



This is to certify that the

thesis entitled

THE EFFECT OF KNOWLEDGE OF RESULTS ON THE MAINTENANCE OF WRITING SKILLS

presented by

J. Gordon Schleicher

has been accepted towards fulfillment of the requirements for

____degree in ______ Ph.D.

Cartille G. C. entry Major professor

Date Nov-5, 1975

O-7639



OVERDUE FINES ARE 25¢ PER DAY PER ITEM

Return to book drop to remove this checkout from your record.

THE EFFECT OF KNOWLEDGE OF RESULTS ON THE MAINTENANCE OF WRITING SKILLS

Ву

J. Gordon Schleicher

A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

College of Education Instructional Development and Technology

ABSTRACT

THE EFFECT OF KNOWLEDGE OF RESULTS ON THE MAINTENANCE OF WRITING SKILLS

By

J. Gordon Schleicher

Overview

This research tested the effect of knowledge of results on the maintenance of writing skills. The 121 writers used in the study were employees of the Michigan Department of Social Services. Their primary job was writing procedural material for various manuals used by other employees. Over a five-month period knowledge of readability level of their writing was provided to half the writers as knowledge of results.' It was hypothesized that writers who knew how well their writing matched the levels of the reader would be better able to maintain acceptable writing levels. Writing levels were measured using a computer adapted readability formula.

Three treatment groups were used according to whether writers received (1) training only, (2) training and knowledge of results, or (3) only knowledge of results. A fourth control group received neither training nor knowledge of results.

J. Gordon Schleicher

Methodology

Readability levels were based on sentence length and syllable count, the most widely reocgnized means for measuring clarity in writing. A computerized readability program called STAR, Simple Test Approach for Readability, was used. The computer printout resulted in providing writers with a copy of the text analyzed, words of three or more syllables, average sentence length, average syllables per word, and the grade level equivalent of the writing. Opposite each of the last three was a goal or standard that each writer was asked to compare his/her score to. This printout became the knowledge of results and the means for measuring the performance level of each group.

The research question was based on discovering the most effective means for achieving readable writing and for maintaining an acceptable readability level after it was achieved. A nonequivalent control group design was used because experimental groups constituted intact groups of writers. Despite the unmatched nature of the groups, control for internal validity was increased by confirming the similarity between the groups. Analysis was accomplished using an analysis of variance and the Greenhouse-Giezer correction factor to account for the violation of score independence within groups. Significant difference was tested at the .05 level of confidence.

Results

The results showed that the trained group provided with knowledge of results maintained acceptable readability levels (between 10-14) throughout the test period. The group receiving only knowledge of results reduced their high readability levels from an initial score of seventeen to a low of fourteen. The extent of the limit of reduction can be attributed to this group's not having had the skills the trained group had. The training only group went from a low of twelve to a high of seventeen (their pre-training level). This increase occurred entirely within the first two months and was not distributed over the five-month period as anticipated.

Conclusions

The particular fluctuations of both groups receiving knowledge of results lends support to goal setting theorists which identify the impact of difficult goals on knowledge of results. These theorists have identified better performance with difficult goals. The readability analysis printout presented a goal for those groups receiving knowledge of results. For the trained group the goal of a grade level equivalent of 12 would not be considered difficult since that group was already writing at level 12. However, for the untrained group receiving knowledge of results that goal would be considered sufficiently difficult since they began the experiment at level 17.

The conclusions affirmed the effect of knowledge of results as an effective means for maintaining readable writing after training. The specific effect of the goal portion of knowledge of results would be identified by a replication of the experiment with two additional groups, one given knowledge of results and the other training and knowledge of results but without the goal statements accompanying the knowledge of results. The results of this experiment will provide meaningful information for organizations with sizeable writer groups who seek to train writers and maintain an acceptable readability level.

DEDICATION

This dissertation is dedicated to my wife, Margie, daughter, Marcia Jo, and son, Andy. It was their support with love, patience, and understanding which gave me the confidence needed to complete this writing.

