	No	6	11.8	192	60.2
4. The number of times I present to audiences per year is (circle one):	<10	0	0	138	43.3
	11-20	2	3.9	104	32.6
	21-30	8	15.7	27	8.5
	31-40	7	13.7	15	4.7
	41-50	4	7.8	16	5.0
	51-100	19	37.3	13	4.1
	100+	11	21.6	6	1.9

	CAPS x	CAPS SD	CAPS N	TI x	TI SD	TI N
5. I have been presenting before audiences for ## years (please insert number of years).	12.9	9.47	51	10.1	8.88	305

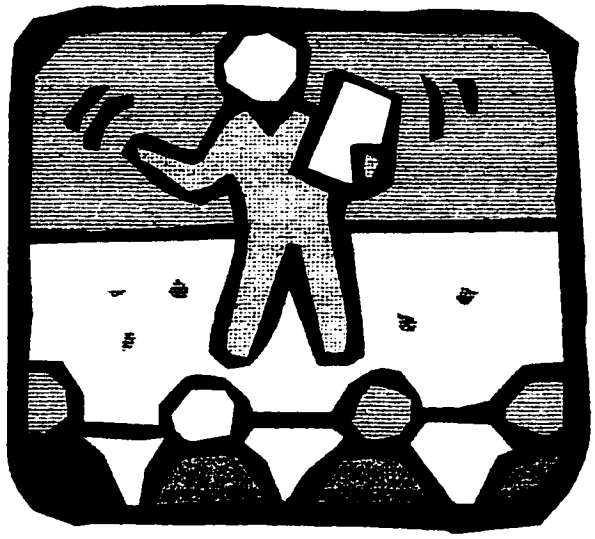
			CAPS		TI	
			N	%	N	%
6.	I am a member of (circle all that apply):	TI			318	99.4
		CAPS	40	78.4		
		Both	11	21.6	2	0.6
7.	a) Toastmasters Branch I attend most – Branch Number:	n/a	n/a		n/a	
	b) Toastmasters designations achieved (check all that apply):	Competent TM	9	17.6	169	52.3
		Advanced TM Bronze	3	5.9	47	14.6
		Advanced TM Silver	1	2.0	18	5.6
		Advanced TM Gold	1	2.0	7	2.2
		Competent Leader	1	2.0	25	7.7
		Advanced Leader Bronze	0	0	1	0.3
		Advanced Leader Silver	0	0	1	0.3
		Advanced Leader Gold	0	0	0	0
		Distinguished TM	3	5.9	15	4.6

		CAPS		TI	
		N	%	N	%
8. a) CAPS chapter I attend the most (circle one):	Vancouver	8	18.2		
	Calgary	7	15.9		
	Hamilton	2	4.5		
	Toronto	20	45.5		
	Ottawa	5	11.4	1	50
	Montreal	1	2.3		
	Halifax	1	2.3	1	50
b) CAPS designations achieved (circle all that apply):	Certified Speaking Professional (CSP)	5	10.2		
	Council of Peers Award of Excellence (CPAE)	1	2.0		
c) I am a member of CAPS – National (circle one):	Professional Member	44	88		
	Associate Member	6	12	1	100
	Not a Member	0			

Appendix B

Appendix B1

Advertisement Brochure – Front Cover



Interested in Improving
Your Presentation Skills
... for FREE?

Join Toastmasters from across the Ottawa-Carleton region.

A *University of Ottawa* Research Study

Appendix B2

Advertisement Brochure – Inside Left

Program Focus:

- This is a University of Ottawa research project focusing on **improving your presentation skills**. Content issues will NOT be addressed in this program.

Eligibility:

- **Participants must be members of an Ottawa-area Toastmasters Club.**

You Will:

- Present a short speech six times over the course of the 6 sessions before 7 other study participants.
- Receive private feedback with a peer on your presentations using videotape of your performance.
- Learn characteristics of constructive feedback, and provide feedback on presentations.
- Participate in a short post-program interview on your experience in the program.

You Need to Bring:

- A 4 to 7 minute speech ON ANY TOPIC for use in the program - be *ready to present at the Oct. 30th meeting*.
- A **strong desire** to improve your presentation skills.
    - A commitment to complete the program by attending all 6 meetings in October and November (program schedule on last page).

Appendix B3

Advertisement Brochure – Inside Right

Your Benefits:

- **Improving your presentation skills.**
- Counting your speech toward your CTM, ATM or DTM designation (optional).
- Receiving a video-tape collection of all speeches that you make in the program. You will literally see your own growth!
- Receiving a video compilation of the best speeches (peer voted) by program participants.
- Hearing **Jack Donohue** (Olympic basketball coach, professional speaker and television personality) speak live on "*The Use of Humour in Public Presentations*" at the final session - included on your video collection.
- Opportunity to submit your best performance for a personal written critique by Jack Donohue (for a nominal fee)

About Your Program Coordinator:

John J. Donohue is a Ph.D. Candidate at the Faculty of Education of the University of Ottawa. He specializes in performance evaluation and improvement, specifically in the area of public speaking. He also serves as the Development Chair of the Ottawa Chapter of the Canadian Association of Professional Speakers.

