The research developed a methodology to code and quantify eye movement patterns as a function of a neuro-linguistic programming strategy elicitation process.

Secondarily the research established methodology that identified the degree of consistency among certain certified Neuro-Linguistic Programming Practitioners while they rated eye movements.

The study reviewed literature in the field, identified problems relevant to research in NLP, defined and described selected rater strategies, and utilized the findings of this study to identify areas which are in need of further research.

Fifteen Certified Neuro-Linguistic Programming Practitioners viewed a videotape of 10 Oregon State University students who had been videotaped from the eyes upward while spelling 10 words. Practitioners observed eye movements of the videotaped students and these were recorded on Individual Rater Charts, with up to seven possible eye movement notations recorded for each spelling task.
Subsequent analysis of these notations provided overall agreement figures among the raters by Eye Movement,¹ a grand mean which reflected agreement among raters over all eye movements and spelling tasks, and Z scores associated with each of the overall ratings by Eye Movement. Greatest agreement among raters was found for the first observed Eye Movement, where overall agreement exceeded the grand mean. Overall agreement for the second observed Eye Movement fell below the grand mean. Overall agreement for subsequent Eye Movements generally increased through Eye Movement 6 and then showed a slight decrease for Eye Movement 7. This increase in agreement for later Eye Movements was largely attributed to what is termed the "NA-effect," which can result in high agreement scores in cases where only one or a few raters note an eye movement while all other raters note no eye movement, and can result in inflated overall agreement scores when the proportion of such cases related to a particular Eye Movement is high in relation to total number of recorded observations associated with the Eye Movement.

¹ An important terminological distinction should be made. Whereas the term 'eye movements' (in all lower case letters) refers to the particular eye movements displayed by individuals performing particular spelling tasks, the term 'Eye Movements' (with first letters capitalized) refers to the ordered set (from 1 through 7) of schema wherein observed eye movements are recorded and ordered. Thus, 'Eye Movement 7,' for example, refers to a schema wherein are recorded particular eye movements observed by raters as being the seventh occurring within a particular spelling task; moreover, since the schema exists even if there is nothing to fill it out, it makes sense to speak about Eye Movement 7 even in cases where no seventh eye movements have been observed.
Anticipated Findings 1, 2 and 3 were not supported by this study, with overall rater agreement lower than predicted for Eye Movements 1 and 2 and higher than predicted for remaining Eye Movements. Anticipated Finding 4 was confirmed by the study, with total number of eye movements observed by all raters for a particular subject found to be inversely correlated with overall agreement among raters in rating that subject. Anticipated Finding 5 was neither supported nor unsupported by the study, since, due to restraints, an appropriate methodology for testing this prediction was not available.
A STUDY FOR DEVELOPMENT OF A METHODOLOGICAL PROCESS AND THE USE OF CERTIFIED NLP-PRACTITIONERS IN ASSESSING THE CONSISTENCY OF NLP-PROGRAMMERS RATING EYE MOVEMENTS

by

Michael B. Sun

A THESIS submitted to Oregon State University

in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Completed April 18, 1988
Commencement June, 1988
APPROVED:

Redacted for privacy

Professor of School of Education in charge of major

Redacted for privacy

Chairman of Department of Post Secondary Education

Redacted for privacy

Dean of the School of Education

Redacted for privacy

Dean of Graduate School

Date thesis is presented April 18, 1988

Typed by Harvey McCloud for Michael B. Sun
ACKNOWLEDGEMENTS

My Thesis Committee:

Tom Grigsby, Education, gave a great deal of assistance in the project. I would like to thank him for his contribution of ideas and hard work.

Charles Carpenter, Education, a good reader and a friend. Thank you.

Carvel Wood, Education. Thank you for your support all along the way.

Tom Murphy, Psychology. Thank you for your contribution and careful reading.

Tom McClintock, History, an excellent reader and organizer. Thank you for your kind and open support.

Suzie Marish for her effort and support. Thank you.

I am grateful to my friends who have been patient with this endeavor: George W. Knox, Harry S. Oxenhandler, Corby Hann, S. Pengelly, Forrest Neil, John DeMeules, D. Woodman, Bella Abzug, Zachary Roo (USA), and Ron Watson and John Steinberg (Sweden).

Finally, thank you to my sister Marlene, and all of my family, and a very special thank you to my patient wife, Susan Marie Sun.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. THE PROBLEM</td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>3</td>
</tr>
<tr>
<td>Importance of the Study</td>
<td>4</td>
</tr>
<tr>
<td>Background of the Problem</td>
<td>7</td>
</tr>
<tr>
<td>II. REVIEW OF THE LITERATURE</td>
<td>11</td>
</tr>
<tr>
<td>Introduction</td>
<td>11</td>
</tr>
<tr>
<td>NLP: A Background</td>
<td>11</td>
</tr>
<tr>
<td>History of NLP</td>
<td>12</td>
</tr>
<tr>
<td>Philosophical Foundation of NLP</td>
<td>14</td>
</tr>
<tr>
<td>The Basics of NLP</td>
<td>16</td>
</tr>
<tr>
<td>Applications of NLP</td>
<td>23</td>
</tr>
<tr>
<td>Psychotherapy</td>
<td>24</td>
</tr>
<tr>
<td>NLP and Education</td>
<td>28</td>
</tr>
<tr>
<td>Other Applications of NLP</td>
<td>31</td>
</tr>
<tr>
<td>Related Research</td>
<td>32</td>
</tr>
<tr>
<td>III. METHODOLOGY</td>
<td>67</td>
</tr>
<tr>
<td>The Proposed Research Model</td>
<td>67</td>
</tr>
<tr>
<td>Selection of Video Tape Subjects</td>
<td>68</td>
</tr>
<tr>
<td>Instructions to Video Tape Subjects</td>
<td>69</td>
</tr>
<tr>
<td>Videotaping Procedure</td>
<td>70</td>
</tr>
<tr>
<td>Eye Movement Indicators</td>
<td>71</td>
</tr>
<tr>
<td>The Sample</td>
<td>71</td>
</tr>
<tr>
<td>Rating Procedure</td>
<td>72</td>
</tr>
<tr>
<td>Analysis of Data</td>
<td>75</td>
</tr>
<tr>
<td>Step 1 - Determination of Agreement by Subject, Word, and Eye Movement</td>
<td>77</td>
</tr>
<tr>
<td>Step 2 - Determination of Agreement by Subject and Eye Movement, Across Words</td>
<td>80</td>
</tr>
<tr>
<td>Step 3 - Determination of Agreement by Eye Movement, across Subjects and Words</td>
<td>82</td>
</tr>
<tr>
<td>Step 4 - Determination of a Grand Mean</td>
<td>82</td>
</tr>
<tr>
<td>Step 5 - Computation of a Z Score and Statistical Comparison</td>
<td>82</td>
</tr>
<tr>
<td>Anticipated Findings</td>
<td>84</td>
</tr>
<tr>
<td>IV. RESULTS</td>
<td>86</td>
</tr>
<tr>
<td>Data Collection</td>
<td>86</td>
</tr>
<tr>
<td>Videotaping of Students</td>
<td>86</td>
</tr>
<tr>
<td>Rating Procedure</td>
<td>88</td>
</tr>
<tr>
<td>Sample of Raters</td>
<td>88</td>
</tr>
<tr>
<td>The Procedure</td>
<td>89</td>
</tr>
<tr>
<td>Analysis of Data</td>
<td>91</td>
</tr>
<tr>
<td>Computation of Z Scores</td>
<td>91</td>
</tr>
</tbody>
</table>
Chapter | Page
--- | ---
Step 1 - Determination of Agreement by Subject, Word, and Eye Movement | 91
Step 2 - Determination of Agreement by Subject and Eye Movement, across Words | 92
Step 3 - Determination of Agreement by Eye Movement, across Subjects and Words | 93
Step 4 - Determination of a Grand Mean | 93
Step 5 - Computation of a Z Score | 95
Anticipated Findings | 98
Anticipated Findings 1, 2 and 3 | 98
Anticipated Findings 4 and 5 | 101
Summary of Results | 106
V. SUMMARY, DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS | 108
Summary | 108
Background | 108
Methodology | 110
Results | 112
Discussion | 113
Videotaping Procedure | 114
Rating Procedure | 115
Scoring Procedure | 121
Use of Certified NLP Practitioners | 125
Conclusions and Recommendations | 127
Conclusions | 127
Recommendations | 128
BIBLIOGRAPHY | 132
APPENDICES | 138
A. Letter to Practitioners, Directions for Certified Raters, Instruction Sheet to Raters | 138
B. Informed Consent - Video Tape Subjects | 142
C. Rater Background Data | 144
D. Equipment | 147
Glossary | 149
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Individual Rating Chart</td>
<td>73</td>
</tr>
<tr>
<td>2. All Raters Chart</td>
<td>78</td>
</tr>
<tr>
<td>3. Agreement by Subjects Chart</td>
<td>81</td>
</tr>
<tr>
<td>4. Overall Agreement Chart</td>
<td>83</td>
</tr>
<tr>
<td>5. Flow Chart of Analysis Process</td>
<td>85</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Table Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Measures of Agreement</td>
<td>94</td>
</tr>
<tr>
<td>2.</td>
<td>Z Scores for Seven Eye Movements</td>
<td>97</td>
</tr>
<tr>
<td>3.</td>
<td>Relation of Overall Agreement by Subject to Number of Recorded Observations per Subject</td>
<td>102</td>
</tr>
<tr>
<td>4.</td>
<td>Relation of Overall Agreement by Subject to Subject Order of Viewing</td>
<td>104</td>
</tr>
</tbody>
</table>
A STUDY FOR DEVELOPMENT OF A METHODOLOGICAL PROCESS AND THE USE OF CERTIFIED NLP-PRACTITIONERS IN ASSESSING THE CONSISTENCY OF NLP-PROGRAMMERS RATING EYE MOVEMENTS

CHAPTER 1

THE PROBLEM

Introduction

Neurolinguistic programming shows us that the complexities of human behavior, like the infinite number of possible well-formed sentences in a language, can be reduced to a finite number of structural elements and a syntax. In the context of the NLP model we maintain that all behavior—from learning, remembering and motivation to making a choice, communication and change—is the result of systematically ordered sequences of sensory representations. Many of the problems and phenomena that have baffled behavioral scientists in the past can be understood, predicted and changed by using the NLP model (Dilts et al., 1980).

This statement presents a generally accepted definition for the relatively new concept of Neuro-Linguistic Programming ("NLP"—a trademark). NLP proponents believe that individuals reveal exactly how they receive, store, process, and retrieve information. Further, they suggest that these processes, which they describe as baffling to behavioral scientists in the past, can be understood, predicted and, where necessary, modified utilizing the NLP model.

The introduction and acceptance of any new idea is often hampered by the lack of research that provides an adequate foundation for the review and analysis of that idea. In the case of NLP, research that would provide support for the relatively broad claims made by proponents is quite limited. Moreover, Robert Dilts (1985)
and Todd Epstein (1985) have questioned the breadth of previous research and claim that current research appears to be misguided in relation to understanding the basics of NLP. Dilts (1985b) suggests that the following areas, or components, need to be examined to provide an adequate research base for NLP:

1) Identification of Accessing Cues
   The behavioral cues that an individual employs to tune his or her neurology to single out cognitive activity within a particular representational system.

2) Strategy Rating
   The identification and assessment of the accessing cues [including eye movements] in the specific sequence of an individual's representational systems that make up a cognitive strategy.

3) Establishing Rapport
   The adjustment of one's language and behavior patterns to achieve conscious and unconscious rapport and the establishment of specific knowledge that indicates that one has rapport.

4) Strategy Elicitation
   The procedure for gathering the necessary information to make explicit the ordered sequence of representational system activity that constitutes a strategy.

5) Anchoring
   The methods of establishing non-verbal cues used to trigger behavioral and cognitive responses.

6) Strategy Installation
   The two basic processes to install a strategy sequence: (1) inserting the steps through a sequence of anchors; (2) rehearsing the strategy sequence (a form of self anchoring).

7) Specific Strategy Applications
   Specific strategy sequences that have been elicited from expert models for the purpose of being installed in novice or inadequate performers; i.e., spelling strategy, math strategy, composition strategy, etc.
8) Sub-Modalities

Properties or qualities of a particular representational system such as color, shape, brightness, volume, etc. Changing or altering these qualities can make a significant impact on subjective experience.

9) NLP's Change Models

All of the NLP processes that bring about a desired change such as: Change Personal History, Reframing, Anchoring Formats, V-K Dissociation, Meta Model, States of Excellence, Belief Systems, and any other specific techniques for producing behavioral change.

The research undertaken here was related most closely to areas 1 and 2 in Dilts' list of areas needing research. More specifically, this study concerned the rating or assessment of eye movements, which, according to Neuro-Linguistic Programming, constitute one of the cues indicating the ways in which individuals process information.

Statement of the Problem

The purpose of this study was to develop a methodology to code and quantify eye movement patterns as a function of an NLP strategy elicitation rating process.

The major objectives of this study were to:

1. review literature in the field;
2. identify problems relevant to research in NLP;
3. define and describe selected rater strategies as used in NLP by certified Practitioners;
4. develop a methodology to code and quantify eye movement patterns as a function of NLP strategy elicitation;
5. develop a methodology to accurately determine degree of consistency among certain NLP Practitioners in their ratings of eye movements;
6. utilize the findings of this study to identify areas which are in need of further research.

**Importance of the Study**

Proponents of the Neuro-Linguistic Programming model claim it to be a useful tool for psychotherapy and education. It is held to be valuable, in fact, at any point where human performance needs improvement. Dilts et al. (1980) maintain that the NLP model yields information quickly about how human beings do things, especially in relationship to "critical strategies such as motivation, creativity, belief, decision making, and remembering strategies." Once gained, this information can, according to NLP adherents, then be employed for the purpose of devising more effective strategies for accomplishing any of a wide variety of tasks. Clearly, if these claims are correct, important implications for psychotherapy, education and other fields follow.

But with respect to the usefulness of Neuro-Linguistic Programming, more is at issue than just the correctness of the model. A large organization has grown up around NLP. This organization, headed by John Grinder, one of the originators of Neuro-Linguistic Programming, undertakes the training of NLP Practitioners who, upon completion of their training, are certified by the organization as being capable of employing the model to
effect behavioral change among individuals. Thus, even if the model is correct, its usefulness is largely dependent upon the abilities of those individuals who have been certified by the NLP organization as being proficient in its application.

According to Grinder (1985), NLP has 80,000 Practitioners at work in many fields. These individuals have successfully completed a course of training, usually 20 days of eight hours a day, or a time sequence equal to this. Part of their training involves learning to rate behavioral cues which are said to indicate an individual's internal information processing systems.

One basic behavioral cue alleged to give such information is eye movements (Dilts, 1984). One objective of individuals' training to become Practitioners is to learn to recognize and rate eye movements. Upon the basis of these observations and ratings, inferences are made about how an observed individual processes information. In particular, inferences are made about the strategies (sequences of information processing) employed by the individual in performing various tasks. It is suggested that once the information processing strategy is understood, intervention can be designed to replace a less than optimum strategy with a new, more effective one. Preparing an individual to make such interventions is one major goal of the training of NLP Practitioners.

Clearly, the efficacy of such an intervention is based not only upon the adequacy of the NLP model but also on the skills of Practitioners in making correct inferences about internal strategies. The correctness of these inferences, in turn, rests upon the
ability of the Practitioner to observe and rate eye movements correctly, since these are one basic indicator of internal processing. Skill in observing and rating eye movements is, thus, a basic underpinning of successful intervention.

Yet, although eye movement rating skills of certified NLP Practitioners are a key component of the claim that NLP methods are effective in producing behavioral change, limited research has been conducted which attempts to determine the importance of eye movements as a component of information processing. Nor has any attempt been made to examine the consistency of certified NLP Practitioners as they rate eye movements. As a result, no procedure or protocol for research in this area has been adequately developed or tested. This research project was designed to provide a tested methodological framework to further research in this area.

In examining consistency among certified NLP Practitioners with respect to observing and rating eye movements, the present study thus attempted to formulate a clear and replicable methodology for examining eye movement rating skills and the consistency among raters in rating eye movements. As such, the investigation involved no attempt to assess the adequacy of the NLP model or concept itself. It involved, rather, the other basic underpinning identified above as being important to claims made by NLP proponents—the skills of that class of individuals who have been certified as able to usefully employ the NLP model.
Background of the Problem

Findings in NLP-related research have been inconclusive and have often contradicted one another. One possible reason for this is that some studies appear to exhibit a lack of clarity in understanding the place and importance of eye movements as cues to internal information processing. Indeed, there are serious disagreements and misunderstandings concerning the way that NLP eye movement rating works as well as the significance of eye movements as behavioral indicators (Dilts, 1985; Epstein, 1985).

Gumm et al. (1982), for example, assume that it is desirable, as a criterion for determination of a person's representational system, that three different types of indicators agree and occur at the same time: 1) eye movements, 2) the most frequently used predicate employed in the individual's speech--Visual (V), Auditory (A), or Kinesthetic (K)--and 3) the person's self report as he or she describes internal, subjective processing. Owens (1977) makes the same assumption, claiming that interaction between self report, observed eye movements, and the most frequently-used predicate needs to occur.

However, Dilts (1983a) claims that subject self report is unimportant and is, in fact, unreliable as an indicator of internal processing. Moreover, Lankton (1980) claims that the different modalities (self report, eye movement and predicate usage) often do not agree and that they do not need to in order to determine internal processing. Grinder and Bandler (1976) seem to concur with this position.
Thomason et al. (1980) failed to demonstrate a correct understanding of eye accessing movements according to Epstein (1985). Epstein states: "[The] experimenter cannot expect to have an entire question result in a single eye position. This expectation indicates a naivete on the part of the experimenter about how NLP actually works." He goes on to say that a "single question may elicit any number of eye movements."

Hernandez (1981) assumes that the beginning of a person's response to a stimulus is revealed in eye movements. However, Dilts (1984) holds that "physiology is more than just eye movements... Moving your shoulders is as much of an accessing cue as moving your eyes." Further, he seems to believe that some other physiological response may precede eye movements as an indicator of internal processing. Hernandez (1981) also singled out the eye movements occurring after her experimental statements had been made as being the relevant cues to internal processing. But in critiquing her methodology she says, "the eye movement that was scored may not have reflected the subject's internal response to the experimental statement... [and this could have] confounded the results."

Ellickson (1980), in studying whether the therapeutic relationship would be enhanced and the subject-client would report being understood if the interviewer-therapist responded with predicates matching the subject's internal representational system, reports a possible incongruency between a subject's "internal" and "external" representations based upon the possibility that "some subjects' eye movements may not have matched their own
Ellickson's study, it should be noted, utilized non-certified trainees who were given only two hours of training to become adept at reading eye movements and asking relevant questions.

It seems clear that within NLP-related research there are not only disagreements on theory, but also in methodology for research and practice. Such confusion indicates the need for more carefully developed methodologies. Though no NLP-related studies have been done to date which deal primarily with consistency among raters in judging behavioral cues, such a study would be a step toward a clearer paradigm for NLP-related research. The research presented here was such a study, involving the behavioral cues of eye movements and concentrating upon the inter-rater consistency among selected NLP Practitioners. The study contributes to the development of clearer methodologies in NLP-related research in three ways:

1. By developing a methodology for recording, rating and reviewing eye movements, the present study helps lay a foundation for future studies which concentrate upon eye movements.

2. By determining the extent to which a group of certified NLP Practitioners agree in their eye movement ratings, this study provides grounds for utilizing or not utilizing Practitioners as eye movement raters in future studies which call for the employment of such raters.
3. By providing a methodology for determining consistency among eye-movement raters, this investigation provides future researchers with a means for testing inter-rater consistency within any group which they employ or desire to employ as eye-movement raters.
CHAPTER II
REVIEW OF THE LITERATURE

Introduction

The Review of Literature is divided into two main sections. The first section is meant to familiarize the reader with the history and main tenets of NLP and includes a discussion of some areas where NLP is currently being applied. It is believed that such an overview is useful in the present study as it better enables the reader to understand the relevance of the research undertaken here. Ending the section is a brief discussion on the relation of representational systems and strategies to their alleged behavioral manifestations. This is of particular interest since the present research concerns behavior (eye movements) which is claimed to be such an indicator.