ACKNOWLEDGMENTS

Many persons assisted me throughout this project. My sincere thanks to my faculty guidance committee, Drs. Kent Gustafson, Steve Yelon, and Frederick Wickert, who helped me learn with every step I took. Particular thanks goes to Dr. "Cass" Gentry who as my faculty chairman provided me with feedback, reinforcement and knowledge of results from start to finish.

I also acknowledge assistance from other specialists in their own fields. Their work and suggestions resulted in many product improvements along the way. Thanks to Judy Charland for her typing and suggestions for information arrangement. Thanks to Fran Miller for her editing of the final copy. And thanks to Dave Solomon who provided help in planning the experimental design and analyzing the results. I also recognized the influence my colleagues in Instructional Development and Technology have in my work. Thanks also to my friends in the Lansing, East Lansing community who accepted my absences and reduced involvement in other activities during this last year. Finally I thank my God whose supreme guidance make⁵ everything possible.

iii

TABLE OF CONTENTS

															Page
LIST	OF	TABLES	•	•	•	•	•	•	•	•	•	•	•	•	vi
LIST	OF	FIGURE	s.	•	•	•	•	•	•	•	•	•	•	•	vii
Chapt	ter														
]	Γ.	IDENTI	FICA	TIOI	N OF	T T	IE :	PROP	BLE	M	•	•	•	•	1
		Intr	oduc	tior	n	•	•	•	•	•	•	•	•	•	1
		Need	for	Thi	is S	Stud	ly	•	•	•	•	•	•	•	2
		The	MATS	TAR	Pro	ojec	ct	•	•	•	•	•	•	•	5
		The	Sign	ific	cand	ce 1	Eor	the	e D	epai	tme	ent	of		
		So	cial	Sei	rvic	ces	•	•	•	•	•	•	•	•	7
		Si	gnif	icar	nce	foi	с Т :	raim	ner	s ar	nđ				
			Inst	ruct	tior	nal	De	veld	ope	ers	•	•	•	•	9
		А	Part	icul	lar	Exa	amp	le V	√it	hin	MDS	SS		•	10
		De	fini	tior	ns					•					13
		Re	sear	ch (.+ ic	n on	•	•	•	•				14
		St	atem	ent,	of	Hvr	$\frac{1}{2}$	hese	• > c	•		•	•	•	14
		T.imi	tati	one	and	<u>γ</u> - β Δ β	2011	mnt	i on		•	•	•	•	17
		Baci	e in	The	anc	7 F16	550	mp c.	1011		•	•	•	•	19
		Dasi	s in shil	1 + 17	eur j and	(1 D2		ahi'	• 1 i +	•	•	•	•	•	17
		Mo	abii	T LY	anc •+		au		LTC	·Y					21
		Dica	asur			•		•	•	•	•	•	•	•	21
		DISS	erla	LIOI	1 01	/er \	Te	N	•	•	•	•	•	•	21
IJ	Γ.	REVIEW	OF	THE	REI	LATI	ED I	LITH	ERA	TURE	2	•	•	•	24
		Poad	ahil	i + 17	and		- A	ahil	1 i +						
		Mo	abii	I LY Omor	anc +		au		LIU	·Y					24
		Comp	asur	Dwa	16	•••	· Do		•	• • • • •	•	•		•	24
		Effo		רב הביד	Juuc	eu ha	Red L	auai		for	Tev	er:	2	•	33
		LILE				iDa(JK (Per	TOL		:е . 1	•	•	22
		Know	reage	e 01	Re	esul	LTS	(K)	K)	and	GOS	L			20
		Se		g	·	•	•	•	•		-	•	•	•	39
		Know	Teag	e 01	C RE	esu	LTS	as	a	Ke11	nioi	cei	-		4.2
		an	a mo	LIVA	ator	-	•	•	•	•	•	•	•	•	43
		Summ	ary	•	•	•	•	•	•	•	•	•	•	•	48
111	[.	DESIGN	MET	HODO	DLOC	ŞΥ	•	•	•	•	•	•	•	•	51
		Tester			-										E 7
		Intr	oauc	LIOR	1	•	•	•	•	•	•	•	•	•	21
		The	Popu.	Lati	Lon	•	•	•	•	•	•	٠	•	•	52
		The	Samp	те	•	•	•	•	•	•	•	•	•	•	53