Important: You MUST participate in all 6 sessions in order to receive the video-tape of your performances and to maintain the integrity of the program.

Appendix B4

Advertisement Brochure – Back Cover

Program Schedule

Participants will meet **6 times** during the program:

Initial Meeting: Saturday, October 30, 1999

Four (4) weekly meetings
Indicate your preference (1st and 2nd choice)

_____ Monday Evenings (7:30 – 9:00 p.m.)
(Nov. 1, 8, 15, 22)

_____ Tuesday Evenings (7:30 – 9:00 p.m.)
(Nov. 2, 9, 16, 23)

_____ Wednesday Evenings (7:30 – 9:00 p.m.)
(Nov. 3, 10, 17, 24)

_____ Thursday Evenings (7:30 – 9:00 p.m.)
(Nov. 4, 11, 18, 25)

Final Meeting: Saturday, November 27, 1999

Meeting Locations:

All meetings will take place on the University of Ottawa campus. Details to be determined.

To Register, Contact:

John J. Donohue, Ph.D. Candidate
Faculty of Education, University of Ottawa
Phone: (613) 271-9522
e-mail: leaders@magi.com
http://infoweb.magi.com/~leaders/present.html

Registration Deadline:

Friday, Oct. 22, 1999 - 4:00 p.m.

Appendix B5

Letter of Informed Consent – Main Study

Principal Investigator: **John J. Donohue**
Affiliation: **Ph.D. Candidate, Faculty of Education, University of Ottawa**
Telephone No.: **(613) 592-3102**

The purpose of the study is to investigate the impact of constructive feedback on public speakers skills and beliefs. Whenever a research project is undertaken with human participants, the written consent of the participants must be obtained. Your signature at the bottom of this form will provide such consent. Please read the following carefully before signing.

If I agree to participate, my participation will consist of attending an in-service training session, four data collection sessions, a final meeting, and a post-interview. At each data collection session I will perform a five-minute speech which will be videotaped, and then I will participate in a feedback session with a peer, which will be audio-taped. I will complete pre- and post-test questionnaires, and I will also complete various rating sheets at each data collection session. I will participate in a post-interview, during which I will be asked to reflect on my experience in the study for the purpose of helping to interpret the findings.

I understand that the researchers will make every effort to avoid or minimize any events that may lead to negative emotional reactions.

I am free to withdraw from the project at any time, and may refuse to participate or to answer questions.

I have received assurances from the researchers that the information I will share will remain **strictly confidential.** I, in turn, assure other participants that I will treat in the same confidential manner any information I may obtain in the context of this project.

Any information requests or complaints about the ethical conduct of the project may be addressed to the Secretariat of the Ethics Committee (562-5800 x4057). **If I have any questions or concerns** about the project, I may contact Professor Brad Cousins of the Faculty of Education, 562-5800 x4075. **There are two copies of the consent form, one of which I may keep.**

_____ _____

Participant's Signature Date

_____ _____

Researcher's Signature Date

I, _____, am interested in collaborating in the study **The Impact of Constructive Feedback on Perceived Self-efficacy in Growth-oriented Appraisal of Public Speakers** conducted by John J. Donohue of the Faculty of Education of the University of Ottawa.

Optional: I wish to receive a summary of the findings of this study which will be available in the Summer of 2000 (approximate date) at the following address:

Appendix B6

Specific Area of Interest Scale

Name:

Please circle one (1) area below where you would most like to improve your performance through participation in the program. This should represent an area that you identify as needing improvement. If the area you wish to work on is not identified below, feel free to write in a description of your area of interest.

General Area:	Specific Area:
Voice	Inflection / Pitch Control Breathing / Support Volume Articulation Pacing / rate of speech Other: _____
Word Usage	Pronunciation Filler Words (err, ahh) Pausing Other: _____
Movement	About the presentation area Posture Gestures Facial Expressions Lectern / Podium Usage Other: _____
Interaction with Audience	Eye Contact Other: _____
Other Areas	Other: _____

On a scale of 1 to 10, please indicate your current level of ability in your chosen area of interest:

1 2 3 4 5 6 7 8 9 10

Low High

Appendix B7

Feedback Expectation Sheet (FES)

Feedback Session Protocol

1. Remember, we are focusing on positive, constructive feedback.
2. Turn on the audiocassette recorder when you are ready to begin.
3. Review the videotape of the performance together. Take notes if you like.
4. Rewind the videotape and begin the feedback session.
5. As the leader of this feedback session, you are responsible for:
   - Controlling the video playback
   - Leading the discussion
   - Supplying 2 positives (behaviors the performer did well)
   - Supplying 2 improvement areas (behaviors the performer can improve)
6. Remember the Elements of Effective Feedback – refer to the sheet if needed – the researchers have extra copies if you need one.
7. After the feedback session is finished, turn off the audio cassette recorder.
8. Each partner separately completes the Feedback Session Evaluation Sheet.