The second section is a review of recent research concerning NLP. Though a number of investigations have been carried out in this area during recent years, problems with some of these studies seem to indicate a need for a more fundamental, circumscribed approach to investigating alleged behavioral indicators of representational systems. The present research should be seen as an attempt at laying groundwork for such an approach.

NLP: A Background

In this section, a background to Neuro-Linguistic Programming is presented under the headings of (1) history of NLP,
(2) philosophical foundations of NLP, (3) the basics of NLP, and (4) applications of NLP.

History of NLP

Neuro-Linguistic Programming was founded in the mid 1970s by Richard Bandler, a psychotherapist, and John Grinder, a professor of Linguistics at the University of California, Santa Cruz. Bandler and Grinder had become familiar with the work of Fritz Perls, Milton Erickson, Virginia Satir and other psychotherapists, largely from viewing videotapes of the therapists at work. In attempting to understand what factors accounted for the unusual effectiveness of some of these practitioners, they began developing their view that knowledge of a person's internal (mental) processing can be gained by paying close attention to certain aspects of the person's language and to various nonverbal cues. Their observations of therapists at work led them to believe that effective therapists were often proficient at "reading" such cues. In gaining knowledge of a client's inner states, such a therapist was better able to achieve rapport with the client and to understand and speak to his/her needs. A more systematic investigation into the relations between behavioral cues and internal processing would, the two men claimed, provide new tools to improve therapist effectiveness.

At the same time, they began developing a new information processing model which they held could account for much of human behavior through its analysis of the internal processing that occurred immediately preceding behavior. This analysis was to be
done largely in terms of representations, both external and internal, and of systems of representations. Thus the behavioral cues which point to internal processing were seen as indicators of ways in which individuals represent the world. The sum of these views, including their application for the purpose of behavior modification, they named Neuro-Linguistic Programming, or NLP.

Bandler and Grinder set up a company to disseminate their views and to train others in their application. The application of the NLP model to psychotherapy was the focus of their first books: The Structure of Magic, Vol. I (1975), The Structure of Magic, Vol. II (1976), and (with Satir) Changing with Families (1976). However, already in The Structure of Magic, Vol. II, the potential applications of their model to education, business and other areas was discussed.

Soon a thriving organization had developed, with greater emphasis being placed upon non-therapeutic uses of NLP. The applicability of the NLP model to business, law, advertising, and in fact to any area involving communication was insisted upon, and training workshops were set up wherein a person could, upon successful completion of the training, be certified by the NLP organization as a Practitioner, or even a Master Practitioner. Special workshops for doctors, lawyers, and business executives were also begun. According to Grinder (1985) NLP had 80,000 Practitioners working in many different fields by 1985.

Bandler left the NLP organization in the early 1980's. Earlier, Robert Dilts had joined the organization and has since
become a leading spokesperson. Dilts, along with Bandler, Grinder and others, wrote NLP, Vol. I (1981). Dilts' Roots of NLP (1983a) and Applications of NLP (1983b), along with Dilts' and Myers-Anderson's Neurolinguistic Programming in Education (1980) are also important publications in the field. Other key works espousing the NLP model are Bandler's and Grinder's Frogs into Princes (1979); Grinder's and Bandler's (1976), Trance-formations (1981) and Reframing (1982); Bandler's Magic in Action (1984) and Using Your Brain for a Change (1985); Lankton's Practical Magic (1980); and Laborde's Influencing with Integrity: Management Skills for Communication and Negotiation (1984). Grinder's, DeLozier's, and Bandler's Patterns of the Hypnotic Techniques of Milton Erickson, M.D. (1975) is helpful in understanding the genesis of NLP.

Philosophical Foundation of NLP

Bandler's and Grinder's view rests largely on the concept of 'representational systems'—systems of internal representations of an external world. Philosophical roots for this notion have been traced by Ellis (1980) to Vaihinger and Korzybski. According to Ellis:

Vaihinger (1924) contends that logical processes (the world of ideas) should not be portrayed as reality, but but should be considered as an "instrument for finding our way about more easily in this world."

Korzybski (1933) made a similar distinction between "maps" (corresponding to cognitive processes) and "territories" (corresponding to what is represented in those processes). In fact, Ellis (1980),
in defining Bandler's and Grinder's notion of 'representational system,' makes use of Korzybski's distinction: "Stated simply, a representational system is a schema or cognitive map that an individual uses to represent his/her experience to him/herself or to others."

The relevance of the notion of representational system for human behavior becomes, at this fundamental level, clear. In fact, Bandler's and Grinder's view seems to imply that these representations are all we have to go on. Vaihinger (1924), as quoted by Mattar (1980), makes such a claim:

Representations are enormously important because they are the real world as far as can be known. Rather than acting on the world as it really is, we must act on the world according to how our representations construe it to be. Representations, then, are the instruments which are used to find our way about in the world.

Dilts (1978) echoes this view: "These representations are our maps or models which form the basis of our interactions with the world, i.e., we act according to our representations" [emphasis added].

Insofar as NLP is committed to such views, conceptual problems seem to arise for the NLP model. On the one hand, representational systems are sometimes conceived of as "instruments" or "maps" for finding our way in the world; that is, they seem to be thought of as mediators between ourselves and the world. On the other hand, a person's representational systems are sometimes treated as if they are, for that person, identical with the world (see, for example, Vaihinger's claim above: "They [representations] are the real world as far as can be known"). An apparent contradiction lies between these two ways of conceiving of representations and representational
systems. Bandler and Grinder do not seem to have dealt with this problem.

It is worth noting that important implications for psychology seem to follow from the above views. Whether subjective experience is taken to be the mediator between ourselves and the world, or is taken to be, in some sense, identical with the world (at least for that person), it arguably follows that to understand people, we must understand them from the standpoint of subjective experience. Thus the structure of subjective experience seems to become a proper goal for psychology. This is, of course, in opposition to traditional behaviorism, some of whose main proponents have claimed that a science of psychology must be based not upon the study of subjective experience, but of behavior (Skinner, 1953).

The Basics of NLP

A fundamental concept in NLP is that of the Test-Operate-Test-Exit sequence, or TOTE. Dilts et al. (1980) define the TOTE as:

a sequence of activities in our sensory representational systems that has become consolidated into a functional unit of behavior such that it is typically executed below the threshold of consciousness.

More specifically, TOTEs are sequences of activities wherein a present state of affairs is tested against a desired state of affairs (the Test phase) with the result that (1) if the two do not match, a further bit of behavior intended to bring about the match occurs (the Operate phase), followed by a further test, or (2) if the two do match, the behavior sequence ends (the Exit phase). For example,
using the illustration provided by Dilts et al. (1980), a person's hammering of a nail can be thought of in terms of a TOTE, wherein after each strike of the nail by the hammer, the actual position of the nail is tested for the desired state of flushness with the wood and as a result either the nail is hit again or else, if the nail is found to be flush with the wood, the nailing behavior ends.

This example also serves to illustrate the idea of nested TOTEs, where one or more TOTEs can be considered part(s) of a more comprehensive TOTE. Hammering a nail, for instance, might be nested within the TOTE of, say, flooring a house, which in turn might be part of the overall TOTE of building the house.

The TOTE as conceived to this point can be thought of as a schematic and as such is the mere form of a behavioral sequence. However, in actual behavioral sequences there is something filling out this form, namely external and/or internal representations. Remaining with the above illustration of hammering a nail, the Test phase could consist of a comparison of an external visual representation of the actual nail with an internal visual representation of a flush nail. If they are not congruent, the carpenter strikes the nail again. This alternation of hammering and testing continues until the external representation matches the internal, whereupon the sequence of behavior ends.

The congruency or incongruency which leads to either a further operation or an exit in a TOTE is also experienced as a representation according to Dilts et al. (1980). For instance, when the comparison of the representation of the actual nail matches the
internal representation of a flush nail, the carpenter in the above example experiences this as a new representation. This could occur, for instance, as a slight kinesthetic sensation in the diaphragm or some other bodily location.

There are other subtleties and complexities concerning the ways in which representations occur in TOTEs. For example, two internal representations can be compared in the Test phase of the TOTE, and a test can be between representations belonging to two different representational systems. However, for the present purposes these need not be elaborated.

What should be mentioned is that, in NLP, all elements of any TOTE are, in an important sense, on the same level. External representations, internal representations, and even the behavior which arises out of the TOTE are all dealt with as equal elements of one behavioral sequence. Dilts and Myers-Anderson (1980), in fact, use the term 'behavior' to cover all parts of the TOTE:

'Behavior' in neurolinguistic programming refers to activity within any representational system complex at any of these stages [input, processing, and output]. The acts of seeing, listening or feeling are behavior. So is 'thinking' which, if broken down to its constituent parts, would include sensory specific processes like seeing in the mind's eye, listening to internal dialogue, having feelings about something and so on. All output, of course, is behavior.

For NLP, the analysis of any particular TOTE into a sequence of representations reveals what is called a "strategy"—"the basic unit of analysis of a particular TOTE, or set of TOTES" (Dilts et al., 1980). The carpenter, for example, is involved in a strategy for hammering nails, with the strategy being just that sequence of representations described above.
One characteristic of strategies is that the same one can occur in a number of different situations. Consider the example of an athlete, say a quarterback, who hears an inspiring speech from his coach. As a result of this auditory stimulus, he gets a feeling of enthusiasm which is experienced as kinesthetic sensations in the diaphragm area and perhaps at other locations in his body, and at once internal images of himself passing the football with pinpoint accuracy occur. Finally, out on the field, the athlete transforms these images into energetic and skillful body movements. This could be a successful strategy for the athlete. He may employ the same basic strategy often in similar situations, and he may in the future seek out coaches who he believes will give inspiring speeches. In fact, if no such speech is forthcoming before a game, he may attempt to recall some previous speech, thereby setting the strategy in motion with an internal auditory stimulus.

However, another characteristic of strategies is the fact that they can be employed in inappropriate circumstances. If the above athlete were himself to become a coach of small children and then one day, upon hearing some supervisor's inspiring words (followed by kinesthetic and visual sensations), go out onto the field and proceed to frighten the children with his overzealousness, he would have taken a strategy useful in one context and employed it in another where it proved less successful.

A further basic notion in the NLP view is that of the "primary representational system." According to NLP, there are three basic representational systems—the visual, the auditory, and the
kinesthetic—and sensations which are one of these three types can be said to occur within the respective representational system. But many individuals favor one representational system over others so that in TOTE sequences the person tends to perform tests and internal operations in that particular representational system mode. The carpenter referred to above, for example, if visually oriented, would likely to look at the nail in testing for flushness, comparing what he observed with an internal visual representation; if kines- thetically oriented, he would more likely pay closest attention to the "feel" of the hammer hitting the nail, with the comparison being made to a stored kinesthetic sensation of what hitting a flush nail would feel like; if auditorily oriented, the carpenter would probably pay closest attention to the sound of the hammer hitting the nail and the corresponding comparison would be with a stored auditory sensation. Sometimes, according to the NLP view, all three kinds of tests and comparisons might be going on at the same time, yet a person with a most highly valued or primary representational system (PRS) will favor one of these over the others.

The possible applicability of NLP to behavioral change can be seen by considering the notion of the PRS, especially in reference to strategies. Dilts et al. (1980) use the example of spelling to point out this applicability. They claim that, for the task of spelling, those who characteristically visualize the word to be spelled are generally better spellers than those whose primary approach is to "sound out" the word. It might be expected, then, that spellers whose PRS is visual will be, on the whole, more
successful in spelling than those whose PRS is auditory. Indeed, this seems to be the view of Dilts et al., though they do not seem to hold that a person with, for example, a visual PRS will always use visual strategies. At any rate, they do claim that employment of visual strategies in spelling leads to greater success than employment of auditory strategies. Specifically, they hold that a "visual" speller, when presented a word to be spelled, will characteristically construct a visual image of the word which is then compared to a remembered visual image of the word (a case of two internal images being compared in the Test phase of the TOTE). A mismatch in this phase will lead to a new visual construction (Operate phase), and this Test-Operate alternation will continue until a satisfactory match with the remembered visual image is made, whereupon a characteristic internal kinesthetic feeling will occur. An "auditory" or "phonetic" speller, on the other hand, sounds out consecutive parts of the word either aloud or internally and compares these sounds to the remembered sound of the word originally presented (either by another or by himself to himself). The required match, if made, is achieved through this comparison of sounds.

The situation for each of these kinds of spellers may be more complicated than this. For example, the visual speller, in generating a new visual image in the Operate phase, may first repronounce the word either aloud or internally; the auditory speller may first be presented a visual stimulus, then pronounce the word to himself, and then go into the Test-Operate phase. Yet
the comparisons themselves are between visual images in the first case and between auditory elements (sensations or actual sounds) in the second, and according to Dilts et al. (1980), a large number of spellers fall into one of these two classes—those who employ a visual strategy and those who employ an auditory (phonetic) strategy. But, they claim, because English is a non-phonetic language, phonetic spellers are generally not as successful as those who use a visual strategy.

The primary significance of such analyses, according to NLP proponents, is that through breaking down strategies into their components and coming to understand which are the most successful strategies for varying tasks, it is possible to replace ineffective strategies with more effective ones. Thus, an important dynamic function of strategies is that they are generalizable: "The same sequencing of representational systems may be applied to a number of different types of behaviors" (Dilts, 1983a). As a consequence, a strategy used successfully by a person in one situation can, upon becoming known, be consciously applied by the individual in a new situation. Moreover, a successful strategy employed by one person or a group of people can in many cases be learned by others. According to Dilts (1983a), "applying a strategy to contexts outside of the one in which it was initially developed is called ... learning to learn."

Though the above provides only a brief introduction to the NLP model, it perhaps serves to highlight some of its chief claims. These are:
1. Overt human behavior can be understood as arising from strategies --Test-Operate-Test-Exit sequences of external and internal representations, with each representation being in either a visual, auditory, kinesthetic, olfactory, or gustatory mode, the first three being the most common of these.

2. Many people have a primary representational system, that is, they favor one of the three most common modes of representation over other modes (the so-called "visual," "auditory" or "kinesthetic" type person) and as a result tend to employ strategies whose Test and Operate phases are primarily composed of representations in the most favored mode.

3. Some strategies are better suited for certain tasks, and thus tend to lead to more successful behavioral outcomes in those tasks, than other strategies.

4. Strategies, once understood, can be taught.

Applications of NLP

Dilts (1983a) states what might be considered to be the "punch line" of NLP:

By making the representational form of test and operate procedures explicit, through strategies, the NLP model makes the transfer of any behavior that may be broken down into TOTE units more accessible and systematic. This is useful in the therapeutic context, because it makes patterns of behavior (whether they are problems or resources) more understandable and controllable. It is also useful in law, business, medicine, and education.
The following is a brief overview of recent applications of NLP to various fields, with special attention paid to applications in psychotherapy and education. Though other applications will be barely touched upon in this review, much recent activity in the NLP movement has been geared toward its use in fields such as advertising, law, and market research.

Psychotherapy. As pointed out in "History of NLP" above, Neuro-Linguistic Programming grew out of observations of unusually effective psychotherapists at work. Bandler and Grinder (1975) make reference to what they found to be common among the therapists they had observed:

The therapeutic 'wizards' . . . use techniques that appear to be dramatically different. . . . [But] they share one thing: They introduce changes in their clients' models which allow their clients more options in their behavior.

According to NLP, these models of reality are often geared to one representational system or another. As a result, an individual's apprehension of the world through other representational systems, and thus his behavioral repertoire, may be seriously limited. The therapist can help the clients develop a broader awareness through their senses and this can lead to more effective behavior. According to Grinder and Bandler (1976): "By adding an entirely new representational system, the client's model of the world is dramatically extended and many new choices become available to him."

Further, the therapist can use NLP methods to enhance communication with the client. The therapist does this by matching
his own language, especially predicates (verbs, adverbs, adjectives) to the client's preferred mode of representation. Grinder and Bandler (1976) illustrate this:

For example, when we are asking for information from a visual, we can phrase questions in the following ways:

How do you SEE the situation?
What do you SEE stopping you?

Or, ... with a kinesthetic, we will ask:

How do you FEEL about this situation?
What do you FEEL stops you?

Switching your predicates in this way will enable your clients to provide you with more information.

When the therapist does not take into account a client's having a different PRS from his own, the following kind of interchange can result according to Grinder and Bandler (1976):

Client (visual): My husband just doesn't see me as a valuable person.
Therapist (kinesthetic): How do you feel about that?
Client (visual): What?
Therapist (kinesthetic): How do you feel about your husband's not feeling that you're a person?
Client (visual): That's a hard question. I just don't know.

An illustration of more effective communication, taking into account the client's representational system, is given by Grinder and Bandler (1976) in reference to the same client:

Therapist: How do you know he doesn't see you as valuable?
Client: I dress up for him and he doesn't notice. (The client is assuming her husband also has a visual model of the world, as she does.)
Therapist: How do you know he doesn't notice?
Client: He just paws me and doesn't even look. (He responds kinesthetically and doesn't stand back far enough to see.)
Therapist: How do you feel as you see your husband not noticing you?
These examples also illustrate the more general NLP notion that language is often so closely tied to a person's PRS that communications between people with different PRSs can be at cross purposes. Grinder and Bandler (1976) claim that such communication gaps cause serious misunderstandings:

Communication between people under these conditions is usually haphazard and tedious. The result is often name-calling, ... Typically, kinesthetics complain that auditory and visual people are insensitive. visuals complain that auditories don't pay attention to them because they don't make eye contact during the conversation. Auditory people complain that kinesthetics don't listen, etc.

Davis and Davis, (1983) claim that the use of NLP in family therapy can help clients to communicate more effectively:

In therapy, it is common to hear sentences such as, 'From my perspective, the future looks very bright,' countered with sentences like, 'That just doesn't ring true; there's a total lack of harmony between us.'

Davis and Davis (1983) further hold that the aim of the therapeutic process in such a case is to help bridge the communication gap between the clients:

The methodology of bridging language differences involves translating to neutral (i.e., nonsensory) words, that both parties readily understand (words like thinking, being, doing, reassuring, happily, nicely, etc.), or repeating the same essential message in each of the two languages, e.g., letting her know that when he gives his word, his commitment and sense of reality is as absolute as when she sets her sights on an idea.

Therapists can also aid clients, according to the NLP view, through helping to alter a strategy which is ineffective for the task they are used for. This can be done through use of "anchoring," a kind of conditioning. Davis and Davis (1983) explain this:
Anchoring in therapy generally concerns internal states rather than external behaviors. . . . The neurological process involved occurs naturally all the time, which is why, e.g., the smell of new-mown grass or hay may immediately evoke a memory of some picture, feeling, some voices or sounds from the past. . . . An anchor is simply any discrete experience in any representational system which consistently evokes another, essentially unrelated, experience in any or several representational systems.

A new anchor is typically "installed" by a touch or some other bit of behavior which the therapist performs at the moment that a certain desired representation is experienced by the client. This bit of behavior is thereby meant to "anchor" the representation at a certain place within a series of representations. In this way ineffective strategies (which, of course, are only, according to NLP, series of representations) are altered. In fact, any undesired association between representations can be altered or erased in this way according to the NLP view. If, for example, the visual representation of oneself entering a room full of people is typically followed by unpleasant kinesthetic sensations, this conditioned association can be extinguished and replaced by a new association by installing a new, more pleasant sensation to follow the visual representation. The new sensation is the anchor.