Hypotheses	•	•	54 56
Allarysis	•	•	57
The form of knowledge of Results .	•	•	58
The Training	•	•	61
Level	•	•	62
			63
Measurement of Writing Level	•	•	63
IV. ANALYSIS OF RESULTS	•	•	67
Introduction	•	•	67
Hypotheses	•	•	73
Analysis of Group One Data			76
Analysis of Group Two Data	_		77
Analysis of Crown Throe Data	•	•	80
Analysis of Group Infee Data	•	•	00
Analysis of Group Four Data	•	•	02
Summary	•	•	85
V. SUMMARY AND CONCLUSIONS	•	•	88
Purpose of Study	_		88
Major Summary	•	•	90
	•	•	02
Conclusions	•	•	33
Implications Gathered from Results	•	•	95
Department of Social Services			97
Department of Social Services .	•	•	100
Recommendation for Further Research	•	•	100
Appendix			
A. ORIGINAL DOCUMENT AND REWRITTEN			
DOCUMENT	•	•	105
B. SOME OBSERVATIONS REGARDING WRITERS	•	•	108
C. FRY READABILITY GRAPH	•	•	111
D. NEWS CLIPPING REGARDING MICHIGAN DEPARTMENT OF SOCIAL SERVICES READABILITY PROJECT	•	•	113

BIBLIOGRAPHY		•				•		•		•		•	•	115
DIDLIUGRAPHI	•	•	•	•	•	•	•	•	•	٠	٠	•	•	4 4 4

LIST OF TABLES

Table			Page
1.	Reading Levels of Employees, Clients and Frequently Used Written Material	•	4
2.	Characteristics of Writers by Group	•	18
3.	Sensitivity Between Sentence Length and Syllable Count in Determining Grade Level Equivalent	•	64
4.	Group Mean Readability Level Scores for Each Measurement Point	•	73
5.	Between Group Analysis of Variances on Variables of Time and Group	•	75
6.	Comparison of Sample Original and Rewrite from a Single Writer at Two Time Points	•	79
7.	Within Group Variances in Grade Level by Time Periods	•	85

LIST OF FIGURES

Figur	e	Page
1.	Relationship Between Number of Correct Problems and Variables of KR and Type of Goal	42
2.	Computer Printout Explanation	59
3.	Sample Computer Printout which Served as Knowledge of Results	60
4.	Plot of Group Mean Scores on Readability Levels of Writing for Groups Three and Four	68
5.	Plot of Group Mean Scores on Readability Levels of Writing for Groups Two and Four	70
6.	Plot of Group Mean Scores on Readability Levels of Writing for Groups Two, Three, and Four	72
7.	Plot of Group Mean Scores on Readability Levels of Writing for Group Two	78
8.	Plot of Group Mean Scores on Readability Levels of Writing for Group Three	81
9.	Plot of Group Mean Scores on Readability Levels of Writing for Group Four	83
10.	Plot of Group Mean Scores on Readability Levels of Writing for Groups One, Two, Three, and Four	86

The text of this dissertation was tested at 12.4 grade level equivalent.

CHAPTER I

IDENTIFICATION OF THE PROBLEM

Introduction

The subject of this research was maintenance of writing skills after training. The hypothesis was that writing skill can be maintained over an indefinite period by measuring writing performance and then providing writers with knowledge of the results. Knowledge of results (KR) in this study was provided by a computer analysis of the readability level of the writing. That level was a measure of how difficult a piece of writing was based on sentence length and syllable count. The measurement of readability was expressed as years of education needed to read the material without difficulty.

The participants in the project were writers in the Michigan Department of Social Services. Over a three-year period prior to the project, writers in the department were given training in clear writing techniques. Tests given immediately after training showed that training did provide writers with the knowledge and skills to enable them to choose words and construct sentences which matched the education level of the intended readers. But a survey of trained writers conducted in

1976 showed that six months after training writers were performing at pre-training levels. This study sought to identify the effect of knowledge of results on maintaining the performance of writers following training.