Switch roles and follow the steps above
for the second feedback session.

Appendix B8

Elements of Effective Constructive Feedback Sheet (EECFS)

Positive:	Use positive statements to make your point. For example, instead of saying "It doesn't look good when you ..." try saying, "I like the way you do this; to make it even better you could try ..."
Specific:	The more specific the feedback is, the better. Be as detailed as possible.
Uses Examples from Performance:	Draw examples directly from the video. For example: "See how you ... here in the video? You may want to try ..."
Focus on the Future:	Give feedback on how to improve performance in the future. For example, "The next time you do this, you may want to consider trying ..."
# of Components Evaluated:	Give feedback on as many components of the performance as possible. Use the Areas of Interest Sheet as a guide.

Probing Questions:

- What do you think would make this presentation better?
- Why do you like this part? Why don't you like this part?
- Is there anything else that you noticed in these areas that you would like to comment on (Voice, Word Usage, Movement, Interaction with Audience)?
- What do you think are your strengths in this performance? Weaknesses?
- What would you like to improve?

When you look at these areas (Voice, Word Usage, Movement, etc.), what techniques have you seen other speaker's use that you could adapt?

Appendix B9

Feedback Session Evaluation Sheet (FSES)

Participant performing on video: _____

Participant submitting this evaluation: _____

1. How constructive was the feedback given in this feedback session?

1	2	3	4	5
Destructive		Mixture		Constructive

2. Who led the feedback session?

1	2	3	4	5
Performer		Mixed / Shared		Peer

3. What did the performer do well during the presentation?

The performer: _____

The performer: _____

The performer: _____

4. What specific areas for improvement were discussed? What suggestions were made for each improvement area?

Improvement Area #1: _____
Suggestions: _____
_____

Improvement Area #2: _____
Suggestions: _____
_____

Improvement Area #3: _____
Suggestions: _____
_____

Feel free to add any comments you may have on the other side of this sheet!

Appendix B10

Post-program Survey Questions

1. What did you learn from participating in the presentation skills program?
   Did you improve in your presentation skills?

2. What are your views on the feedback you received during the program?

3. What did you like / dislike about the feedback sessions?

4. What did you like / dislike about the program in general (location, format, etc.)?

5. How would you change the program (esp. the feedback sessions)?

6. Did you incorporate the suggestions made in the feedback sessions into your performances? If yes, how, and with what effect?

7. I have just about finished analyzing the results of the program, and I could use your help in explaining some of the findings.

   During the program, to what extent did you talk with other members of the program about your performance, OTHER THAN during the normal feedback sessions?

   For example, you may have talked about your performance walking out of the building, after the tape was turned off, etc.

   Please indicate the content of this feedback, too. For example, you may have been more evaluative during these "informal" sessions than you were on tape.

   Also, the number of times (of the 4 feedback sessions) you did this would be useful to know.

Appendix B11

Peer Performance Rating Scale (PPRS)

Please indicate the extent to which each of the following statements reflects your perceptions of the speaker's behavior by circling ONE of the response codes in the right hand column.

The response codes are defined as follows:

N = Never R = Rarely S = Sometimes F = Frequently A = Always NA = not applicable
Try to use NA as infrequently as possible.

1. The speaker increases volume and slows down on key words to emphasize their meaning in a sentence.	N	R	S	F	A	NA
2. The speaker has difficulty articulating words.	N	R	S	F	A	NA
3. The speaker has difficulty maintaining eye contact.	N	R	S	F	A	NA
4. The speaker uses dynamic, varied and natural facial expressions.	N	R	S	F	A	NA
5. The speaker varies vocal pitch to suit the material and maintain audience interest.	N	R	S	F	A	NA
6. The speaker has correct posture when giving a presentation.	N	R	S	F	A	NA
7. The speaker has difficulty using open and varied gestures when giving a presentation.	N	R	S	F	A	NA
8. The speaker uses the entire stage or podium during the presentation.	N	R	S	F	A	NA
9. The speaker makes very few, if any, pronunciation errors.	N	R	S	F	A	NA
10. Some audience members have difficulty hearing the speaker..	N	R	S	F	A	NA
11. The speaker moves smoothly from idea to idea within the presentation.	N	R	S	F	A	NA
12. The speaker varies the rate of speech to suit the material and maintain audience interest..	N	R	S	F	A	NA
13. The speaker maintains a resonating, dynamic vocal quality, fully supported by breath during the presentation.	N	R	S	F	A	NA
14. The speaker speaks from behind a lectern.	N	R	S	F	A	NA
15. The speaker finds it difficult to avoid using filler words.	N	R	S	F	A	NA
16. The speaker has difficulty pausing during the presentation.	N	R	S	F	A	NA

NOTE: These questions are NOT meant to reflect ALL aspects of speaker behavior, nor are they necessarily the most important behaviors. They are meant to address the types of behavior of interest in this study.