This writer observed a powerful intervention, involving anchors, into a couple's repetitious arguing styles. These patterns quickly became obvious in language repetitions and hand gestures. The therapist first had the couple reduce their argument to a few words and the hand gestures. Then she literally stepped in between the two people and gestured like one of them. In doing this the therapist "stole" their anchors. She then asked the couple to do
something different. The behavioral patterns of these two clients could be considered a "loop" which was broken by the intervention of the therapist. (Simulated demonstration by therapists Feldman and Grask at the AAMFT conference, Washington, D.C., October 1983.)

NLP and Education. If the Neuro-Linguistic Programming model is substantially correct, then far-reaching implications seem to follow for education. According to Dilts and Myers-Anderson (1980), "Neuro-Linguistic Programming turns the teacher from a passive presenter of information to a skilled behavioral technician." One way in which NLP is alleged to aid the educator is through providing tools for determining the representational systems and strategies employed by students. This enables the teacher to design instruction to match the ways in which students actually apprehend and process information.

Understanding students' learning strategies, according to Dilts and Myers-Anderson (1980), helps prevent misapplication of the term "slow learner." They claim that many students thus labeled have as great an ability to learn as do their classmates, yet their learning strategies are not understood by their instructors:

In our experience many young people who have been labeled 'slow,' 'handicapped' or 'disabled' in this [school] context are far from 'stupid'--they simply have different strategies for learning that are not utilized by present techniques of education.

For example, kinesthetically oriented students, Dilts and Myers-Anderson claim, often have a difficult time in school due to the fact that classroom situations, methods of teaching, lesson plans, etc., are usually primarily oriented toward students who
favor visual and auditory ways of understanding. In this way the kinesthetic student is, in an important sense, discriminated against in the typical classroom. The teacher who recognizes this can make efforts to meet the kinesthetic student's special learning needs.

Dilts and Myers-Anderson (1980) cite the example of a kines-thetically-oriented student who had been having considerable difficulty with algebra until he had occasion to begin working with a blind fellow-student's braille:

Not so surprisingly the student's ability to pick up algebra using the braille and raised surface material was many times more rapid than when he attempted to do it visually. The braille paced his natural abilities and strategies with his tactile system.

Understanding how strategies are employed, and for what ends, also enables the educator to teach strategies which are appropriate to a particular task. Dilts et al. (1980) claim that the teacher should decide what the outcome of a teaching segment is to be and employ strategies accordingly:

One of the initial distinctions in outcomes to make is whether (1) you have a specific set of contents you must teach—that is, do you have to teach X number of people N amount of information or (2) you want to teach learning skills specifically targeted for the subject matter of your courses. In the first case you will want to pace and utilize your students' strategies and feed in the content. . . . In the second case you will want to install in your students the strategy most appropriate for the task or behavior in question.

Such a view presupposes one of the most significant claims of NLP, viz., there are best strategies for particular tasks and one can, through using NLP principles and methods, "explicitly map out those strategies which have proven to be most universally effective for learning the subject in question" (Dilts and Meyers-Anderson,
One claimed example of such a task—spelling—was given in the section on the Basics of NLP. The spelling teacher's role thus becomes, in this view, more than simply teaching spelling—it includes, as a main component, teaching strategies for spelling.

The spelling strategy devised by NLP has had notable success with learning-disabled children, according to McCoy (1981). In an article in The Rocky Mountain News, reporting on the work of a Denver consulting group using NLP methods to deal with brain damaged and learning-disabled children, McCoy states: "The result ... is breakthroughs in learning for the average student and even more dramatic results for the learning-disabled and the previously unresponsive student." A member of this group, Garth Johnston, notes the continuing effect on the learning-disabled student of gaining some success in spelling. As cited in the article: "[once the child learns] to spell a word like 'encyclopedia' or 'dictionary,' a change takes place inside him. He thinks, 'Hey, I'm not so dumb after all.'"

Neuro-Linguistic Programming's spelling strategy has also provided one focus for NLP's movement into the field of educational software. This has been one of the more visible developments of NLP and one to be expected given NLP's view of learning. Programs have also been developed for math, typing and writing, among other subjects. They are designed on the basis of NLP claims about what are the most successful strategies for learning in each of these task areas.
There are other areas of importance concerning the use and potential use of the NLP in education which will not be considered here. However, two further points can be noted:

1. Even if the NLP model is substantially mistaken, the emphasis that NLP places upon the process of learning seems commendable.

2. The educational setting provides an excellent arena for testing NLP claims. If the NLP model is substantially correct, then clear-cut and effective learning strategies for various tasks should be an expected outcome of that model; failure of such applications to develop might be considered evidence against the NLP model.

Other Applications of NLP. Although the earliest applications of Neuro-Linguistic Programming were in the areas of psychotherapy and education, its application to other areas, as diverse as medicine and advertising, has occurred rather rapidly. As early as August, 1983, at an NLP workshop in Larkspur, California, organizers made the following claim in an information pamphlet, "Who Should Attend this Seminar?":

The NLP Programmer Training has been created for all people who want to change themselves and help change others, including: Professionals, Professional Communicators, Educators, Psychotherapists, Physiotherapists, Holistic Health Practitioners, Sales Professionals, Counselors, Executives, Students, Social Workers, Writers, Organizational Development Consultants, Managers, Psychologists, Medical Professionals, Real Estate Professionals (NLP Center for Advanced Studies, 1983).

Clearly, the NLP organization takes its principles to be applicable to a large number of fields and professions.
This can be seen as an outgrowth of the NLP model itself. An information processing model which purports to explain why people behave as they do, why they miscommunicate, and how more effective communication can be brought about, can be expected to be applicable in any area where behavior modification or communication is considered important. This includes those areas listed above and many others. Moreover, because of the relative simplicity of the model as set out by the chief NLP proponents, the rudiments of the view can easily be taught. The result is that a number of efforts have been geared to instructing particular professional groups in NLP theory and methods.

Related Research

NLP offers a model of inner experience and its relation to behavior. One major aspect of this model is the view that individuals encode, store, process, and access information through three main representational systems—visual, auditory, or kinesthetic. A related aspect is that many individuals have a preferred or primary representational system (PRS). According to NLP advocates, these features of experience are often exhibited in behavior. Indeed, though the NLP model concerns the always problematic area (from the standpoint of scientific investigation) of subjective experience, it has the admirable quality of making some rather specific claims about the relations of inner experience to behavior. This makes it possible to test the model through paying attention to those behaviors which are alleged to indicate facts about inner experience.
Research attempting to test the NLP representational system model through investigation of behavioral indicators has generally been concerned with either patterns of predicate usage, patterns of eye movements, or both. It might seem curious that these have been the two main avenues of choice for NLP-related research to date, since NLP proponents have maintained that other kinds of behavior—for example posture, breathing rate, and voice quality—also vary with respect to representational systems. Yet claims by NLP advocates about ways in which predicate usage and eye movements vary with representational system modality have generally been more detailed and explicit than claims about other behavioral indicators. Moreover, of the various kinds of behaviors which NLP proponents have said are related to representational systems, predicate usage and eye movements seem to be the most easily observable and quantifiable in an experimental situation.

The claimed relationship of predicate usage to PRS is that individuals typically employ a greater number of predicates (verbs, adverbs and adjectives) reflective of their own PRS modality than of predicates which reflect other modalities. Such predicates include, for example, "see," "show," and "focus" as visual words; "tell," "heard," and "quiet," as auditory words; and "touch," "feel," and "angry" as kinesthetic words. ("Angry" and other words indicating emotions are kinesthetic words according to the NLP view because emotions are experienced as physical sensations). Thus a person whose speech typically exhibits a higher number of auditory predicates than either of the other two types of words would
properly be classified, according to the NLP model, as having an auditory PRS.

A second kind of behavior held to be related to representational systems is that of eye movements. According to this view right handed individuals, when asked non-sensory-specific questions, typically exhibit eye movements which indicate concurrently occurring modes of representation. Specifically, the following associations of eye movements and representational modes are claimed by NLP proponents (Dilts et al., 1980) (indicated directions are specified as relative to the subject):

- Up and to the right or left — visual
- Straight ahead — visual
- Level and to either side — auditory
- Down and to the left — auditory
- Down and to the right — kinesthetic

The bulk of NLP-related research to date has involved one or both of these alleged behavioral indicators. The results of that research have been mixed. While some studies seem to lend support to the NLP model, others do not. Indeed, findings of some studies seem to contradict those of others. Reasons for these discrepancies are not clear. However, it is notable that methodologies employed by researchers in this area have varied substantially (Dorn, 1983), and it is reasonable to assume that variations in results may sometimes be due to differences in and problems within experimental design.
Throughout this review particular attention is paid to the matter of methodologies. This is appropriate for the present study since it is largely an attempt to help lay a foundation for proper methodologies for eye movement investigation. However, the review is not limited to studies that have dealt with eye movement behavior. Because the amount of NLP related research done to date is limited, a wide range of investigations will be discussed here. In the course of the review some problems with prior research will be identified and suggestions will be made regarding methodological approaches in future NLP-related investigations.

Owens (1977), in one of the earliest NLP-related studies, presented nine stimulus questions to 128 undergraduate students with raters observing the subjects' responses and evaluating each for PRS based on predicate usage and eye movements. Each subject was also asked to complete a self-report identifying his or her own PRS. Owens found no significant correlation between self report and either of the other two measures, but he did find a positive correlation between the predicate usage and eye movement measures. These results seem to contradict Bandler's and Grinder's claim that one need only pay attention to predicate usage in order to determine PRS.

Ellis (1980), in reviewing Owen's (1977) study with respect to the discarding of "unspecifiable predicates" in the judgment of predicate usage, claimed that Owens failed to make clear how he decided which predicates were "readily identifiable." Moreover, he pointed out that Owens failed to clearly specify his criteria for
rater accuracy. It can also be noted that Owens' inclusion of self report might seem to imply that it is an accepted NLP method for determining PRS. However, although Dilts (1978) claimed that in his own experimentation self report was of "extreme importance" in clearing up observed discrepancies in eye movements, NLP proponents do not seem to consider self report to be an adequate independent measure of PRS. Dilts (1983c) has further stated, "Generally speaking, self report is not reliable... People don't really know their own primary representational system."

Shaw's (1977) study made use of some of the same subjects and data as did Owens' research. The presumed PRS of each participant was first determined by employing the predicate-usage method, with all participants identified as either visual (V) or kinesthetic (K). Each participant was then shown, via videotape, visual, auditory and kinesthetic versions of the same story. Results of this study showed that the participants showed no particular preference for the version of the story which was presented in their presumed PRS.

Ellis (1980) also criticized Shaw's (1977) study, claiming that it had problems in rating methodology similar to Owens'. Whether predicates are to be rated as visual, auditory, or kinesthetic is a central question in such studies, and Ellis noted that words like "light" are ambiguous and that context has to be considered in order to clarify which of the representational system categories the word falls under in a particular instance. He suggested breaking scoring down into scoring units and contextual units. He claimed that once a predicate has been identified, the meaning of a given phrase or
word is still open to interpretation. That is, though a linguistic element may be given its obvious meaning, it may actually refer to or mean something else not so obvious.

Ellis' own (1980) study also focused on predicate usage. Subjects for the study were 60 students from three majors—Art, Music, and P.E.—presumed to have visual, auditory, and kinesthetic orientations respectively. The students were interviewed and instructed to talk on four topics while they were audio taped: (a) describe a friend, (b) describe an hour, (c) describe a problem, and (d) describe an activity you enjoy. When predicate usage was scored, results indicated that two of the variables, visual predicates and auditory predicates, showed significant differences among majors in expected directions. However, these effects were not independent of topic.

Birholtz's (1981) study attempted to determine whether people have preferred modes of expression corresponding to the preferred modes of representation of the NLP model. Twenty-seven students were exposed to six stimulus questions on audio tape with their responses also taped. Birholtz recognized that stability of predicate usage is an important factor in judging preferred mode of expression, so in order to test for stability over time, another six similar questions were presented to each student a week later. Stability of reports over positive and negative experiences and over reports of past, present and future experiences was also tested. In the study's third week, the California Personality Inventory was given in order to identify possible correlations.
between modes of expression and personality characteristics. In the fourth week each subject filled out a self-report questionnaire in order that categorization by predicates could be compared to any self-categorization by subjects of preferred mode of expression.

Results showed a high proportion (20) of the students with a preferred mode of expression. However, the preferred mode in each case was the kinesthetic mode. Still, in Birholtz's view, this "adds credence to the assumption that individuals have a preferred way of representing the world." Stability of these preferred modes of expression occurred across time and across all other investigated variables. Further support for Bandler's and Grinder's claim, Birholtz added, came from finding that there was a correlation of predicate usage with personality characteristics. There were nine such significant correlations, five of these being between proportion of auditory predicates used and CPI measures. Also supporting the NLP model, Birholtz held, was the finding that the three most identified predicate modes were visual, auditory, and kinesthetic. She found no correlation between categorization by means of predicates and categorization by self report.

It is noteworthy that Birholtz's study dealt with one of Ellis's (1980) concerns, namely, the possible variance in rating of a predicate depending upon whether it is considered as standing alone or is taken in its context. Birholtz found that no significant difference in categorization of subjects according to predicate usage occurred whether words were checked for use in context or were not. Further, her findings indicated that whether
nouns were included or not included as being among the rated predicates made no difference to categorization.

Birholtz also recognized and dealt with another problem in predicate rating, namely the frequency with which the various predicates were used. She suggested two possible criteria for determining preferred mode of expression with respect to frequency, one in terms of the percent of usage of predicates belonging to each mode, the second in terms of the deviation of the usage of that mode by a particular subject from the mean usage of that mode across all subjects. Because the mean usage of kinesthetic predicates was so high among the subjects, Birholtz concluded that it would have taken an extremely high score in order to be classified in the kinesthetic preferred mode according to the second criterion. Therefore she chose the first. Specifically, she defined the preferred mode as "used 20 percentage points more frequently than the next most frequently used mode."

Such attempts at clear specification of criteria for judging preferred mode of expression are of utmost importance in research on predicate usage. The criteria must take into account both the categorization of individual words and the frequency of the words occurring within each category. These criteria are of course subject to change in the light of further research and as the NLP model itself is more clearly specified, but investigations in this area must begin with some reasonably precise stipulation of what having a "preferred mode of expression" consists in behaviorally, and the more clearly those criteria are set out, the more useful will be the
results of the research. It is perhaps especially important to strive for greater clarity in the case of investigations concerning predicate usage, since NLP proponents do not themselves seem to have specified clear criteria for making such judgments.

With respect to the problem of determining which categories predicates fall into, Gumm, Walker, and Day (1982) reported that the two predicate raters in their own investigation encountered numerous instances of difficulty in discriminating between two modalities based on predicate usage. Their study attempted to identify PRS through three methods: predicate usage, eye movements, and a self-report questionnaire. Predicate usage was judged on the basis of subjects' responses to instructions to talk for at least one minute on each of five topics. Eye movements were rated with the subjects' heads secured by a padded restraint. They concluded that there was no significant agreement among the three methods, and in reviewing Owens' (1977) study, they reported an error in his finding that a combination of eye movements and predicate usage was significant in determining PRS.

Several questions concerning methodology can be brought up with respect to the Gumm et al. investigation. First, apparently the raters did not record the number of predicates detected in the verbal reports of the subjects. Without this information it is difficult to judge not only the nature of the difficulties faced by the raters, but also to suggest any remediation.

Further, the statements eliciting discussion of the five topics may have introduced significant unwanted factors into the situation.
For example, the statement "I would like for you to describe a place, a scene or an activity" (Gumm, personal communication) includes the word "scene," which appears to be a visually oriented word. But the intention of Gumm et al. here should have been to provide neutral statements which would not bias a subject toward one mode of expression or another. Also, some of the statements (e.g., "Describe one of your most pleasant childhood experiences" and "Describe a place, a scene or an activity that provides you peace of mind") could have elicited mood states which might have affected later replies. Moreover, the statement, "I would like you to tell me about a time when you knew you were trusted," is quite personal and requests a dissociation state of the subject. Problems with stimulus questions or directives will arise again in the case of several studies discussed below.

With respect to the eye movement rating procedure, the technique of placing the subject's head in restraints is also questionable. Generally speaking, controlled experimentation involving human subjects is and must be an unnatural situation. Yet, since the intention of experimentation is usually the extrapolation of findings into the world outside the laboratory, each element of "unnaturalness" should be weighed carefully with respect to the possibility of biasing results. If that possibility is significant, then ways to avoid the element should be sought. Since NLP proponents have stated that non-language behaviors other than eye movements are associated with representational systems, it seems unwarranted to restrict the subject's ability to freely move a part
of his or her body when investigating NLP claims about representational systems. Note Dilts' (1984) statement: "Moving your shoulders is just as much of an accessing cue as moving your eyes." Moreover, head movements and eye movements are so closely connected that it is reasonable to suppose that restriction of the first could be experienced by a subject as a restriction on the second. Other methods of assuring correct rating of eye movements should be sought by the experimenter.

A final problem with the Gumm et al. study involves the question of what, upon presentation of a stimulus item, should be counted as the initial eye movement. In their study the first eye movement after a question was posed was rated as the initial eye movement. But it has been noted by Hernandez in relation to her own study (1981) that sometimes an eye movement will occur before a stimulus question has been completed. Indeed, it is arguable that this is to be expected, for a person's understanding of a spoken sentence is generally being formulated as the sentence is being heard. Specifically, if the sentence is a question, the formulation of a possible answer to the question will often begin before the question is finished. Or if the sentence is an instruction to do something, especially a task involving internal imagery, the beginning of carrying out the task may well begin before the instruction is completed. It thus appears that Gumm et al., by restricting initial eye movements to the first eye movement after the stimulus question, failed to take account of important facts about how we actually understand sentences.
Hernandez' (1981) investigation did not deal with predicates, but with eye movements alone. Sixty-four undergraduate students were given stimulus directives to either see, hear, feel, or think of a certain item (e.g., "Think of your first day at Ball State"), with six directives presented for each modality. It was supposed that the subjects, in carrying out the sensory-specific directives, would exhibit eye movements associated with those specific modes, whereas in carrying out the non-sensory specific directives to "think of" something, they would be freer to use any of the modes. The first eye movement exhibited following the request was recorded. Results showed significant correlations between NLP-predicted eye movement direction for all six visually-oriented questions (five at the .008 level and one at the .05 level), between eye position and four of the auditorily-oriented questions (three at .008, one at .05), and between eye position and two of the kinesthetically-oriented questions (at .05).

As mentioned above, Hernandez pointed out that in some cases eye movements were exhibited before the experimental statement was completed and, thus, that the eye movement which was scored may not have always reflected the subject's internal response to the experimental variable. She further noted that in some cases eye movements were quite rapid and the videotape could not be played back slowly enough in all cases for raters to make an accurate assessment.

In relation to this study, Johnson (1983) suggested a reason for the variances in number and significance of correlations among the three types of stimulus questions. He pointed out that visual
requests in Hernandez' study were an average of 8.83 words long, auditory requests an average of 9.5 words long and kinesthetic requests an average of 12 words long, suggesting that longer stimulus directives or questions may allow a subject to activate a sensory modality through synesthesia. More generally, Johnson held that the way in which a directive or question is presented may have a significant effect upon responses. For example, he noted Galin's and Ornstein's (1974) finding that beginning a request with "Tell me how many" seemed to elicit a different response than a question that began "How many."

Dorn, Atwater, Jered and Russell (1983) conducted a study investigating the stability of eye movement patterns over time. Twenty-six female undergraduates were interviewed twice in the space of one week by three male doctoral-level students in counseling psychology who were "familiar with the PRS model." Interviewers alternated, with each conducting approximately one third of the interviews while the other two observed through a one-way mirror. During the first interviews six stimulus questions were asked of each subject with the subject asked to raise her hand when she experienced an "internal response." The eye movement just prior to the hand signal was recorded by all three experimenters. A week later the subjects were again interviewed, using the same six questions, and rated.