Need for this Study

Ever since its formation, government has generated its own technical vocabulary. The public, as users of government services, is directly and indirectly guided by the written language which is imposed upon it. Edwin Newman, writing in his best seller, <u>Strictly</u> <u>Speaking</u>, asks the question, "Will Americans be the death of English?"¹ Newman goes on to suggest that the use of simple English is in a decline. Despite the humor that the author pokes at words and phrases that we Americans coin, there is a certain seriousness given to the problems we sometimes cause in not clearly communicating our thoughts.

With an ever increasing frequency, persons are calling for clearer, more easily understood government policies and procedures. As the number of government programs, employees, and users of government services increases so does the opportunity for miscommunication

¹Edwin Newman, <u>Strictly Speaking</u> (New York: Warner Books, 1974).

and its resultant errors, fraud, lost time and frustration.

Within Michigan state government a movement is underway to clear up this writing. A bill has been proposed in the state Senate which would require all commercial and public documents to be able to be understandable to "a person of average intelligence."² The Michigan Department of Social Services, the second largest department in state government, has recognized the need for clearer, simpler communications.

In 1975 the department recognized the need to train its writers to write clearly. But it wasn't until 1977 that the extent of the problem was evident. It was that year that this researcher applied a readability formula to 200 samples of departmental writing. The samples tested ranged in readings from grade level 17 to grade level 21. Subsequent testing of employees and clients of the department (Table 1) showed that between 20%-100% of certain groups had reading skills below the reading levels of written material that they were expected to comprehend and work with.

During the three years preceding this project this researcher trained over 500 writers using a

²State of Michigan, <u>Senate Bill 1061</u> (1977).

TABLE 1.--Reading Levels of Employees, Clients and Frequently Used Written Materials.

Group	Average Reading Level*	Reading Level of User Material*	Percentage Reading Below Material
Services Workers	17 (extrapolated)	15.0 Services Manual	20
Assistance Payments Worker Trainees	15	17.1 Assistance Payments Manual	86
All Employees	14	20.0 Administrative Handbook	100
*Measured as	years of educati	on (grade level equivalent)	

twelve-hour mediated training program called "Put It in Writing."³ Satisfactory completion of the course required the writer to write at a grade level no higher than 12. The level 12 was chosen as that level which could be read by the employee population. The training course did meet the objectives of enabling writers to choose those words and sentences which would result in a readability level of grade 12 or less.

In early 1976 this researcher did follow-up testing on fifty writers who had completed training between four and six months earlier. The results showed that 68% were writing at pre-training levels.⁴ The writers had failed to maintain the clear writing skills taught them. The need for an ongoing skill maintenance program had been identified.

The MATSTAR Project

This researcher submitted a proposal for a skill maintenance study project in June 1976. The proposal received no response. The following year it was resubmitted with some minor modifications and was accepted.

³Albert Joseph, <u>Put It in Writing</u> (Cleveland, Ohio: Industrial Writing Institute, 1972).

⁴J. Gordon Schleicher, Michigan Department of Social Services, memo to Director of Bureau of Program Operations, 9 November 1976.

The project was titled MATSTAR, which stood for Management and Administrative Training (the departmental unit to which this researcher was assigned) Simple Test Approach for Readability (the name of the readability computer program from General Motors).

The project used a computerized readability formula which enabled faster and more accurate calculation of the readability level. General Motors' (GM's) STAR program is based on the simple premise that comprehension does not depend solely on the skills of the reader.⁵ The way words are set down on paper can make reading hard or easy. Since 1923, methods have evolved for identifying reading difficulty as a numerical standard. These "indexes" can be equated to educational grade levels, allowing comparisons between reading levels of the reader and reading material. A discussion of the evolution of readability formulas is found in Chapter II.