For each interview the three raters determined a PRS from the six ratings for each subject. This was calculated upon the basis of the category (visual, auditory or kinesthetic) indicated most often
by observed eye movements. Following this, an overall rating of PRS was determined for each subject based upon the category "most often assigned" by the raters. Inter-rater coefficients were calculated for the first interview and yielded the following values: between experimenters A and B, 0.66; between A and C, 0.88; and between B and C, 0.78. Yet, despite some evident disagreement, Dorn et al. maintain that "At no time . . . was there a failure to agree upon a specific category between any two of the interviewers."

Correlating assigned PRSs for each student over the two interviews yielded a non-significant relationship according to the study, thus indicating instability of PRS over time. The investigators concluded that the results "discourage acceptance of the hypothesis of Bandler and Grinder . . . that eye movement will follow a pattern distinct to the PRS."

In discussing these results, Dorn et al. noted the possibility that rating of eye movements and assessment of PRS may be a highly specialized skill and suggested that some NLP proponents may be better able to employ their own theoretical model. Moreover, they pointed out that during debriefing of the subjects for this experiment, some indicated that in the second interview they remembered responses which they had given at the first interview. "Thus, the internal sequencing process might have taken a different form." This does seem to be an important methodological problem for the study.

It should be pointed out, however, that Dorn et al. treated the question of inter-rater reliability as an important aspect of their
study. This has not always been the case in NLP-related research. Radosta (1982), for example, although utilizing two raters, apparently did not determine their inter-rater reliability because "both the researcher and the independent rater were considered qualified by training to correctly rate eye movements." Also, Falzett's (1981) investigation, although employing two raters, apparently included no inter-rater reliability measure.

Krim's (1983) study also investigated eye movements through asking stimulus questions in two separate interviews. Indeed, by asking 21 similar but different questions in the two interviews he seems (1) to have avoided the problem of Dorn et al. and (2) to have had the opportunity to investigate stability of patterns over time. Yet in regard to (1) he did not, as will be discussed below, succeed in avoiding the same problem, and in regard to (2), although he had the opportunity to investigate stability, he apparently failed to do so. In his study, 30 subjects were videotaped during both interviews with their eye movements later scored by three raters. Although Krim's hypotheses were not clearly stated in this study, he apparently claimed to have found statistically significant correlations between visually-oriented questions and eye movements indicative of visual processing, and between auditorily-oriented questions and eye movements indicating auditory processing.

Notably, as the second main question in his study, Krim included an investigation of inter-rater reliability. He thus gave the matter a place of central importance which is apparently unique in NLP-related studies excepting the present one. He claimed to
have found consistency among his raters at a statistically significant level.

Unfortunately, there are many problems with Krim's study. These problems begin, as noted above, with his statements of hypotheses. Further, his compilation of data for the first part of the experiment included, without explanation, only six questions of each sensory-specific type, whereas 14 of each type were apparently asked each subject. This loose handling of data continued for the part of the study dealing with inter-rater reliability. Some columns in tables presenting this data were misadded. More importantly, given his presentation of data and very brief discussion, it appears that Krim based his calculations of inter-rater reliability solely upon a comparison of raters with respect to the total number of eye movements of each type recorded by each. Such a method is clearly inadequate. Any calculation of inter-rater reliability must compare judgments of the raters with respect to when and for what subjects the judgments occurred. Krim seems not to have done this.

Also, Krim's method of rating eye movements involved rating the first movement observed after a stimulus question has been completed. This practice has been called into question above, with respect to Hernandez (1981) and Gumm et al. (1982). Moreover, there are problems with Krim's stimulus questions. Perhaps the most obvious is that in his dissertation he did not categorize them according to stimulus type. Perhaps he considered this to be obvious from the question itself. But one wonders what category the
question, "On a telephone, what letters correspond with the number two?" falls into, since a telephone is an instrument which is so closely connected with sound. Further, several questions seem to have required a dissociation state—for example, "If you were writing a poem, make up the next line that would rhyme with: 'Now you've just turned twenty-nine.'" Finally, although Krim's second set of questions was not identical with the first and he, thus, seems to have avoided the problem of Dorn et al., several questions in the second set were very similar to questions in the first. Compare "How were you feeling when you first entered this room? (first set) with "How were you feeling when you entered the room?" It seems likely that some questions in the second set might have elicited memories, and thus strategies different from those employed by subjects in answering corresponding questions in the first set. In fact, it is by no means difficult to suppose that a subject might have spent a good deal of time during the second interview in the cognitive exercise of attempting to determine whether in fact he or she were being asked the same set of questions.

Johnson's (1983) study avoided the use of stimulus questions, holding that "questions to elicit a scorable response should be limited because of their proven unreliability." His investigation attempted to determine a correlation between eye movements and predicate usage by having each of 30 subjects (15 male and 15 female) converse with an interviewer about one or two pleasant remembered experiences while being both video and audio taped. Afterwards the audio tapes were transcribed. Then, while an
experimenter-trained rater observed the videotape and called out eye position changes, the experimenter noted each position at its corresponding place on the transcript. A "scorable unit" was a recorded eye position, other than up, center, or down, that was followed within two seconds by a sensory-based spoken word. Johnson's main finding was a significant correlation, at the .0001 level, between NLP-predicted eye movements and sensory-based words spoken. Thus his findings seem to support the NLP model.

The overall methodology of Johnson's study appears to be unique in NLP-related research to date, and it deserves further consideration. Johnson indicated that an important factor in designing eye-movement studies is the naturalness of the situation, and he apparently believed that "drawing out" a subject in conversation, as opposed to simply presenting stimulus questions, increases this quality. He seems to be correct here. If the subject is engaged in a rather unstructured conversation he is likely to be less self conscious and less affected by artificial elements of the situation such as special equipment or procedures being used. Moreover, it goes without saying that a conversational situation approximates more closely the interaction of a counselor and a client than does the presentation of stimulus questions. Generally, observed behaviors in this kind of situation can be expected to approximate more closely those outside the laboratory.

Such a relatively free-flowing interchange might seem at first to result in a much increased difficulty in structuring a valid experiment. But Johnson's method of identifying eye movements and
predicates independently in a taped conversation and then correlating the two seems not only to be a valid test of one prediction of the NLP model but also to be a procedure that can be carried out in a sufficiently rigorous way. This "tightening up" would include a fuller specification of what predicates were to be considered associated with what sensory-modes, something which Johnson's report failed to do.

Another apparently unique element in Johnson's study was his method of measuring inter-rater reliability. His procedure was to have a second trained rater score the tapes of three of the subjects along with random tape segments for all of the subjects. The three subjects who were rated in their entirety were apparently selected because, due to their particular characteristics, they could be expected to be more difficult to rate correctly than most of the subjects. Johnson's inter-rater agreement measure showed an overall 86% agreement on a total of 398 rated movements.

It should be noted that Johnson's study also made the important distinction between lead system and representational system, a distinction which some of the studies reviewed failed to make. In NLP theory, the "lead system" is that representational system which an individual uses to access other systems, whereas the primary representational system is the representational system which people tend to favor in bringing information to consciousness (Lankton, 1980). That these may differ according to the NLP model is apparently sometimes not recognized by researchers or else, if recognized, is ignored (see, e.g., Carbonell, 1985). Yet the
Positing of a lead system is only partly the acknowledgement that an individual's having a PRS does not imply that all his/her representations are in that modality. For example, a person with a kinesthetic PRS could well employ, according to NLP theory, a spelling strategy which began with a visualization of the word. This immediately points to what seems to be a problem in some NLP-related studies, namely, the apparent assumption that NLP theory holds that the first eye movement after, for example, a stimulus question reveals the PRS. But this does not seem to be the case, since in NLP theory a first eye movement can indicate not the PRS but rather the lead system used to access another system.

Petroski's (1985) study attempted to provide support for the concept of representational systems by seeking correlations between eye movements, predicate usage, self report and several other variables. Petroski noted the importance of providing as natural a setting as possible for such investigations and thus attempted to closely replicate certain aspects of a clinical interview. Experimenters provided each of 64 subjects with both sensory-specified and unspecified stimulus items of various types (questions, directives, and tasks) while being video and audiotaped. Afterwards, each subject completed a self-report questionnaire on internal experience, had height and weight measurements taken in order to determine somatotype, and wrote a brief paper describing his or her experience in participating in the study.

Three independent raters unfamiliar with the NLP model and unaware of the hypotheses of the study rated the subjects' eye
movements, for each stimulus item, at four different points: at the beginning of the stimulus item, at the end of the stimulus item, during latency, and at the beginning of the subject's verbalization. Of most interest to Petroski were the eye movements during latency and during verbalization. For individual eye movements, ratings assigned to subjects were determined by applying criteria of agreement among the three raters and then assigning the agreed-upon rating when those criteria were met. The raters also scored the subjects' predicate usage, both during the interview and within the written exercise.

A hypothesized PRS was assigned to each subject for each of the five main measures: eye movements during latency, eye movements at verbalization, predicate usage during verbalization, written predicate usage, and self-report. Results showed no significant correlations among these different methods of categorizing the subjects. Moreover, overall assignments of PRS based upon various combinations of the five main measures showed little agreement. However, a significant relationship was found between eye movements during latency and oral predicates used (at the .05 level).

Petroski's study is admirable in offering different kinds of stimulus items, seeking correlations among a number of different variables, and attempting a thorough statistical analysis of results. This is surely a substantial improvement in methodology over some studies reviewed. He also did not require any head restraints and, in fact, did not even instruct the subjects to hold their heads in a particular position, arguing that the eye movements
of interest were "based on naturalistic observations of a subject's eyes without any artificial constraints." Further, an inter-rater reliability check was performed for both eye movement and predicate ratings, yielding a rater agreement significant at the .001 level. Finally, the investigation also paid attention to the concept of the lead system and entertained the possibility that some of the data could be attributed to a lead system, different from the PRS, being in operation.

It was noted above (p. 36) in relation to Owens' (1977) investigation that self-report does not seem to be an accepted means for determining PRS according to NLP theory. Petroski, however, treated it as an approved method of NLP. This led him, in determining whether some combinations of the five main measures might agree with one another in indicating PRS, to include self report within all his tested combinations. Not including self-report in these combinations may have led to results more in keeping with NLP predictions. On this same line, it seems notable that in the midst of findings otherwise apparently unsupportive of the NLP model, the significant relationship found between eye movements during latency and oral predicate usage was perhaps the one relationship that, given NLP theory, would be most expected.

A number of studies have dealt with NLP and psychotherapy. These include investigations by Allen (1982) and Liberman (1984) on the treatment of phobias using NLP techniques and research by Ehrmantraut (1983) and Thomason (1984) on comparisons of NLP-trained and non NLP-trained psychotherapists. One rather specific
prediction of the NLP model is that counselor effectiveness is increased when the counselor matches his language to the PRS of the client. This technique, called "predicate matching," consists in the counselor using verbs, adverbs, and adjectives which match the mode of the client's PRS. The method is alleged to increase rapport between counselor and client and to aid in counselor-client communication. If shown to work, the technique would not only provide some corroboration for the NLP model, but would likely also prove to be of great value in the practices of counselors and psychotherapists. It is thus often taken to be one of the more significant predictions of NLP, and a number of studies have been centered upon testing it.

Mattar (1980) attempted to test the claim by determining whether identification of PRS could provide a basis for improved comprehension and communication. Eighty undergraduate psychology students were tested for PRS through predicate usage. No subjects with an auditory preference were found, so the remainder of the study was conducted on those who had been identified as having a visual or kinesthetic PRS. Twenty from each group listened to a taped script of a therapeutic session which used either primarily visual or primarily kinesthetic predicates. Results showed no difference between the groups with respect to overall comprehension of either of the tapes. However, dividing comprehension into General Information (GI) and Specific Predicate Usage (SPU) segments resulted in the finding that visually-oriented subjects comprehended visual SPU segments better than did the kinesthetically-oriented
subjects and that kinesthetically-oriented subjects comprehended kinesthetic SPU segments better than did the visually oriented subjects. Mattar concluded that a therapist who uses visual predicates for subjects with a visual PRS and kinesthetic predicates for subjects with a kinesthetic PRS cannot expect better general information comprehension, but can expect that specific bits of information contained in the message will be better understood.

Cody (1983), asking stimulus questions and observing eye movements in two separate interview situations, attempted to determine whether representational systems of 105 college students could be reliably determined. He found that few of the subjects exhibited more than marginal preferences and that no more than a modest degree of temporal stability was found in preferences noted.

On the basis of these findings and utilizing the same students, he conducted a second experiment wherein subjects evaluated audiotaped vignettes in which commonplace experiences were presented in each of three versions—visual, auditory and kinesthetic. He found no relationship between representational preferences and subjects' preferences among versions. In a third experiment, the same subjects were asked to evaluate therapists as they listened to what were alleged to be actual counseling sessions. These excerpts, which were staged, varied with respect to linguistic matching between the therapist's language and representational preferences. Results showed that the therapists who matched predicates were evaluated by the subjects as less trustworthy and effective than therapists who did not match predicates.
Given the population that Cody worked with in his second and third experiments, namely one wherein few subjects exhibited more than "marginal preferences" for representational systems and where only a "modest" degree of temporal stability was found, it is doubtful that the results of those experiments are of much value. Moreover, it does not seem to follow from the NLP model that an outside observer of a counseling session will judge a counselor who matches predicates to be more trustworthy or effective. The NLP model prediction at issue concerns the client's judgment of the counselor.

Ellickson's (1980) study also dealt with predicate matching and counselor effectiveness. Her investigation attempted to determine whether interviewer-therapists responding with perceptual predicates congruent to the PRS of subject-clients would lead to the therapeutic relationship being enhanced and the subject-clients reporting being understood. The population consisted of 72 undergraduates. One half were interviewed and had their predicates matched by the interviewer, as indicated by their eye movements during the interview, while the predicates of the other half were mismatched. Specifically, the interviewer presented each subject with stimulus statements intended to (1) evoke a response from the subject and (2) be neutral with respect to representational system. The interviewer noted the subject's first eye movement following presentation of the stimulus. Then, following the subject's response, the interviewer asked for more detail, including within this request at least one predicate matching or mismatching the representational system.
indicated by the subject's eye movement. Four interviewers trained in the model of representational systems and in accurately identifying eye movements and responding with appropriate predicates were used, with one male and one female interviewer assigned to each of the treatment conditions.

Findings showed that sex of interviewer by sex of subject interaction was statistically significant. Males were at ease with males more fully than with female interviewers, while females appeared equally comfortable with male or female interviewers. No other significant findings came out of this study.

Ellickson noted a possible problem with the experiment. In some cases the subject's eye movements may not have matched his or her own verbalizations. Not controlling for this may have affected the subject's ability to perceive interviewers who were in the congruent condition as easier to communicate with and/or more empathic.

There are also problems with the stimulus statements used in the study. For example, "I'd like you to think about your earliest memory on your first day of school" seems a clumsy and therefore possibly confusing request for information. Another question asks the subject to think and experience at the same time: "I'd like you to think about experiencing a day at the beach." This is different than to "remember" experiencing. Moreover, though Ellickson claimed to use no specific predicates in stimulus statements, "I'd like you to think about a time when you knew that someone really cared about you" is asking for recall of an emotional (K) experience and thus possibly produces a bias toward a downward and right eye movement.
An even more fundamental criticism of Ellickson's study, however, is pointed out by Carbonell (1985), who maintains that the procedure of identifying an eye movement and then responding with a statement congruent with that movement (with respect to indicated representational system) is not a valid test of NLP technique. Carbonell points out that NLP theory stipulates that predicate matching by a counselor should be based upon a prior identification of the client's PRS. In fact, this criticism can be developed even further. Ellickson's method seems to presume that occurrence of an eye movement after a stimulus item is, according to NLP theory, a sufficient indication of PRS. But although admittedly NLP writings are unclear upon the matter of just how preponderant a behavioral indicator must be in order to confidently assign PRS, nowhere in these writings is it claimed that an individual with one preferred way of representing the world will not have representations in other modes. (See the discussion of "lead system" above, pp. 50-51). NLP theory would thus presumably hold that some of the eye movements in Ellickson's experiment were likely not indicators of PRS and thus that interviewers in the matching condition probably used mismatching predicates part of the time. This clearly is counter to NLP approved technique for establishing rapport.

Carbonell's (1985) study paid close attention to methodology in prior NLP-related research. Noting what he considered to be serious methodological problems of that research, he attempted to avoid those difficulties in investigating the effects of predicate matching on perceived counselor qualities. His study involved 64
undergraduates who were videotaped while responding to 10 sensory-ambiguous questions and whose eye movements were later scored by two independent raters. Primary representational systems were determined on the basis of these ratings and at a second interview the students were exposed to a fully scripted interview with predicates either matched or mismatched by one of two female interviewers. Each subject afterwards rated each interviewer on the Truax-Carkhuff empathy scale and on three subscales (attractiveness, expertness, and trustworthiness) of the Counselor Rating Form. Carbonell found no significant differences in counselor perception between subjects in the matched condition and subjects in the unmatched condition on any of the scales.

One unique element of Carbonell's experimental design is the absence of the interviewer from the experiment room. During both the questioning session and the scripted interview session, interviewers communicated with subjects via intercom. But use of this procedure for the second part of the experiment seems highly questionable. Placing the interviewer outside the room during this part of the study constituted a significant departure from the normal counselor-client situation. It is not clear that any useful conclusions about, for example, empathy with a counselor in a normal interview situation can be gotten from comparing measures of subjects' empathy with a disembodied voice. In short, Carbonell may have constructed an inadequate analog for the counselor-client interview.
What makes this procedure puzzling is Carbonell's rationale for it. Citing Kinsbourne's (1972) review, Carbonell suggested that "the presence of an experimenter in front of the subject tends to enhance the frequency of eye movements to the side." He held that placing the interviewer outside the room during both sessions was for the purpose of avoiding this effect. But his stated reason seems to have no bearing upon placing the interviewer outside the room during the second session, for, although subjects' eye movements were also videotaped and rated for that session, this was apparently done only as a secondary check for inter-rater reliability. Indeed, the experiment required determination of subjects' PRSs based upon first session tapes alone in order that the matched and mismatched conditions could be applied during the second session. Thus, assuming that inter-rater reliability was also adequately determined on the basis of first-session tapes—which it appears to have been—there seems no good reason for requiring the interviewer to speak via intercom during the second session.

However, there is one possible rationale for utilizing this procedure. It concerns the nature of Carbonell's "fully-scripted" interview. There were three versions of this interview (visual, auditory, and kinesthetic), each running about three pages long. Further, there were various decision points within each version wherein the interviewer was offered several options, depending upon the subject's response. Memorizing and then correctly and naturally applying such complex scripts in a face-to-face interview situation
would have been a formidable task for interviewers. At the same
time, for the interviewer to have referred to the script in the
face-to-face situation would certainly have been an even more
radical departure from the normal counselor-client situation than
having the interviewer speaking over the intercom. But, given this
difficulty, the proper solution seems not to have been placement of
the interviewer in the next room, but rather to either have simpli-
fied the script or perhaps, better yet, to have done away with it
altogether and to have trained interviewers in the method of
matching and mismatching predicates and to have allowed a more free-
flowing interview. This would, of course, have required devising
ways to adequately compare the manipulations of the interviewers in
the various conditions, a requirement which Carbonell rightly set
out. But it should be noted that it would not have violated another
important requirement which Carbonell made for such experiments,
namely that the interviewers who apply matched or mismatched
predicates be blind to the condition (the PRS) of the subject in
order to avoid the effects experimenter expectation.

Another interesting aspect of Carbonell's procedure was the use
of a chin rest during the first session in order to ensure that
subjects' eyes can be clearly videotaped. The use of such arti-
ficial head-movement restraints was discussed above (pp. 41-42) in
the study of Gumm et al. (1982). Again, since eye and head movement
so often occur in unison, such a restriction seems too likely to
confound experimental results.
Unlike some other researchers, Carbonell established a minimum number of eye movements (three) that must occur in order for a PRS to be assigned, with the most frequently displayed eye movement being the basis for the assignment. He claimed that this is more in keeping with Bandler's and Grinder's writings than the practice of determining PRS on the basis of a single eye movement. He seems correct in this, yet at the same time it is difficult to be confident that a mere summation of eye movements is the correct method for determining PRS. This uncertainty is due to the fact that NLP proponents simply have not been clear on the matter.

Using the minimum of three eye movements, Carbonell was able to assign a PRS to only 42 out of 64 subjects. This is indeed surprising, since a total of 10 questions were asked of the subjects. Carbonell suggested that the presence of the video camera may have focused the attention of subjects, thus reducing the frequency of lateral eye movements. He recommended placing the video camera in a position where it is not visible to the subjects.

What Carbonell calls "a more basic criticism of the present study" is that it tests a single NLP technique in isolation from the total range of counselor-client interaction. He holds that although results of the experiment do not support the NLP model, this should not be taken as implying that therapists using NLP techniques will have no positive results. But he suggests that these will be due to "other factors," such as the confidence displayed by some NLP practitioners and the image of NLP as a new and potent approach to therapy.
In the above discussion of NLP-related research a number of important methodological problems have been brought to light. These include:

1. The treatment of self-report as an NLP-approved method for determining PRS, although it does not seem to be regarded so by NLP proponents.
2. In studies attempting to determine PRS on the basis of predicate usage, the failure to specify clear criteria for categorizing predicates.
3. In predicate usage studies, failure to specify clear criteria of frequency of predicate usage upon which to base judgments assigning PRS.
4. Failure to adequately control stimulus questions. (This includes using sensory-specific words in what are intended to be sensory-ambiguous questions, using questions which request dissociation states, using confusing questions, and using identical or near-identical questions in two interviews of the same subject.)
5. The problem of determining which is the first eye movement that should be counted as being relevant to a stimulus question.
6. The fundamental problem of determining which and how many eye movements are indicative of PRS. (For example, whether only one—the "first"—as several of the studies seem to assume, or whether Carbonell is correct in saying that it is a series of eye movements).
7. Related to both 5 and 6, the problem of integrating the NLP concept of the "lead system" into experiments. Only two of the above studies seem to have done this at all. Carbonell (1985), it should be noted, claims that the concept has perhaps been ignored because researchers recognize that it makes the NLP model unfalsifiable. Though he provides little argument for this claim, he does ignore the concept in his own investigation.

8. The problem of designing the experimental situation so that artificial constraints do not invalidate the experiment's result. Use of subject restraining devices, video cameras, interview techniques (including use of stimulus questions as opposed to open-ended interview techniques) are all features of the experiment that must be weighed carefully by the experimenter.

9. Failure to use well-trained raters and/or to adequately measure inter-rater agreement.

Given these methodological problems it is not surprising that findings of NLP-related studies have often contradicted one another. An underlying reason for some of the problems seems to be a certain unclarity about just what the NLP model does predict. Such fundamental matters as the role of the "lead" system—including its relation to PRS and its effect upon eye movements—and whether PRS should be associated with the first eye movement after some stimulus or rather with a series of eye movements, do not seem to have been thoroughly dealt with by any of the studies reviewed. One
result is that different studies have, to some extent, been investigating different NLP models, while the assumption in all of the studies seems to be that there is only one NLP model. Any future research must attempt, in the face of a good deal of unclarity in some of the writings of NLP proponents, to determine what that one model is.

Of particular relevance for the present study is the last methodological concern listed above. Although the importance of adequately measuring interrater agreement in NLP-related eye-movement research seems to have been more fully recognized in recent studies, there is still need for standardized methodologies for rating eye movements and for determining inter-rater reliability. This will become even more important if future research should follow Carbonell (1985) in holding that the NLP prediction concerning eye movements and PRS involves a series of eye movements and not just one.

With respect to the training of raters for research projects, it is notable that there exists a large population of previously trained raters available for NLP-related studies--NLP Practitioners. Yet none of the studies reviewed seems to have made use of this population. Determining the inter-rater agreement among a reasonably large sample from this group will not only be of interest in itself, since many of these individuals are putting NLP techniques into practice in real-world settings, it will also help future researchers to decide whether raters for their studies should be selected from the group.
Utilizing NLP practitioners would of course be inappropriate in studies where it is important that raters be ignorant of the NLP model. Petroski (1985), for example, claims this to be the case in his own study, although it is not clear that he is correct. In any event, a study such as Johnson's (1983) investigation seems not to be biased in any way by the rater's knowledge of NLP. It can also be noted in relation to Johnson's study that use of NLP Practitioners as interviewers in an open-ended interview situation such as he described might be of particular benefit in drawing out subjects.
CHAPTER III

METHODOLOGY

The purpose of this study was to develop a methodology to code and quantify eye movement patterns as a function of a Neuro-Linguistic Programming (NLP) strategy elicitation rating process. A secondary purpose was to establish a methodology that would determine the degree of consistency among certain certified NLP Practitioners while they were rating eye movements of a given strategy.

The Proposed Research Model

Research is limited in NLP to such a degree that no adequate models existed that might be replicated or modified for the purposes of this project. Further, the areas of needed research are so broad (p. 2) and involve the interaction of so many complex factors, it was necessary to identify one relatively narrow component of the NLP concept, the rating of eye movements, as a manageable point of initiation.

Several important design steps were implemented in this research:

1. Development of a protocol to record eye movements of videotaped subjects designed to provide a consistent view of taped-subject eyes with all distractors eliminated.
2. Selection of the eye movement indicators that the raters were asked to identify -- A (audio), V (Visual), etc.
3. Development of a data recording procedure that the raters utilized to record or chart observations on each subject.

4. Development of a procedure to quantify the degree or percentage of agreement among raters in recording the observed eye movements.

5. Identification of a research population of certified NLP raters.


The following sections of this chapter present the details of the proposed methodology.

Selection of Video Tape Subjects

Individuals who were videotaped were drawn from the population of students enrolled in General Psychology classes at Oregon State University during the Spring quarter of 1987. Students in General Psychology classes were informed, by written notices posted on the Psychology bulletin board, of the opportunity to earn extra credit for participating in the present research project.

Interested students were given a Video Tape Subject Information Form to complete. At this time, the students were assured that, if chosen for the study, they would be guaranteed confidentiality through the assignment of an identification number which would be used to identify them throughout the study. The Video Tape Subject Information Form was used to screen out those students with a history of:

1) Hearing deficiencies
2) Corrected vision with eyeglasses

3) Physical conditions requiring medication

Those with hearing deficiencies were screened out to lessen confusion between elicitor and student. Those with vision corrected by eyeglasses were screened out because eyeglasses might create interference for video equipment. Those requiring medication for physical conditions were screened out to eliminate possible confusion due to drug influences in listening and response.

Further, all video tape subjects were:

1) Right handed

2) Native English speakers

Only right handed individuals were used because proponents of NLP believe that left handed individuals elicit different eye movements than those who are right handed (Dilts, 1981). Those who were not native English speakers were screened out in order to lessen possible confusion between the elicitor and the video tape subject.

**Instructions to Video Tape Subjects**

The 18 individuals who were videotaped were told, prior to the testing procedure, that they would be asked to spell some words and that during the spelling time they would be videotaped. They were told that there would be no penalty of any kind for misspelling, that the video tape would be viewed by professionals concerned with the issues of the study, and finally that they would be welcome to a copy of the summary of the results of the study when it was completed.
Videotaping Procedure

Videotaping was done on the Oregon State University campus. Each of 10 video tape subjects (see "Selection of Video Tape Subjects") was asked by an elicitor to spell the following 10 words:

- pharaoh
- license
- gene
- surprise
- receive
- sincerely
- beautiful
- lacerate
- antique
- cologne

These words were chosen as being in a middle range of difficulty for college students. That is, they are generally words that the student would be required to "think about" in order to spell correctly; yet they are fairly common one-, two-, or three-syllable words, whose correct spelling was expected to be not overly difficult for a substantial percentage of college students.

Each video tape subject was videotaped from the eyes upward as he or she spelled each word. The same procedure, including the order of words to be spelled, was followed for each. After the video tape subject had been seated and while the video equipment was operating, several trial words were given to the student. When the student was judged to be comfortable with the procedure, the first of the 10 words was stated by the elicitor. When it had become clear to the elicitor that the last letter had been stated by the subject in his or her attempt to spell the word, the elicitor said "thank you." After five seconds, the elicitor stated the next word
to be spelled. This procedure continued until the video tape subject had made an attempt to spell all 10 words.

Eye Movement Indicators

In NLP, eye movements are said to be indicators of types of representational systems employed in internal processing (Dilts, 1980). Generally, upward eye movements to either the right or left are held to be indicators of visual (V) processing. Eyes level but to either the right or left or downward to the viewer's left are held to be indicators of auditory (A) processing. Eye movements directed downward to the right are held to be indicators of kinesthetic (K) processing. Since the V-A-K notations were presumed to be familiar to all certified NLP Practitioners involved in this study, they were used as indicators of observed eye movements. A fourth notation, "C", indicating that the subject's eyes were in a centered and defocused position, was also employed.

The Sample

The sample for this study was derived from directories of certified NLP Practitioners published by Grinder, Delozier, and Associates (1987). Names of currently licensed Practitioners practicing on the West Coast were obtained from the list. Fifteen in all took part in this study.

The researcher met with each Practitioner at a mutually agreed upon time and location in or near the Practitioner's city of
residence. The video tape review and data collection process related to each Practitioner took place in this location.

It should be noted that Practitioners are licensed for a period of two years, at which time they must be retested. The sample for this study was chosen from currently licensed NLP Practitioners.

**Rating Procedure**

Certified NLP Practitioners were shown video tapes of each student performing the 10 spelling tasks. Eye movements were recorded by the Practitioner on the Individual Rater Chart (IRC) for each student at the beginning of, during and upon completion of each spelling task (see Figure 1).

For each word, the video tape subject's responses were rated on the IRC in the seven rating boxes corresponding to the word. Immediately after pronunciation of the word, the first observed eye movement response was recorded in rating box number one for that word, using one of the eye movement indicators: A—audio, V—visual, K—kinesthetic, or C—centered or defocused. Any subsequent observed eye movements of the video tape subject as he or she spelled that word was recorded by the rater in adjacent rating boxes until the rater judged that the individual's information processing strategy had ended for the word. At that point the rater recorded "NA" (Not Applicable) in the next adjacent rating box and continued to record NA in any remaining rating boxes corresponding to that word.

Prior to observing the video tape subjects, three practice scoring runs were given to the Practitioner with the researcher
### Figure 1. Individual Rater Chart

<table>
<thead>
<tr>
<th>Word</th>
<th>Subject 1</th>
<th>Subject 2</th>
<th>Subject 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Eye Movement</td>
<td>Eye Movement</td>
<td>Eye Movement</td>
</tr>
<tr>
<td>1</td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Word</th>
<th>Subject 4</th>
<th>Subject 5</th>
<th>Subject 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Eye Movement</td>
<td>Eye Movement</td>
<td>Eye Movement</td>
</tr>
<tr>
<td>1</td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rater</td>
<td>Subject 7</td>
<td>Subject 8</td>
<td>Subject 9</td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td>Eye Movement</td>
<td>Eye Movement</td>
<td>Eye Movement</td>
</tr>
<tr>
<td>1</td>
<td>2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye Movement</td>
</tr>
<tr>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>

Figure 1. Continued
present. This was to enable the Practitioner to become at ease with the equipment and the setting. In addition, the practice scoring runs helped ensure that Practitioners, who would ordinarily rate several aspects of behavior, understood that only selective rating of eye movements was to be carried out in the research situation.

Analysis of Data

Data analysis involved a five-step process:

1. Determination of agreement among raters by subject and eye movement—the initial agreement measures

2. Determination of agreement among raters by subject and eye movement, across words—second level agreement measures

3. Determination of agreement among raters by eye movement, across subjects and words—third level agreement measures

4. Determination of a grand mean—an overall measure of agreement among raters across all eye movements

5. Computation of a Z score allowing one to compare movements with respect to agreement.

At the outset of explaining the analysis process it should be noted that the data collection procedure described in the preceding section was a forced-choice procedure. That is, each rater was required to record one of the five eye movement notations—V, K, A, C, or NA—in each rating box. This ensured a data point in each of the 700 cells of each Individual Rater Chart, with a total of 10,500 data points for the experiment.
However, an important limitation was set on the use of these data points for analysis. Since it was expected that for most spelling tasks fewer than seven eye movements would be noted by each Practitioner and, consequently, that the latter rating boxes for most tasks would be composed of "NA" notations, a limitation was set on using data from the latter rating boxes; otherwise, a meaningless inflation of agreement scores might occur. So, for each task, data for a particular set of rating boxes (1 through 7) was taken into account only so long as at least one rater had not marked "NA" for that box. For example, in spelling task one of Subject 1, if 14 raters noted three eye movements and each then recorded "NA" in the remaining four boxes, while one rater noted four eye movements and recorded "NA" in the remaining three boxes, only the first four sets of rating boxes were relevant for that spelling task. Boxes wherein all raters marked "NA" were not taken into account.

The expectation that fewer than seven eye movements would in most cases be noted was based upon results from previous studies (see, e.g., the discussion of Carbonell [1985] on p. 62 of the Literature Review). Yet this expectation was not conclusively justified by prior research, and thus seven rating boxes were used in this study to help ensure adequate space for recording observations. If the expectation proved correct, this of course might provide evidence that similar future studies should use fewer rating boxes.

Given the above-described limitation on data points held to be relevant, the overall process of analysis was simply one of developing, beginning with the Practitioners' notations, a series of
agreement measures which covered larger and larger arrays of data points, leading finally to an overall agreement measure. The five steps of the process are outlined in greater detail below.

Step 1 - Determination of Agreement by Subject, Word, and Eye Movement

The purpose of this step was to measure agreement among raters with respect to each rating box on the IRC. This was achieved by first transferring data from the 15 IRCs to the 10 All Raters Charts (ARC) (Figure 2). Each ARC pertained to one subject and allowed a direct comparison of notations made by all raters for that subject. On the basis of this compilation, the percentage occurrence of each rater notation for each rating box on the IRC was determined. These percentages were, in turn, the basis for deriving an agreement measure among all raters for each rating box on the IRC. The initial agreement measures (M) were determined according to the following formula:

\[
M = \frac{V^2 + C^2 + A^2 + K^2 + 'NA'^2}{100}
\]

where, with respect to each rating box on the IRC,

\[V = \% \text{ occurrence of the "V" notation}\]
\[A = \% \text{ occurrence of the "A" notation}\]
\[K = \% \text{ occurrence of the "K" notation}\]
\[C = \% \text{ occurrence of the "C" notation}\]
\['NA' = \% \text{ occurrence of the "NA" notation (in cases where at least one rater had marked something other than "NA" for that box)}\].
<table>
<thead>
<tr>
<th>Rater</th>
<th>Subject</th>
<th>WORD 1</th>
<th>WORD 2</th>
<th>WORD 3</th>
<th>WORD 4</th>
<th>WORD 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>1 2 3 4 5 6 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. The All Raters Chart
Figure 2. Continued
These agreement measures for each rating box amounted to measures of agreement by subject, word, and eye movement; they were recorded in the last row of the ARC.

An example of this process would be a case where, for a particular subject and Eye Movement, notations of the 15 raters included three C's, two A's and 10 V's. Dividing by 15 yields the following percentages: 20% C's, 13.3% A's, 66.7% V's, 0% K's, and 0% NA's. Entering these figures into the above formula gives:

\[
\frac{20^2 + 13.3^2 + 66.7^2 + 0^2 + 0^2}{100} = 50.23
\]

The result, 50.23, is the initial agreement measure in this case.

**Step 2 - Determination of Agreement by Subject and Eye Movement across Words**

This step's purpose was to develop more comprehensive measures of rater agreement by eye movements, namely, measures of agreement across all words spelled by a particular student. To help accomplish this, the initial agreement measures recorded on the ARC were transferred to a third chart, the Agreement by Subjects Chart (ASC) (Figure 3). Summing the values of each column of the ASC and dividing by 10 gave an averaged figure representing agreement among all raters by eye movement and across all words spelled by a particular student. These second level agreement measures were recorded in the last row of the ASC.
Figure 3. Agreement by Subjects Chart
(agreement by subject, word, and eye movement)
Step 3 - Determination of Agreement by Eye Movement, across Subjects and Words

The purpose of this step was to determine agreement among raters across all subjects and words, that is, with respect to each eye movement alone. In order to generate these values, second level measures recorded on the 10 ASCs were compiled on a fourth chart, the Overall Agreement Chart (OAC) (Figure 4). Summing each column of the OAC and dividing by 10 gave a third level agreement measure—an averaged figure which represented, for the 1,500 data points corresponding to each of the seven eye movements, the degree to which raters agreed.

Step 4 - Determination of a Grand Mean

Determination of a grand mean was done by summing the seven third level measures obtained in step 3 and dividing by seven. This averaged figure represented the overall measurement of agreement among all raters with respect to the 10,500 data points obtained in the study.

Step 5 - Computation of a Z Score and Statistical Comparison

A Z score was computed \((Z = (X - \bar{X}) / S)\) where \(X\) was a given raw score, \(\bar{X}\) was the raw score mean and \(S\) was the standard deviation of the raw scores). This allowed a computation of the number of standard deviations above and below the grand mean for each eye movement.

The analysis process led finally, in Step 5, to an indicator of inter-rater consistency across the entire project. The flow chart
<table>
<thead>
<tr>
<th>Subject</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M.E.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. Overall Agreement Chart (agreement by subject and eye movement)
on the following page (Figure 5) presents a second view of the data collection and analysis process.

It should be noted that the purpose of this research was to develop methodology that would allow the examination of the level of agreement of certified raters as they viewed eye movements. No attempt was made to determine an acceptable level of agreement.

**Anticipated Findings**

1. There would be a minimum M.E. score (see p. 95) of 95 among all raters on first observed Eye Movement of each video tape subject on each word.

2. There would be a minimum M.E. score of 70 among all raters on second observed Eye Movement of each video tape subject on each word.

3. There would be no greater than an M.E. score of 50 among all raters for the remainder of all observed Eye Movements (those Eye Movements for which at least one rater has not noted "NA").

4. There would be an inverse relationship between measures of agreement and number of observed responses. That is, as the number of observed responses increased there would be a decline in values of agreement measures.

5. There would be a decline in agreement measure values across the experiment. That is, measures of agreement would be higher among raters for observations of initial as compared to later video tape subjects.
Data recorded on the IRC (Fig. 1)  
Eye movements of 10 video tape subjects rated by 15 certified NLP raters

Bottom row of the ARC (Fig. 2)  
Agreement of all raters by eye movement computed ...
a) for each word spelled by each subject

Bottom row of the ASC (Fig. 3)  
b) across all words spelled by each subject

Bottom row of the OAC (Fig. 4)  
c) across all subjects and words

Grand mean computed—agreement of all raters across all data points

Z score computed to analyze distribution of deviations

Basic projection of inter-rater consistency

Figure 5. Flow Chart of Analysis Process
CHAPTER IV

RESULTS

This chapter is divided into three sections. The first section deals with the two phases of the data collection process: the videotaping of students spelling 10 different words and the eye movement rating process undertaken by 15 certified Neuro-Linguistic Programming Practitioners. The second section provides an analysis of the data in accordance with the method outlined in Chapter III, with the five anticipated findings listed in Chapter III also discussed in this section. The final section of the chapter will summarize the results of the study.