The Flesch method⁶ for testing reading ease was adopted for use in the GM program because it provided an accurate scale of readability and was easily adapted to a computer system. The researcher chose the STAR program because it is faster than manually counting the

⁵General Motors Corporation, <u>S.T.A.R. General</u> <u>Motors Computerized Simple Test Approach for Readability</u>, 1976.

⁶Rudolph F. Flesch, <u>The Art of Readable Writing</u> (New York: Harper and Row, 1949), p. 213.

words, sentences and syllables. Also, STAR was uncomplicated and easily learned by persons not previously familiar with operating a computer terminal. It could be adapted for the batch mode or it could read the paper tape prepared for typesetting machines. Overall, the computer readability program was the least expensive to acquire, install and maintain. It also had the flexibility to adapt to expanding needs, and it was easy for non data processing people to use and understand.

The Significance for the Department of Social Services

The reason for studying this approach to skill maintenance rested on the benefits to the Department of Social Services. Readability analysis of departmental material showed evidence that the writing was more difficult than it needed to be. Such writing is often called heavy writing. It is not only harder to read but harder to write.⁷ When writers labor over the passages they construct, it costs time. And it takes more time to make long sentences grammatically correct.

From the reader's standpoint, heavy writing takes more time to read because of its greater length and less clear meaning. Increasing the difficulty of the writing

⁷Albert Joseph, op. cit., p. 12.

beyond a person's level to comprehend leads to misunderstanding.⁸ And misunderstanding can lead to errors.

A federal paperwork study of welfare forms found that heavy writing and unclear forms cost government six billion dollars a year.⁹ The study found this type of writing imposes unnecessary hardship and confusion on welfare clients. It can convey feelings of indignity and appear demeaning to the reader. The study also found evidence that such writing invited welfare fraud.

Audits of this state's welfare case files showed a high degree of errors. Many were attributed to unintentional misunderstandings on the part of clients and employees. It seemed reasonable that if simpler, clear English were used in instructions, fewer misunderstandings would result. And since the rate of errors determined in large part the amount of federal funding for state programs, the small dollar investment was well worth the cost. MATSTAR and this skill maintenance project were, to a major degree, an effort to improve departmental efficiency.

⁸Thomas G. Sticht, et al., <u>Determination of</u> <u>Literacy Skill Requirements</u>, as cited in "Maybe They Can't Read the Manual," <u>Training in Business and</u> <u>Industry</u>, June 1974, pp. 36-37.

⁹<u>Report of the Commission on Federal Paperwork</u>, Frank Horton D-NY, chairperson (Washington, D.C.: U.S. Government Printing Office, 1976).

Significance for Trainers and Instructional Developers

Maintenance of a skill back in the work area after being learned in a classroom has been an ongoing concern for instructors and managers alike.

In professional technical writing, particularly in government, there exists little opportunity to use monetary rewards as a reinforcer. In many organizations where writers work independently, supervisors may have few opportunities to provide writers with feedback in a timely and uniform manner. This is particularly true in the case of the Michigan Department of Social Services where over two hundred writers are found in more than a dozen separate departmental units. The problem is compounded by the highly technical and procedural nature of the writing. Most of the terms used in the writing are particular to the department. Such terms as "redetermination" and "unemancipated" are assumed to have no clearer substitutes.

Knowledge of results (KR), as used in this research, has a potential to capitalize on the writer's own internal ability and desire to write clearly. It does not require imposing an outside requirement and separate approval process for all written material.

This research will use a computer-generated printout as a uniform means for providing KR. Compared

to a supervisor's less objective analysis, computer analysis is not subject to personal preferences to writing style which may have little or no effect on readability. The computer provided KR can also provide more regular analysis. Such analysis is less threatening to the writer as compared to supervisory feedback in which case there may be a perceived threat to one's job security.

Although the computer program will greatly reduce the time required by hand calculation of readability some individuals, such as teachers, may not have access to a full-time computer; writers in large organizations do. It is where written material is printed in thousands of copies and read by tens of thousands of people that computer speed can result in substantial cost savings.