Data Collection

Videotaping of Students

A total of 18 Oregon State University students enrolled in psychology classes volunteered to be individually videotaped from the eyes upward while spelling 10 different words. Of the videotaped results, five students were rejected either because the student turned his or her head inappropriately during the session or because sound quality on that part of the videotape was judged inadequate. Three of the remaining 13 students were chosen to be viewed by Practitioners in practice runs, while the other 10 students were chosen to be viewed by the Practitioners for the purpose of gathering data for this study. The video tape subjects were simply used in the sequence that they were taped in. This
group of 10 students consisted of seven males and three females. As mentioned in the section Selection of Videotaped Students in Chapter III, each student completed an Informed Consent Form prior to the videotaping (Appendix), with handedness determined by a question on the form and by observation of which hand the student used to fill out the form. However, it was later determined that one of the students, a male, was, in fact, ambidextrous. This student's eye movements were also rated by the 15 certified NLP Practitioners, but, since handedness is one variable which, according to NLP theory, affects eye movement direction, ratings of that student were omitted from the analysis.

Each of the remaining nine students spelled at least seven of the 10 words correctly. Three students spelled all 10 words correctly; one spelled nine words correctly; three spelled eight words correctly; and one spelled seven words correctly.

The word, "pharaoh," presented a problem. Several of the students apparently took this word, when spoken, to be the word "farrow" and spelled it, correctly, as such. The Practitioners observed and rated the eye movements of students spelling this word in the same way that they did those of students spelling the other words. Moreover, since any confusion concerning this word had no bearing upon the inter-rater consistency of the Practitioners in their recorded observations of students spelling the word, the recorded observations of the raters were included within the data analysis.
Rating Procedure

Sample of Raters. Fifteen certified Neuro-Linguistic Programming Practitioners took part in this study. Each of the 15 viewed the videotape individually with only the researcher otherwise present in the room. The viewing took place in three different cities on the West Coast over the course of 10 days.

Six Practitioners were male and nine were female. Their ages ranged from 25 to 51 for the males, with the median age being 38, and 31 to 48 for the females, with the median age being 35. The Practitioners held degrees ranging from solely a high school diploma, to a Ph.D., including degrees from higher institutions located in the Eastern and Western United States and Europe. The Practitioners fell into three basic categories occupationally: six in education, five in business, and four in therapy.

The Practitioners indicated that they had used Neuro-Linguistic Programming for as little as six months and for as long as eight years, with the median time being two years and 10 months. The fields indicated by the Practitioners as areas in which they used Neuro-Linguistic Programming, and the number for each field are as follows (many reported more than one field):

- Sales: eight
- Education: nine
- Business: nine
- Therapy: eight
- Other: seven
There is reason to believe that the motivation of the raters was generally high. Several drove out of their way to attend the rating session and all gave up weekend time for purposes of the study. All wanted to know the results of the study.

The Procedure. To familiarize Practitioners with the research setting and procedure, each was given a minimum of three practice runs which consisted of observing three of the videotaped students each spelling the 10 words. If continued practice was requested by a Practitioner, the Practitioner was allowed to repeat viewing these three students until he or she indicated being comfortable with the procedure. At that point the rating process began.

During the familiarization period each rater practiced recording observed eye movements. Recording consisted of the rater drawing, on a blank sheet paper, consecutive arrows indicating the directions of observed eye movements for a particular subject and word. Directions were indicated by arrows pointing up, down, to the left or to the right, and with a circle signifying eyes being in the centered position. Arrows were drawn consecutively either from top to bottom or from left to right on the paper. Eye movement notations—A, V, K, and C—corresponding to the directions indicated by the arrows were immediately recorded on the rater's Individual Rater Chart by the researcher after each spelling task. This method was preferable to having raters themselves write eye movement notations on the Individual Rater Chart, for it did not require the rater to shift his or her eyes from the screen to the paper in order to record observations. The method was judged fully satisfactory,
with no ambiguities arising in determining meanings of the raters' notations. If one were going to rate an entire strategy via eye movements, then a mechanical eye tracking device coupled with a video taping process might have to be developed.

Raters were instructed to rate all observed eye movements of a particular videotaped student spelling a particular word, beginning with the point when they thought the student's spelling strategy for that word had begun and ending with the point where the student said the last letter of the word. At times the number of eye movements observed by a rater was greater than the seven allowed for on the Individual Rater Chart. In such cases observed eye movements beyond the seventh were ignored.

After the rating process of each Practitioner was completed, comments were solicited. Some raters indicated that some of the videotaped students' eye movements were more difficult to rate than others. For example, several raters commented that the eyes of students number three and six were off center, making it more difficult to rate these two. It was suggested by four raters that a short period of time for calibration, i.e., observation of each videotaped student before the spelling tasks began, could have helped remedy this problem. Two raters said that some students moved their heads (although they had been instructed not to), making rating more difficult. Other notable comments, each made by one or two Practitioners, included the following:

1. "One subject traversed so quickly through the eye movements that it was difficult to distinguish them."
2. "It would be easier to see the eye movements if one could see the whole person."

3. "Keeping one's mind on the task was difficult because it was hypnotic." (That is, the rater would continue looking at the television display, but his/her mind would drift off to something else.)

4. "It would be easier to rate eye movements if some of the students did not close their eyes; blinking may give the illusion of movement."

5. "The spelling words formed a good range of difficulty."

In addition, three raters made statements during the process which indicated that they were tired of looking at the screen. Several of these comments are useful in determining ways in which the research procedure could be improved. They will be discussed in greater detail in Chapter V.

Analysis of Data

Computation of Z-scores

Step 1 - Determination of Agreement by Subject, Word, and Eye Movement. After transferring data from the Individual Rater Charts to the 15 All Raters Charts (ARC), with one ARC pertaining to each subject, the percentage occurrence of each rater notation for each spelling task was determined. Application of the formula

\[ M = \frac{v^2 + c^2 + a^2 + k^2 + 'NA'^2}{100} \]

to each column of each ARC then yielded an agreement measurement, M, reflecting consistency among raters in their observations for each of seven possible eye movements associated with each spelling task.
M was determined for those Eye Movements\(^1\) where at least one rater noted the occurrence of an eye movement.

**Step 2 - Determination of Agreement by Subject and Eye Movement, across Words.** Transferring M values from each ARC to a corresponding Agreement by Subjects Chart (ASC), with each ASC pertaining to one subject and each column of the ASC pertaining to a single Eye Movement, allowed development of a more comprehensive measure of agreement, $\overline{MSE}$. Since each column of the ASC contained the values of M for a single Eye Movement of a particular subject across all words spelled by that subject, summing the values of each column and dividing by 10 gave the mean value of M ($\overline{MSE}$), representing agreement among all raters by Eye Movement by subject and across all words.

The 10 ASCs indicate that some subjects were more difficult to rate than were others. The values of $\overline{MSE}$ for videotaped Students 3 and 7, in particular, were generally lower than those for the other students. With respect to Student 3, it was pointed out in the

---

\(^1\)An important terminological distinction should be made. Whereas the term 'eye movements' (in all lower case letters) refers to the particular eye movements displayed by individuals performing particular spelling tasks, the term 'Eye Movements' (with first letters capitalized) refers to the ordered set (from 1 through 7) of schema wherein observed eye movements are recorded and ordered. Thus, 'Eye Movement 7,' for example, refers to a schema wherein are recorded particular eye movements observed by raters as being the seventh occurring within a particular spelling task; moreover, since the schema exists even if there is nothing to fill it out, it makes sense to speak about Eye Movement 7 even in cases where no seventh eye movements have been observed.
previous section that some raters expressed having difficulty rating
this subject because the subject's eyes were "off center."

Step 3 - Determination of Agreement by Eye Movement, across
Subjects and Words. Transfer of the second level agreement
measures, \( M_\text{SE} \), to a single Overall Agreement Chart (OAC) enabled
the calculation of a measure of agreement among raters by Eye
Movement and across all subjects and words. This third level
measure, \( \bar{M}_E \), was determined for each Eye Movement by summing the
values of the column and dividing by nine (for nine videotaped
students). Exceptions to this occurred for the sixth and seventh
Eye Movements, represented by columns 6 and 7. Since there were no
recorded observations for the sixth Eye Movement on all words
spelled by Subject 8 or for the seventh Eye Movement on all words
spelled by Subjects 1, 2, 8, 9, and 10, there were no entries to be
made in the boxes corresponding to these subjects and Eye Movement.
The totals for columns 6 and 7 were therefore divided only by the
number of entries in each column—eight and four respectively.
Third level measures, \( \bar{M}_E \), are shown in the bottom row of the OAC,
which is presented in Figure 4. Detailed discussion of \( \bar{M}_E \)
measures shown in the OAC occurs below, under the heading of
Anticipated Findings.

Step 4 - Determination of a Grand Mean. The grand mean, \( \bar{M} \),
was determined by adding the seven \( \bar{M}_E \) measures and dividing by
seven. The result of this calculation showed that \( \bar{M} = 67.80 \).
At this point in the analysis, there were four measures of agreement—\( \bar{M}, \overline{\text{MSE}}, \overline{\text{M.E.}}, \) and \( \bar{M}. \). For the sake of clarity, a short description of these is given in Table 1 below.

Table 1. Measures of Agreement

\[
\begin{array}{|c|}
\hline
\text{MSEW} = \text{Raw Score} \\
\hline
\text{MSEW} & \text{Initial agreement measure reflecting agreement among raters for subject S, eye movement E, and word W, where} \\
 & \text{Subject} \ S = 1, 2, \ldots, 10 \\
 & \text{Eye Movement} \ E = 1, 2, \ldots, 7 \\
 & \text{Word} \ W = 1, 2, \ldots, 10 \\
\hline
\text{\overline{MSE}} & \text{The mean value of agreement scores averaged over all words for Subject S and for Eye Movement E.} \\
\hline
\text{\overline{M.E.}} & \text{The value of \overline{MSE}, averaged over all subjects for Eye Movement E.} \\
\hline
\text{\overline{M}} & \text{The grand mean—the mean of the seven \overline{M.E.} measures, reflecting agreement among raters over all eye movements, spelling words, and subjects.} \\
\hline
\end{array}
\]
Step 5 - Computation of a Z score. Statistical analysis of a set of quantities normally involves two basic kinds of measures: measures of central tendency and measures of variability. In the present case, the analysis to this point had concentrated on measurements of the first type, culminating in determination of the grand mean in Step 4. The analysis then moved to the second kind of measure, that of variability. This required determining the distance from the grand mean of each of the M.E. scores (representing each of the seven Eye Movements). This was accomplished by calculating first the standard deviation for the set of M.E. quantities as a whole and then, on this basis, a Z score for each of the M.E. values individually.

The standard deviation of a set of quantities is, as the name implies, a measure of how great is the variance or deviation of the quantities on the average from the mean value for the set. Determination of the standard deviation, then, allows one to determine, for any particular member of the set, whether it is above or below the mean and by how much. The standard deviation for a set of quantities is determined according to the following formula:

\[ s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \]

where

- \( x_i \) is the \( i \)th number of the set
- \( x \) is the mean of the \( n \) numbers of the set

and where

- \( n \) is the number of members in the set.
In the present case, where the seven M.E. values ranged from 52.47 to 77.07 and where the grand mean equaled 67.80, the standard deviation, s, was found to be equal to 9.67.

On the basis of the standard deviation, a Z score was then calculated for each of the M.E. values. Since the Z score is a measure of how many standard deviations each member of a set of quantities falls above or below the mean for the set, it allowed, in the present case, the M.E. values to be directly compared to one another with respect to variance from the mean. The Z score was computed according to the formula

$$ Z = \frac{X - \bar{X}}{s} $$

where

- X is a member of a set of independent quantities,
- $\bar{X}$ is the mean of those quantities, and
- s is the standard deviation of the quantities.

In the present case the quantities of interest were those designated M.E., where $\bar{M}...$, the grand mean, was the mean of those quantities and where s was the standard deviation of the M.E. values. Substituting these notations into the above equation yielded

$$ Z = \frac{\overline{M.E.} - \bar{M}...}{s} $$

This formula allowed calculation of a Z score for each measure $\overline{M.E.}$. Table 2 shows these values of Z.
Table 2. Z scores for Seven Eye Movements

<table>
<thead>
<tr>
<th>Eye Movement</th>
<th>M.E.</th>
<th>Grand Mean</th>
<th>Standard Deviation</th>
<th>Z Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>77.07</td>
<td>67.80</td>
<td>9.67</td>
<td>.96</td>
</tr>
<tr>
<td>2</td>
<td>52.47</td>
<td>67.80</td>
<td>9.67</td>
<td>-1.59</td>
</tr>
<tr>
<td>3</td>
<td>56.49</td>
<td>67.80</td>
<td>9.67</td>
<td>-1.17</td>
</tr>
<tr>
<td>4</td>
<td>67.46</td>
<td>67.80</td>
<td>9.67</td>
<td>-.04</td>
</tr>
<tr>
<td>5</td>
<td>71.49</td>
<td>67.80</td>
<td>9.67</td>
<td>.38</td>
</tr>
<tr>
<td>6</td>
<td>74.94</td>
<td>67.80</td>
<td>9.67</td>
<td>.74</td>
</tr>
<tr>
<td>7</td>
<td>74.68</td>
<td>67.80</td>
<td>9.67</td>
<td>.71</td>
</tr>
</tbody>
</table>

As pointed out above, the Z scores shown in Table 2 indicate how far the M.E. values, i.e. the mean rating for each Eye Movement, diverged from the grand mean. The value of .96 for the Z score associated with Eye Movement 1, for example, indicated that the mean ratings for this Eye Movement fell .96 standard deviations above the grand mean. The value of -1.59 for the second Eye Movement indicated that the mean ratings for this Eye Movement fell 1.59 standard deviations below the grand mean.
Anticipated Findings

The five anticipated findings of this study fell into two groups. In general, the first three anticipated findings predicted greater agreement among Practitioners in rating early eye movements than in rating later eye movements. Anticipated findings 4 and 5 predicted inverse relationships between (1) rater agreement and number of total observed eye movements for a videotaped student, and (2) rater agreement and the order in which the student was viewed.

Anticipated Findings 1, 2 and 3. Each of the first three anticipated findings was not supported in this study. Anticipated Finding 1 stated: There will be a minimum agreement of $\overline{M.E.} = 95$ among all raters on first observed eye movement over all videotaped students and words. As indicated by Table 2, this mean value of $\overline{M}$ for Eye Movement 1, indicated by the measure $\overline{M.E.}$, was, although higher than for any other Eye Movement, lower than predicted by Anticipated Finding 1. $\overline{M.E.}$ in this case equaled 77.07.

Anticipated Finding 2 stated: There will be a minimum agreement of $\overline{M.E.} = 70$ among all raters on second observed Eye Movement over all videotaped students and words. In this case, the actual value of $\overline{M.E.}$ was 52.47, again lower than the value predicted.

Anticipated Finding 3 stated: There will be no greater agreement than $\overline{M.E.} = 50$ among all raters for the remainder of all observed eye movements over all videotaped students and words. In this case the actual value obtained was substantially greater than the predicted value. The mean value of $\overline{M.E.}$ for the third through
seventh Eye Movements, determined by adding the values of $\bar{M}_E$ for these Eye Movements and dividing by five, was 69.01. Thus, Anticipated Finding 3 was also not supported.

In understanding these results, it is helpful to recall, as mentioned above, that the first three anticipated findings expressed the general expectation that there would be higher overall agreement among raters' observations of earlier eye movements than of later eye movements. Inspection of Table 2 shows that this general expectation was at least partially upheld by the results of the experiment, since overall agreement on Eye Movement 1 was greater than for any other Eye Movement, with a mean of 77.07 and a Z score of .96. However, the next highest agreement was for Eye Movement 6, with a mean of 74.94 and a Z score of .74, and in general, the expected decline in rating agreement over the seven Eye Movements did not occur. Though overall agreement decreased substantially from Eye Movement 1 to Eye Movement 2, it then increased rather steadily over the remaining Eye Movements, only showing a slight decline from Eye Movement 6 to 7.

This increase in agreement among raters over the later eye movements is largely explained by what can be termed the "NA-effect." The NA-effect is that effect upon agreement measures which occurred when some raters noted the occurrence of an eye movement corresponding to a particular rating box while others did not. When only one or a few raters noted such an occurrence and most did not—with 'NA' thereby indicated for that rating box—a high agreement score resulted. For example, where one rater noted an eye movement
and 14 raters noted none, the value of M, the initial agreement measure, was 87.11. Where two raters noted an eye movement and 13 noted none, the value of M was 75.14. Both figures were considerably above the grand mean. The proportion of such cases in relation to total number of observations recorded for the Eye Movement was substantially higher for later Eye Movements than for earlier ones.

That an NA-effect would be present in this study was anticipated. In fact, the effect was present in data for all Eye Movements except the first, since only for the first Eye Movement did raters agree, for every spelling task, that some eye movement position occurred. The presence of the NA-effect for all other Eye Movements was simply the reflection of the fact that agreement and disagreement were not only about what was observed, but also about whether anything was observed. By stipulating that cases wherein all raters agreed that no eye movement occurred would not be used in the analysis (Chapter 3, p. 76), one limitation of the NA-effect was achieved. Moreover, this seemed to be the only legitimate limitation available prior to the study, since, in the absence of data, any other limitation would have been arbitrary. This will perhaps be seen upon recognizing that any of a large number of possible sets of data would have resulted in the NA-effect being negligible for later Eye Movements. As it turned out, given the actual data collected, it was not negligible for the later Eye Movements. With this data now available, recommendations for dealing with the NA-effect in future similar studies are possible, and several will be made in Chapter V.
Anticipated Findings 4 and 5. While both development of a Z score and testing of the first three anticipated findings required determining overall means for each Eye Movement, testing of anticipated findings 4 and 5 called for determining an overall agreement mean for each subject. Thus it is appropriate to deal with these two under one heading. Anticipated Finding 4 stated: There will be an inverse relationship between overall agreement measures and number of observed responses for individual videotaped students. In order to test this anticipated finding, an overall agreement measure was first determined for each subject. Since each row of the Overall Agreement Chart (Fig. 4, p. 83) pertained to one videotaped student, overall agreement measures by subject were determined by adding up the values of \( \bar{N}_{SE} \) in each row of the OAC and dividing by the number of Eye Movements for which an entry was made in that row. This gave a set of values, \( \bar{N}_S \)\ldots, which were the overall agreement measures by subject. Next, the number of eye movement notations made by the 15 Practitioners were totaled for each subject. Finally, these totals were compared with the \( \bar{N}_S \)\ldots values in order to determine if the predicted correlations occurred. These values and comparisons are presented in Table 3.

Table 3 shows that, although total number of recorded observations per subject and overall agreement by subject were not perfectly correlated in the predicted inverse relationship, there was a clear trend indicating that a correlation did exist. Videotape Subjects 5 and 3 can be considered as cases where, compared to other subjects, a high number of total observations were
<table>
<thead>
<tr>
<th>Videotaped Subject</th>
<th>Total Number of Recorded Observations</th>
<th>Overall Agreement by Subject (Ns..)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>741</td>
<td>53.16</td>
</tr>
<tr>
<td>3</td>
<td>611</td>
<td>52.42</td>
</tr>
<tr>
<td>7</td>
<td>493</td>
<td>56.35</td>
</tr>
<tr>
<td>1</td>
<td>421</td>
<td>70.72</td>
</tr>
<tr>
<td>2</td>
<td>340</td>
<td>73.30</td>
</tr>
<tr>
<td>4</td>
<td>325</td>
<td>76.37</td>
</tr>
<tr>
<td>10</td>
<td>322</td>
<td>73.01</td>
</tr>
<tr>
<td>9</td>
<td>274</td>
<td>75.44</td>
</tr>
<tr>
<td>8</td>
<td>221</td>
<td>80.09</td>
</tr>
</tbody>
</table>
recorded, while Subjects 7 and 1 can be considered to fall within a high midrange, Subjects 2, 4, and 10 to fall within a low midrange, and Subjects 9 and 8 to be cases with relatively low numbers of total observations made. Mean values of $\bar{N}_g$ for these four groups were 52.79, 63.54, 74.23, and 77.77 respectively. Anticipated Finding 4 was therefore supported by the results of this study, although the predicted inverse correlation was not perfect.

Anticipated Finding 5 stated: There will be a decrease in percentage of agreement across the experiment. Since all Practitioners viewed the videotaped students in the same order, testing of this anticipated finding required determining whether the overall agreement measures $\bar{A}_g$ showed a decline from earlier to later subjects. Table 4 shows this comparison, where videotaped student number is identical with videotaped student order of viewing.

The comparisons shown in Table 4 offered no support for Anticipated Finding 5. The mean of the last two overall agreement measures, 74.23, for example, was higher than the mean of the first two overall agreement measures, 72.01.

It should be noted, however, that the procedure of this study did not provide a proper test of the fifth anticipated finding. Variables having nothing to do with viewing order could have resulted, for example, in videotaped Student 10 being an easier subject for Practitioners to rate than was videotaped Student 1. In such a case, since Student 10 was viewed last by all Practitioners,
Table 4. Relation of Overall Agreement by Subject to Subject Order of Viewing.

<table>
<thead>
<tr>
<th>Videotaped Student</th>
<th>Overall Agreement by Subject (Rs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>70.72</td>
</tr>
<tr>
<td>2</td>
<td>73.30</td>
</tr>
<tr>
<td>3</td>
<td>52.42</td>
</tr>
<tr>
<td>4</td>
<td>76.37</td>
</tr>
<tr>
<td>5</td>
<td>53.16</td>
</tr>
<tr>
<td>7</td>
<td>56.35</td>
</tr>
<tr>
<td>8</td>
<td>80.09</td>
</tr>
<tr>
<td>9</td>
<td>75.44</td>
</tr>
<tr>
<td>10</td>
<td>73.01</td>
</tr>
</tbody>
</table>
any tendency for rater agreement to be less for the final subject than for the first would have been more difficult to detect. Randomizing of videotaped students for viewing by the Practitioners, where different Practitioners viewed the students in different orders, may have lessened possible effects of other relevant variables. Such randomizing would have required either (1) the use of 10 video tapes, each containing the taped results for a single subject, or (2) the use of 10 video tapes, each containing the record of the 10 subjects in a unique order. The first method would have necessitated switching video tapes nine times while the study was being conducted with each Practitioner. This would have been an unacceptable procedure, for it would have likely broken the concentration of the Practitioner nine times, would have added an element of disruption throughout the study, would have resulted in the researcher being unable to remain mostly in the background while the study was being conducted, and might have resulted in a substantial distraction increasing possibility of error. The second method would have required the services of a professional videotape editor in order to create 10 videotapes of the students in 10 different orders. This was far outside the budget of the present research.

Since the present study provided no adequate test for Anticipated Finding 5, it must be considered as neither supported nor unsupported by this research.
Summary of Results

Fifteen certified Neuro-Linguistic Programming Practitioners viewed a videotape of 10 Oregon State University students who had been videotaped from the eyes upward while spelling 10 words. Practitioners observed eye movements of the videotaped students and these were recorded on Individual Rater Charts, with up to seven possible eye movement notations recorded for each spelling task.

Subsequent analysis of these notations provided overall agreement figures among the raters by Eye Movement, a grand mean which reflected agreement among raters over all eye movements and spelling tasks, and Z scores associated with each of the overall ratings by Eye Movement. Greatest agreement among raters was found for the first observed Eye Movement, where overall agreement exceeded the grand mean. Overall agreement for the second observed Eye Movement fell below the grand mean. Overall agreement for subsequent Eye Movements generally increased through Eye Movement 6 and then showed a slight decrease for Eye Movement 7. This increase in agreement for later Eye Movements was largely attributed to what is termed the "NA-effect," which can result in high agreement scores in cases where only one or a few raters note an eye movement while all other raters note no eye movement, and can result in inflated overall agreement scores when the proportion of such cases related to a particular Eye Movement is high in relation to total number of recorded observations associated with the Eye Movement.

Anticipated findings 1, 2 and 3 were not supported by this study, with overall rater agreement lower than predicted for Eye
Movements 1 and 2 and higher than predicted for remaining Eye Movements. Anticipated Finding 4 was confirmed by the study, with total number of eye movements observed by all raters for a particular subject found to be inversely correlated with overall agreement among raters in rating that subject. Anticipated Finding 5 was neither supported nor unsupported by the study, since, due to restraints, an appropriate methodology for testing this prediction was not available.
CHAPTER V

SUMMARY, DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS

This chapter is divided into three sections. A summary of the study is presented in the first section, followed by a discussion of several major aspects of the study in the second section. The final section includes both conclusions and recommendations for further research.

Summary

The purpose of this study was to develop a methodology to code and quantify eye movement patterns as a function of a NeuroLinguistic Programming (NLP) strategy elicitation process.

A secondary purpose was to establish a methodology that would identify the degree of consistency among certain certified NeuroLinguistic Programming Practitioners while they were rating eye movements included within a given strategy.

Background

Neuro-Linguistic Programming (NLP) offers a model of human behavior, where the term 'behavior' refers both to overt, publicly observable behavior and to so-called inner processes such as thinking and imagining. According to the NLP model all human behavior can be understood as "systematically ordered sequences of [internal and external] representations" (Dilts et al., 1980). Behavior which aims for an appropriate outcome, such as spelling,
reading, and negotiating, can therefore be understood as consisting of such sequences, and these are termed 'strategies.'

The model further asserts the existence of relationships between certain overt behavioral items, or cues, and inner representations. Understanding these relationships enables one to determine, through observation of overt behavior, facts about nonovert behavior. Out of this arise two claims that constitute the practical importance of the NLP model: (1) it is possible to understand the strategies employed by people in performing tasks, and (2) through use of various techniques, it is possible to alter people's strategies to lead to more successful outcomes.

The bulk of NLP-related research to date has dealt with two behavioral cues alleged to provide knowledge of internal processing -- eye movements and predicate usage. A substantial portion of that research has been concerned with what these cues are alleged to indicate about internal processing, namely, whether the processing is auditory (internal representations of sounds), visual (internal visualization) or kinesthetic (feelings, moods, and internal kinesthetic sensations). However, an overview of this research shows inconclusive and often contradictory results. Such discrepancies may largely be due to methodological problems present in some studies. A number of these were pointed out in the Review of Literature in Chapter II.

The methodological issues involved are, in fact, so numerous and serious that there seems ample justification for a study which deals primarily with methodology. The present study undertook to do
this, while concentrating on eye movements as behavioral cues. In doing so, eye movements were not investigated in an attempt to test some feature of the NLP model. Rather, the rating of eye movements, i.e., the process of determining which eye movements are displayed by a person, was singled out for investigation as an area which has been prone to methodological confusion. By isolating this process and determining what problems and possible solutions arise in investigating eye movements, the study attempted to lay groundwork for future NLP-related research involving eye movements as behavioral cues.

In pursuing this objective, the study was concerned specifically with agreement among raters when observing eye movements. In fact, measuring inter-rater agreement is a central aspect in the methodology of rating eye movements, since any substantial rater disagreement in observations of an individual seriously weakens inferences drawn from those observations. Moreover, since NLP techniques based on eye movement ratings are currently being used by many individuals licensed by NLP training establishments, it is appropriate to ask the question of how well those raters agree among themselves. For this reason, the raters used in this study were members of the population of certified Neuro-Linguistic Programming Practitioners.

Methodology

The sample for this study consisted of 15 currently certified Neuro-Linguistic Programming Practitioners residing on the West
Coast of the United States. The researcher met with each Practitioner individually in one of three West Coast cities where the Practitioner observed a videotape of 10 Oregon State University students spelling 10 words. The students were videotaped from the midpoint of the nose upwards, and the Practitioners rated eye movements of each student spelling each word.

Results of these observations were first recorded on an Individual Rater Chart using the V-A-K-C system, where V indicated upward eye movement, A indicated eye movement to either the left or right, K indicated downward eye movement and C indicated eyes being in the centered, defocused position. Data were then transferred to an All Raters Chart, allowing comparison of the 15 raters for each spelling task. An agreement measure, M, was determined for each spelling task, by Eye Movement, based upon the percentage occurrence of each notation. Determination of the mean of M measures, by Eye Movement and across all words spelled by a videotaped student, allowed a second level agreement measure, \( \bar{M}_{SE} \), to be determined. A third level measure, \( \bar{M}_{E} \), representing overall agreement among raters by Eye Movement, was determined by calculating the mean of all second level measures across all videotaped students. A grand mean, \( \bar{M}_{...} \), representing overall agreement across all Eye Movements, was then calculated, and Z scores were determined for each Eye Movement based upon the \( \bar{M}_{E} \) measures and the \( \bar{M}_{...} \). This allowed overall agreement measures by Eye Movement to be compared with the grand mean. A second third level measure, \( \bar{M}_{S...} \), was also determined.
for the purpose of testing the fourth anticipated finding of the study.

Results

Five findings were anticipated in this study.

Actual Finding 1: It was anticipated that there would be a minimum mean M value of 95 for first observed Eye Movement across all raters, videotaped students and words. This prediction was not upheld by the data. Although the mean M value (M₁E.) for first observed Eye Movement was higher than those for any of the other observed Eye Movements, it was, at 77.07, lower than predicted.

Actual Finding 2: It was anticipated that there would be a minimum mean M value of 70 for second observed Eye Movement across all raters, videotaped students and words. This anticipated finding proved incorrect in light of the data, since the value of M₂E. in this case, 52.47, was again lower than predicted.

Actual Finding 3: It was anticipated that across the remainder of Eye Movements and across all raters, videotaped students and words, mean M value would not exceed 50. This prediction was not upheld by the data. The mean M value in this case was identical with the mean M₂E. value for the final five Eye Movements, and was determined to be 69.01.

Actual Finding 4: It was anticipated that there would be an inverse relationship between measures of agreement and total observed eye movements for each videotaped student. This
anticipated finding was upheld by the data. Although the correlation was not perfect, there did exist a clear trend for agreement measures to decrease as total number of eye movements observed for a student increased.

Actual Finding 5: It was anticipated that there would be a decline in agreement scores across the experiment. This prediction was not adequately tested in this study and thus was considered as neither supported nor unsupported by the data.

In general, results showed that highest overall agreement among raters occurred for first observed Eye Movement. Agreement declined considerably for second observed Eye Movement, then rose consistently, showing only a slight decline from sixth to seventh observed Eye Movement.

Discussion

This study was concerned primarily with developing a methodology for rating eye movements in NLP-related research. This section, therefore, will deal mainly with the methodology used, pointing out where it was satisfactory and where it was not, and discussing possible alternatives in the latter case. The issues to be discussed are the videotaping procedure, the rating procedure, the scoring procedure, and the use of certified NLP Practitioners as raters. Based largely on this discussion, conclusions will be drawn and recommendations made in the final section.
Videotaping Procedure

In the Review of Literature attention was paid to the unnaturalness of the experimental situation in several previous NLP-related studies. In the present investigation there were unnatural elements also. These included the operation of a video camera while the students were spelling, along with the more general fact that they were asked to spell words aloud in an unfamiliar and contrived situation. These factors could be expected to have produced a certain amount of performance anxiety in the students, leading to behavior—including eye movement patterns—which they would not otherwise have exhibited while spelling words aloud.

Given the purposes of the present study, however, since these involved only the methodology of rating eye movements, this element of unnaturalness did not present a problem. At the same time, it should be remembered that any study concerned with the use of eye movement observations to test the NLP model ought to pay close attention to the element of naturalness. In the final section several recommendations will be made for dealing with this element, especially for studies which use videotaping procedures and also seek to test the NLP model.

Several raters said that it would have been easier to discern eye movements if the students' entire faces had been videotaped. Since changes are sometimes easier to detect if the object observed is seen in some larger context, this objection may have been well-taken. At the same time, the objection should be weighed against the fact that, according to NLP theory, eye movements are only one
relevant behavioral indicator among others, including head and shoulder movement, breathing rate, and posture, etc. Thus, in studies where only eye movements are being rated, videotaping a larger segment of the person may run the risk of confusing the issue by presenting any of a number of other behavioral indicators. Especially when raters have been trained in NLP techniques, attention paid to these other indicators might tend to break the raters' concentration on eye movements. On the other hand, videotaping a larger segment of the individual would provide a more natural situation of observation from the rater's point of view, and such simulation could be important in some future studies. Moreover, for any studies which seek to deal with the full strategies employed in, say, spelling, videotaping the whole person or at least the upper body would seem to be correct, since in this way a much wider range of behavioral indicators could be observed.

Rating Procedure

Before the rating process began, a period of time was devoted to practice runs for the raters. Practitioners were given as much time as they wanted to view three of the practice videotaped students and were able to view them at both normal and slow speeds. This could be seen as giving the raters an unwarranted advantage which they would not have had in a real-world rating situation. The researcher's view, however, is that the practice runs are more correctly seen as having been a way to partially counteract some obvious disadvantages presented by the research setting which raters
would not have had to face in a real-world rating situation. A period of time for practice, for getting used to the research environment, and for making sure that they understood fully what was expected of them, was held to be clearly necessary for the raters. Moreover, in a real-world situation a rater would not normally begin rating eye movements until he or she had gained some familiarity with the situation. The practice time alloted raters seems to have been warranted and by their own comments sufficient.

Several raters complained of having a difficult time rating two of the videotaped students because their eyes were "off center"—that is, in what would normally be taken to be the centered, defocused position, the directionality of the pupils of these students' eyes was skewed from the straight-ahead direction. These same raters suggested that a short period of time for observing each videotaped student before he or she began to spell would have been helpful, allowing them to get used to any peculiarities associated with the directionality or position of a student's eyes. This suggestion seems reasonable, especially upon realizing that in a normal rating situation the rater would likely have the opportunity to become aware of such peculiarities and would take them into account during the rating process.

The method of recording observations, with the rater drawing directional arrows on a blank sheet of paper and the researcher recording those observations on an Individual Rater Chart, was also practiced during the trial runs. The researcher noted no ambiguities or other confusions in understanding the meanings of the
raters' notations, and, by all indications, accurate data resulted from application of the method.

It is relevant to note that the ease and effectiveness of using this method seemed to largely reside in the fact that the sheets of paper upon which raters made their notations were blank. Requiring the use of any form or chart, even with the simplest internal structure, would have made it more difficult for the rater to make notations while at the same time continuously viewing the screen. The rater was able, with the blank sheet of paper, to use his or her tactile sense to make sure that notations were being made on the paper, while kinesthetic awareness in general enabled the rater to make sure that notations were made consecutively either across or down the paper.

Several times the researcher was asked by a rater during the rating process whether he or she should rate what could be termed 'minimal eye movements'--i.e., cases where directionality of the eyes appeared to change slightly.

In each case, the rater was told simply to "rate what you see." This was the appropriate answer to such questions since what was of interest in this study was how the Practitioners rated eye movements and how well they agreed among themselves based upon their own standards; for the researcher to have imposed strict standards for determining what constituted an eye movement or a change in directionality would have defeated this purpose.

It should be noted, however, that the issue of minimal eye movements is an important one, especially for any research which
studies eye movements in order to test the NLP model, i.e., how much change constitutes significant, scorable change. More basically, and further clouding the matter, NLP theorists do not seem to have clearly addressed the issue. What does seem clear is that if it turns out that very small changes in eye position are held to be relevant to inner processing according to the NLP model, then rating via videotape, since the tape can be both rerun and run in slow motion, would likely be necessary in order to ensure correct ratings. Possibly even use of eye movement monitoring systems—television based eye trackers—would be required by raters. In any event, what constitutes enough eye position change to count as significant change seems to be a central question awaiting clarification within the NLP model.

A separate but related issue that arose during the course of this study involves the matter of eye position or direction. Though attempts to precisely define eye directionality have been made in NLP-related studies, some, e.g. Johnson's (1983) investigation, seem not to have taken into account possible deviations from the norm which may be exhibited in an individual's typical eye positions and/or movements. In the present study several raters remarked, as was pointed out above, that two of the subjects' eyes were "off-center." This suggests that a wholly adequate treatment of what counts as a scorable eye movement cannot rely entirely upon a model of "normal" eye position and movement; it should recognize and make allowances for individuals whose eyes are, when looking straight ahead, skewed from center, as well as those
whose eyes are skewed with respect to one another in either position, movement, or both (e.g. cross-eyed individuals). This was a problem in that one rater noted a temporarily cross-eyed speller.

These three issues—eyes being "off-center," minimal eye movements and strabismus—are basic matters that have gone unrecognized or unaddressed by some NLP-related eye movement studies whose objectives were much "broader" than those of the present investigation. That such issues did arise and were recognized in the present, "narrow" study helps warrant the view that quite limited investigations are appropriate and needed in NLP-related research. Indeed, very specific, focused, probing studies which are confined to a small area of inquiry are essential in all research fields, since such investigations have the best prospect of identifying, isolating and dealing with "small" issues which prove to be of great relevance. In the field of NLP-related research there is ample reason to suggest that further limited and focused studies are needed to help clarify concepts and sharpen methodologies.

A critical point regarding the agreement measures arising from this study should be taken into account in any similar future studies. This concerns the fact that not only were Practitioners given little guidance in deciding which eye movements were scorable, but it was also left to them to determine when to begin rating eye movements associated with a particular spelling task: raters were simply instructed to start rating when they believed the videotaped student had begun his or her strategy for spelling. The consequence
of this procedure was that agreement measures were composed of two things: agreement about which eye movements occurred and agreement about which were the first relevant eye movements.

A different procedure could have been used, wherein raters would have been instructed to begin rating at a particular point in time, say at the moment the word to be spelled was first pronounced for a student. Agreement scores arising from this procedure would have more nearly reflected agreement just about what was observed. However, as in the case of deciding what counts as a scorable eye movement, this study was interested in the extent to which Practitioners agreed, based upon their own skills and understanding of NLP theory and techniques. Telling raters precisely when to begin rating would have simplified the meaning of the agreement scores and would have likely led to higher scores, but it would have taken out of the raters' hands a decision which they would ordinarily make. As a result, it would have provided less information on how well raters agree among themselves in the world outside this study.

In this context, it is also useful to point out that designating a particular time to begin recording eye movements would have been inappropriate for this study in another sense. In the Review of Literature the problem of deciding which is the first relevant eye movement was discussed, especially in reference to Hernandez' (1981) investigation. As was made clear, this remains an undecided though important question for NLP theory. It is doubtful that the NLP model holds that there is a particular time or eye
movement which should be considered the first relevant eye movement for every person responding. Such a designation in the present case, therefore, would have been arbitrary.

Equipment used in this study was satisfactory, although a few sound difficulties occurred with the videotaping of some of the students. Sound quality for the 13 students selected for the study was, however, sufficiently high. No combination of recording and playback equipment, of course, no matter how high its quality, will provide the clarity of direct observation of a person, but use of the highest quality equipment available is important in a study such as this. Particular attention should be paid to the video equipment, as it is perhaps at this point where the greatest decrease in resolution occurs. Monitor resolution was deemed a very important issue in providing clarity.

Scoring Procedure

The agreement measures developed for this study—\(\bar{M}\), \(\bar{M}_{SE}\), \(\bar{M}_{E}\), \(\bar{M}\), and \(\bar{M}_{S}\)—worked well, with the basic measure, \(\bar{M}\), being sensitive to small differences in agreement. There was a possible drawback to the family of \(\bar{M}\) measures, however, and it was psychological rather than mathematical. This consisted in the fact that the value of an \(\bar{M}\) measure might have appeared to reflect less agreement than actually existed on a particular item. For example, consider a case where 12 raters note a V eye movement and three raters note an A eye movement and thus where 80% of the raters are in agreement. The \(\bar{M}\) value associated with those ratings, however,
is only 68. To counteract any tendency to think of M values as having indicated less agreement than they actually did, it would be helpful to remember that those measures, though based on percentages, were not themselves measures of percentage of agreement. Thus, the overall agreement among raters on the first Eye Movement, \( \bar{M}_1 = 77.07 \), did not reflect 77% agreement, but rather the fact that, on the average, better than 13 of 15 raters, or 86%, agreed in their notations on first observed Eye Movement. The value for the second Eye Movement, \( \bar{M}_2 = 52.47 \), indicated that, on the average, better than 10 out of 15 raters—close to 70%—agreed on the second Eye Movement.