A Particular Example Within MDSS

Within the Michigan Department of Social Services, the Bureau of Assistance Payments regularly receives the most attention since it administers the Food Stamp, Medicaid, Medicare, Aid to Dependent Children, and General Assistance Programs. Within the Bureau's manual writing unit, six writers write new material and update existing policy. Their work is then printed and sent to over 3,000 holders of the manuals. These manuals

allow the Assistance Payments worker to determine eligibility and amount of assistance to be received. This manual series totals more than 2,000 pages. Over the six months of this study from January through June, 1978, Assistance Payments workers received on the average twenty-five pages of new or revised material each month.

Applying a readability formula to written procedures and policies found in the manual resulted in a grade level average of 17.1 years of education. Assistance Payments workers tested for reading ability showed a reading level average of 14.2 years of education.¹⁰ This three-year difference between the reader and reading material is one reason for the concern over readability. This difference is increasing because the educational requirements for new workers was recently reduced from 14 to 12 years of education. Recent testing of new worker-trainees found that they averaged only 12.8 years of education.¹¹

Trainers of new Assistance Payments workers were concerned enough by the increases in the number of trainees who do not pass the three-week training course that they asked that both trainees and training material

¹⁰J. Gordon Schleicher, Michigan Department of Social Services, memo to the Director, Bureau of Management and Staff Development, 17 November 1977.

be tested. The department's management was also concerned with the number of errors made by both workers and clients. Errors in case records can result in overpayment or under-payments to clients. And errors discovered by auditors result in reduced federal monies for these programs. Manual material is very technical. New employees and clients have voiced concern over the complexity and general lack of clear directions offered in manuals and forms. Research has shown that reading materials which exceed the reading ability of the reader will not be used.¹² Since just about every decision on eligibility for services is found in the manual, it is imperative that the manual be read.

The MATSTAR project was designed to make administrators and manual writers aware that very technical material requires greater clarity. Even before the particulars of the project were reported to department employees, reports from new trainees and workers began to feed back to department management. At the end of the sixmonth project, over thirty verbal reports were received from individuals who said they could recognize the improvement in readability.

¹²Thomas G. Sticht, et al., <u>Determination of</u> <u>Literacy Skill Requirements in Four Military Occupational</u> <u>Specialties</u> (Alexandria, Virginia: Human Resources Research Organiztion).

Definitions

<u>Acceptable Readability Levels</u> - The acceptable level for satisfactory completion of the training is a grade level of between ten and twelve grade level equivalent.

<u>Goal of Writing</u> - Identified on the knowledge of results printout as sentences which average between 15-20 words each, a syllable count average of less than 1.6 syllables per word, and an overall grade level equivalent of twelve.

<u>Grade Level Equivalent</u> (GLE) - An index of reading difficulty expressed as equal to years of education. For example a grade level equivalent of twelve means the writing is suitable for persons reading at a grade level of twelve.

<u>Heavy Words</u> - Difficult words and sentences which burden the reader unnecessarily. Any words which can be replaced with shorter, more commonly used words without changing the meaning.

Knowledge of Results (KR) - Describes the information that the person gets following the completion of a task. KR for this study was provided in the form of a computer printout analysis of a written sample.

Maintenance of Acceptable Readability Level -Readability levels which fall within the limits of ten and fourteen grade level equivalent.

<u>Readability</u> - Reading difficulty as measured by assessing the workload of words (using syllable count) and sentences (using sentence length) and expressed as readability index.

<u>Technical Writing</u> - Policy and procedural writing. Writing which conveys official direction or decisions as compared to opinion. Writing aimed at a group or class of employees rather than a single person.

<u>Training in Clear Writing</u> - A twelve-hour course primarily focused on the skills needed to choose the simple words and write short sentences.¹³

Research Question

Of the three alternatives listed below, which method provides the most effective means for reaching and/or maintaining the readability skills of manual and procedural writers?