It seems appropriate at this point to answer a question which may arise, namely, why were not simple percentages used to measure agreement instead of developing the family of M measures? The answer to this begins with understanding that the only candidate for using percentages as agreement measurements in this case would be to have taken the notation upon which there was highest agreement for a particular spelling task and Eye Movement and to have called that the percentage of agreement (as was done in the discussion just above). However, it should be noted that tying agreement to such percentages would have failed to adequately take into account the distribution and variance of scores, both of which were relevant to the notion of agreement in these cases. In illustration, consider the following two possible distributions of ratings of 15 raters for a particular spelling task and Eye Movement:
To use the highest occurring percentage as the agreement measure in these cases would result in agreement being the same for both, i.e., 33.3%, whereas agreement is actually clearly lower in the second case due to the wider distribution of noted observations. The M measure, on the other hand, takes account of this difference, yielding a value of 33.3 for the first case and 23.6 for the second.

The NA-effect, since it led to an inflation of agreement scores for later Eye Movements, presented some problems in the quantitative analysis of this study, and it deserves further discussion. As pointed out in Chapter IV, the NA-effect is that effect upon agreement scores which resulted from cases where only one or a few raters noted an eye movement and all others noted none. Although the NA-effect was present for all Eye Movements but the first in the present study, the effect was smaller for the second and third Eye Movements than for the later ones. This was because the proportion of spelling tasks where only one or a few raters noted an observation, compared to total number of tasks where observations were noted, was greater for each of the later Eye Movements than it was for the second or third Eye Movement.
This suggests a way in which future studies might deal with the NA-effect. To explain this approach, it is helpful to first define, with respect to a particular Eye Movement (Eye Movement 1 through 7), the value T as equal to the total number of spelling tasks for which at least one rater notes an eye position change, while the value N equals the total number of spelling tasks where only one or a few raters note an eye position change. Then, by defining a variable V whose value increases as the ratio N/T increases and by subtracting the value of this variable from the mean overall M score (M.E.) for that Eye Movement, increases in the NA-effect could be counteracted.

Aside from effects on scoring, allowing Practitioners to rate up to seven eye movements presented no problems for the present study. It also had the advantage of diverging from the "single eye movement" methodology. As was pointed out in the Review of Literature, it is highly questionable whether any adequate test of the NLP model is provided through observation of single eye movements occurring at specified times (e.g., the first eye movement which occurs after presentation of a stimulus item). On the contrary, the main thrust in both NLP theory and NLP practice seems to involve strategies, and these are alleged to be accessible primarily through series of behavioral indicators. Developing methodologies that examine such series—as the present study has done—will perhaps open the way for more adequate tests of the NLP model.
Allowing Practitioners to rate up to seven eye movements also made clear that some individuals exhibited many more eye movements than others while performing the same spelling tasks. A few Practitioners seemed to have difficulty recording all of the eye movements of these individuals, an impression corroborated by the support given to the fourth Anticipated Finding of the study. This also brings into question the comments of some NLP proponents (see Dilts, 1980), namely, that adequate training in NLP enables an interviewer to observe eye movements of an interviewee in a person-to-person situation well enough to be able to detect strategies. Results of the present study suggested that for some interviewees, this may be more difficult than these proponents have indicated. This may be especially true considering that the interviewer may be attempting not only to observe eye movements, but also to remember those which have already been exhibited and to infer a strategy from that series. If this is indeed a problem, then videotaping arises once again as being an important, perhaps indispensable tool for those who wish to rate strategies.

Use of Certified NLP Practitioners

Because certified Neuro-Linguistic Practitioners are the individuals who make use of NLP theory and techniques in real-world settings, it was of special interest in this study to determine the extent to which members of a sample of this population agreed among themselves when rating eye movements. Since the sample was not a true random sample of the population of certified NLP Practitioners,
any extrapolations from the sample to the larger population must be considered as only suggested and as yet insufficiently supported by the data. Moreover, since it was not a purpose of this study to attempt to determine what is an acceptable level of agreement among raters, only the most general comments can be made about the agreement exhibited among the Practitioners.

Given the above, it is the researcher's view that the inter-rater agreement among the Practitioners was high enough to warrant considering using certified NLP Practitioners in future studies. Also, as mentioned in Chapter IV, interest and a professional attitude toward their role was notable among the Practitioners. Moreover, their comments and observations about the rating process itself represented an important contribution.

The matter of minimal eye movements was perhaps the most significant to come up in discussions with the Practitioners after the rating process was completed. These have been discussed in some detail above, but it is worth noting that agreement among raters would have likely been substantially higher if there had been agreement upon whether minimal eye movements were or were not to be rated. This should be interpreted not so much as weakness in the raters' skills but rather as a lack of clarity in NLP theory about what constitutes a ratable change in eye position.

Another important matter which came up in discussion with Practitioners involved videotaped students whose eyes were skewed from normal, a subject which was also discussed above. Here again,
if more time had been given to Practitioners to become familiar with these subjects, higher scores would have likely resulted.

The question of what constitutes an acceptable level of agreement must wait further studies. It should be pointed out, however, that what turns out to be acceptable will depend greatly upon what is being rated. Replications of the present study, especially if they take into account recommendations made in the final section and if they continue to use certified NLP Practitioners, would provide data allowing this question to be more adequately addressed, at least for the case of rating series of eye movements.

Conclusions and Recommendations

Conclusions

The following conclusions were drawn from this study:

1. Overall agreement, by Eye Movement, among the certified Neuro-Linguistic Programming Practitioners who were used as raters in this study was highest on the first Eye Movement observed by the Practitioners.

2. Because of inflation of agreement scores due to the NA-effect, overall agreement scores among the raters in this study for Eye Movements 3 through 7 may not have adequately measured agreement among raters.

3. Overall agreement among raters in this study generally declined as the total number of observed eye movements increased.
4. One major reason for disagreement on eye movement ratings among raters in this study was disagreement over whether minimal eye movements should or should not be rated.

5. Certified Neuro-Linguistic Programming Practitioners used as raters in this study exhibited sufficient inter-rater agreement to warrant their use in future studies which do not require raters blind to the NLP model.

Recommendations

The following recommendations are made for future research.

1. The present study should be replicated with the following changes:
   a) A period of time should be allowed raters to become familiar with the characteristic eye positions and eye movements of each videotaped student before rating begins.
   b) A more truly random sample of certified NLP Practitioners should be used if possible.

2. Further research should be conducted using certified NLP Practitioners not only as raters of eye movements but also as raters of other behavioral cues because:
   a) They are the primary individuals who employ NLP theory and techniques, and it is important to understand how well they agree among themselves in practicing their skills.
b) They constitute a population which, given their prior training, may provide a natural and highly-skilled source of raters for some NLP-related studies.

c) Insights offered by these individuals can be extremely useful in the design and interpretation of studies which seek to test or which are otherwise concerned with the NLP model.

3. The issue of minimal eye movements and their relation to the NLP model, i.e., what constitutes a ratable change in eye position given that model, should be recognized as an essential issue to be addressed by any research seeking to test the NLP model through observation of eye movements.

4. The fact that some individuals' customary eye positions and/or movements are skewed from normal and some individuals' eyes are skewed with respect to one another should be recognized as important issues to be addressed by any research seeking to test the NLP model through observation of eye movements.

5. High-quality videotaping procedures should be used and eye movement monitoring systems (television based eye tracking systems) should be considered for all NLP-related research which involves rating eye movements. Perhaps a separate division of NLP set up to test these systems might be appropriate.

6. Research seeking to test the NLP model through observation of eye movements should concentrate on series of eye
movements rather than upon single eye movements associated with a stimulus item or task, as the latter do not seem to provide adequate tests of the model.

Finally, on a more general note, it is recommended that researchers seeking to test the NLP model take extreme care in design of their studies. Serious methodological problems were found in a number of prior studies which sought to test the model, making it highly questionable whether they provided an adequate test of that model.

This is unfortunate, for the NLP model may provide valuable insights into human behavior and information processing. Indeed, the researcher, himself an NLP Practitioner and psychotherapist, has found NLP techniques useful in his own practice. Yet, what is needed in this area is solid, replicable research to test the broad claims made by NLP proponents.

Cautious researchers who seek to test the NLP model should first become as familiar as possible with NLP theory and practice. Where confusion in the model arises they should seek clarification from NLP theorists.

It is also recommended that, where possible, researchers concentrate on NLP claims about strategies, for it is here, if anywhere, that the meat of NLP lies. That it is possible to determine the strategies (series of internal and external representations) used by those who perform certain tasks successfully and to teach those strategies to others is not only a broad claim, but is also, if true, of great significance. The
present study, in fact, was designed with a view to such future investigations--as one step toward the development of research that can adequately test and even, perhaps, eventually clarify and expand upon NLP claims about strategies.


APPENDICES
Appendix A

Letter to Practitioners
Directions for Certified Raters
Instruction Sheet to Raters
Dear NLP Practitioner,

The field of NLP is expanding and more information is being made available as a result of new research. A study is being conducted here at Oregon State University to code and quantify eye movement patterns as a function of an NLP strategy elicitation rating process. The study further attempts to provide a methodology for determining inter-rater consistency among Certified NLP Practitioners. You have been selected from a sample of NLP Practitioners to participate in this study.

The study will take place during the latter part of April at a mutually agreed upon time and location near you. Participation will consist of viewing video tapes of 10 subjects involved in the activity of spelling 10 different words. You will be asked to indicate on a chart each observed eye movement for each of the 10 spelling tasks. Complete instructions on the rating process will be reviewed with you by the researcher at the viewing site. The time required for your participation will be approximately 1 hour.

In order to control the many variables that could affect a study of this sort, you will be asked to complete a Raters Information form (sample enclosed). However, you are assured of complete confidentiality. Each form contains an ID number for this purpose. This number will be used to identify your ratings on all charts and forms filled out by you. In return for your participation, results of this study will be made available to you at the completion of the study.

Your participation is important for the success of this project, and your prompt reply will be greatly appreciated. I have enclosed a return envelope for this purpose and will be contacting you by phone during the next 10 days to answer any questions you may have and to talk with you about time and location for meeting.

I have also enclosed a sheet entitled "Directions for Certified NLP Raters" which will give you a more detailed idea of the rating procedure. At the bottom of that sheet is a list of the 10 words that each subject will be asked to spell.

Thank you in advance for your assistance. I eagerly look forward to meeting you in the near future.

Sincerely,

Michael B. Sun, MA
DIRECTIONS FOR CERTIFIED NLP RATERS

1) Before the study to rate eye movements begins, you will have three or more trial runs to rate spellers. The researcher will be present to assist you in questions and clarifications.

2) This is a limited observation eye movement exercise for a rater inasmuch as you will only be recording eye movements. Other responses will not be relevant to this study.

3) Please observe the three spellers on the trial video and record your observations. Feel free to ask questions.

4) During the trial run you may stop the tape at any time. During the study itself the tapes will run non-stop until the end of each word. The tape involved in the actual study will be viewed only one time.

5) There will be video sound of the students during viewing.

6) You will hear the experimenter ask for the spelling of the word. You should begin recording eye movements when you believe the person has begun his or her strategy for spelling a word and continue until you believe that the subject has finished spelling. There are five categories of possible responses from you. You must select one of the 5 possibilities for each rating box on the chart:

   V   A   K   C   N/A

The "V" notation indicates visual; "A" indicates auditory; "K" indicates kinesthetic; "C" indicates eyes centered and defocused; and "N/A" indicates "not applicable," i.e., no eye movement noted. When you believe that the subject has finished spelling the word, place N/A in the next available box and in all remaining boxes for that word.

7) Note that you are to rate all visual eye movements as "V". "Vc" and "Vr" will not be used in this study. All auditory eye movements are to be recorded as "A." "Ac," "Ar," and "Ai" will not be used in this study.

8) Note also that it will be important for you not to respond to the elicitations yourself, for example by spelling the words silently. This would of course interfere with your concentration on the subjects' eye movements. In order that you be thoroughly familiar with the words to be spelled, they are included below:

   pharaoh    license
   envelope   surprise
   receive    sincerely
   beautiful  lacerate
   antique    cologne
INSTRUCTION SHEET

1. HELLO. THE FIRST THING I WOULD LIKE YOU TO DO IS TO FILL OUT THIS FORM. AFTERWARDS, I'LL EXPLAIN WHAT WE'RE GOING TO BE DOING TODAY.

Give subject a form. As he or she fills it out, check to make sure that the right hand is being used. After the subject has completed the form, say:

2. THANK YOU. FIRST, I'D LIKE TO TELL YOU THAT I APPRECIATE YOUR TIME AND ATTENTION TO THIS STUDY. WHAT WE WILL BE DOING TODAY IS VIDEOTAPING YOU WHILE YOU SPELL SOME WORDS. THE WHOLE PROCEDURE WILL TAKE ABOUT HALF AN HOUR, AND I'D JUST LIKE YOU TO MAKE YOURSELF COMFORTABLE DURING THAT PERIOD. DON'T WORRY IF YOU'RE NOT SURE ABOUT THE SPELLING OF SOME OF THE WORDS. YOU AREN'T BEING GRADED ON HOW WELL YOU SPELL. SO JUST RELAX AND DO THE BEST YOU CAN WITH EACH WORD.

During this time show the subject to his/her seat and make an effort to see that the subject is at ease with the situation. When the subject is comfortable, start the video tape.

3. AND NOW WE'RE READY TO SPELL THE WORDS. HERE'S HOW WE WILL PROCEED. THERE WILL BE TEN WORDS. I WILL SAY THE FIRST WORD AND THEN I WOULD LIKE FOR YOU TO SPELL THAT WORD FOR ME OUT LOUD. IF YOU DON'T UNDERSTAND ME WHEN I SAY THE WORD, TELL ME AND I'LL REPEAT IT. AFTER YOU HAVE FINISHED SPELLING THE FIRST WORD, I WILL GIVE YOU THE SECOND WORD TO SPELL, AND SO ON, UNTIL YOU HAVE FINISHED SPELLING THE TEN WORDS.

Begin presenting the words. Say the first word to the subject, and after you have judged that the subject has completed spelling the word, wait five seconds and then present the second word, and so on. If the subject indicates that he/she does not understand the word to be spelled, repeat the word.

When the ten words have been spelled, say:

4. THANK YOU VERY MUCH. AGAIN, I APPRECIATE YOUR TIME IN HELPING WITH THIS STUDY.
Appendix B

Informed Consent - Video Tape Subjects
INFORMED CONSENT - CONFIDENTIAL RELEASE

Video Tape Subject Information Form

This information is for mailing records only. You will be assigned a number which will be the only method of identification used from this point on.

Name ________________________________

Address ________________________________

_____________________________________

Telephone ______ Age______ Sex: M____ F____

Mark True or False (T or F)

1) ___ I am right handed.

2) ___ I am a Native English language speaker.

3) ___ I take no medications.

4) ___ I do not wear glasses.

I am taking part in this research for extra credit in class ________.

Instructor ____________________________________________

Assigned Student Identification Number for this research:

[__________]
Appendix C

Rater Background Data
Background of Rater

Your Name__________________________ Rater No. ______

Age_______  Sex__________

Please print or type in the name of the individual and school that trained and certified you as an NLP Practitioner.

Trainer and name of program________________________

Date certified as NLP Practitioner____________________

How long have you been using NLP?____________________

What is your educational background?
   Last degree earned________________________
   Date graduated________________________
   Name of educational institution________________
   Address of above institution________________

Current vocational position________________________

I use NLP in:                  Sales
                                  Therapy
                                  Education
                                  Business
                                  Military
                                  Other

How important is the Eye Rating process of NLP to the overall strategy rating process in your opinion? Circle one.

  1    2    3    4    5
Very Important Unimportant Neutral Important Very Important

How many people's strategies have you rated using NLP techniques?_______
Has your opinion about the value of the eye rating meaning changed since becoming certified? Circle one.

No change  Increased  Decreased

What setting do you believe holds the most promise to achieve full implementation of NLP?

Education:  Elementary________________________
            Secondary________________________
            2 or 4 yr. college____________________

Business:  Interpersonal Skills________________
            Sales________________________
            Training____________________

Military:  Intelligence_______________________
            Training______________________

Psychology:______________________________

Other:__________________________________
Appendix D

Equipment
EQUIPMENT

1. Canon VR-30--capable of accurate freeze framing.
2. Canon VR-30--line remote control--capable of accurate freeze framing.
3. TDK-EHG Tape - 2 hour speed.
4. TV - Zenith 19 inch.
GLOSSARY

Accessing Cue - A behavioral cue that an individual employs to tune his or her neurology to single out cognitive activity within a particular representational system.

Anchoring - A discrete experience which evokes another experience. For example, at the moment a desired representation is being experienced by a client, a therapist may "anchor" that experience with a touch or some other bit of behavior which thereby becomes associated with and tends to bring about the desired representation in the future.

Behavior - Activity within any representational system, including input (e.g. the act of seeing), processing (e.g. listening to internal dialogue), and output (physical movement).

Behavioral Indicators - See Accessing Cues.

Certified NLP Practitioner - An individual certified by the NLP Organization as having completed an Organization-approved course of training in Neuro-Linguistic Programming designed to enable the individual to practice NLP techniques.

Cognitive Map - Internal representation of the external world.

Eye Movement - Change in position of the eye relative to the eye socket.

Eye Position - Position of the pupil of the eye relative to the eye socket. In NLP, an upward eye position of a subject, to either the right or left, is held to be an indicator of the occurrence of internal visual representations; eyes level to either the right or left and eyes down and to the subject's left indicate the occurrence of internal auditory representations; eyes downward and to the subject's right indicate the occurrence of internal kinesthetic representations.

Inter-rater Consistency - A measure of the agreement among the ratings of a group of raters.

Inter-rater Reliability - A measure of the overall correctness of the ratings of a group of raters.

Lead System - A representational system which an individual uses to access another representational system.

Neuro-Linguistic Programming - A set of beliefs and practices arising from the view that all behavior (in the broad sense defined above) is either identical with or else the result of systematically ordered sequences of sensory representations.
NLP Organization - The organization begun by John Grinder and Robert Bandler to develop and promulgate the Neuro-Linguistic Programming view.

Predicates - Verbs, adverbs and adjectives. According to NLP, an individual's usage of predicates typically reflects his or her Primary Representational System (PRS). For example, individuals with an auditory PRS will typically employ a greater proportion of predicates with auditory associations (for example "hear," "quiet," "sound") in their speech than predicates with visual or other sense modality associations.

Predicate Matching - The practice of using verbs, adverbs and adjectives which are associated with another individual's presumed Primary Representational System. This is one means, according to the NLP view, for one individual to establish rapport with another. Other ways of matching the individual's behavior, for example with respect to posture, are also held to be means for establishing rapport.

Primary Representational System - The representational system that a person favors in his or her strategies.

Rapport - See Predicate Matching.

Representation - A sensory percept (external representation) or sensory image (internal representation) held to represent some feature of an independently existing external world.

Representational System - A system of sensory representations of the external world. The representations are linked to one of the sensory modalities, e.g. auditory or visual.

Strategy - A sequence of representations which form a Test-Operate-Test-Exit (TOTE) pattern.

Strategy Elicitation - Determination, on the basis of accessing cues, of the strategy used by an individual to perform a certain task.

Strategy Installation - Teaching an individual or oneself a strategy for performing a certain kind of task.

Synesthesia - Access of one sensory modality through a representation(s) in a different sensory modality. For example, use of a visual representation of a phone to "bring to mind" the sound of the phone, i.e. an auditory representation.
TOTE or Test-Operate-Test-Exit - A behavior sequence wherein a present state of affairs is tested against a desired state of affairs (Test phase) with the result that (1) if the two do not match, a further bit of behavior occurs (Operate phase) followed by a further test, or (2) if the two do match, the behavior sequence ends (Exit phase).