- a. Knowledge of results without training in clear writing
- b. Training in clear writing without post training knowledge of results
- c. Training in clear writing followed by knowledge of results after training

Statement of Hypotheses

The most effective method to achieve and maintain acceptable readability skills will be one which combines

¹³Joseph, op. cit.

training and regular periodic knowledge of results. Knowledge of results alone will allow a person to reduce high readability levels but only to a lesser extent, over a greater period of time, and with greater variability of readability levels. Training alone will result in writers initially meeting acceptable levels, but without periodic knowledge of results the readability levels will revert to pre-training levels.

To test these hypotheses, four treatment groups will be used:

Group	1	A control group given no training and no knowledge of readability results	NO T/NO KR
Group	2	Given knowledge of results but no training	NO T/KR
Group	3	Given training only with no knowledge of results	T/NO KR
Group	4	Given both training and knowledge of results	T/KR
		Null Hypothesis 1: There will be no a	difference

in readability level of writing between the group receiving training only, and the group receiving both training and knowledge of results at each measurement point.

A positive relationship exists between knowledge of results and the maintenance of reduced readability levels. This is such that the trained group without knowledge of results will show increased readability levels following training. The trained group with knowledge of results will show little difference at each measurement point, therefore maintaining their clear writing skills throughout the experimental period.

Null Hypothesis 2: There will be no difference in readability level of writing between the group receiving knowledge of results only, and the group receiving both knowledge of results and training at each measurement point after the first month.

Although the knowledge of results only group is expected to show a reduction in readability levels during the experimental period, it is not expected to meet the readability level shown by the group receiving both training and knowledge of results.

> Null Hypothesis 3: There will be no difference in readability level of writing among the group receiving only knowledge of results, the training only group, and the group receiving both training and knowledge of results at each measurement point after the first month.

It is expected that each of the three experimental groups will have meaningfully different readability levels at each measurement point after the first month.

> <u>Null Hypothesis 4</u>: There will be no difference in readability level of writing between treatment groups and the control group at each measurement point after the first month.

It is expected that each of the three treatment groups will differ meaningfully from the control group. The control group should function as would any writing group not participating in the experiment. It receives neither training nor knowledge of results. And the readability of material written by group members is measured without their knowing they are part of an experiment.

Limitations and Assumptions

This study used intact writer groups, each assigned to a different treatment. Also, written material analyzed for readability was identified by group and not individual writer. Because the groups preexisted without random sampling or assignment, research control is affected.

Yet investigation of group members' ages, length of writing experience with department, length of time with the department, plus writing and reading ability, show similarities within and between groups. Table 2 identifies the characteristics of subject writers and sample writing expressed as group averages. On the basis of these confirmed similarities participants in the experiment are assumed to be representative of the total population of writers and that individual differences are equally distributed among groups.

Group No. Treatment h	l No T/No KR	2 No T/KR	3 T/No KR	4 T/KR
Group size	35	25	33	28
Age	40.1	37	39	38.2
Years in job	4.4	3.3	5.8	5.1
Years writing greater than 50% of time	8.1	9.4	8.3	9.1
Percent of males/females	51/49	60/40	45/55	64/36
Highest grade reached in school	16.7	16.6	16.4	16.7
Ratio of samples per person per month	1.16	1.18	1.15	1.25
Ratio high/low month	1.34/1.05	1.44/1.12	1.33/1.00	1.39/1.10
Writing level before training	17.2	16.8	16.7	17.6
Average sentence length	20.1	19.4	21.2	19.7
Average syllables/word	1.74	1.76	1.67	1.78

TABLE 2.--Characteristics of Writers by Group.

Since writing samples could not be identified by individual the use or nonuse by each individual writer cannot be established. For this experiment the use or nonuse of the knowledge of results was assumed to be equally distributed among groups. Thus it is assumed that any variable other than the independent variable would work equally within each group so as not to affect the outcome of any one group over another. It is further assumed that since no group gained new persons during the study and two groups lost only one person that these changes had no effect on writing level scores.

Basis in Theory

Knowledge of results theory states the subjects given feedback in the form of knowledge of task results will improve their performance on that task. Goal setting has also been shown to have an effect on performance. In particular when subjects are given specific information regarding performance following a task, and have a sufficiently difficult goal to achieve, they tend to make more improvement. That is, being able to compare knowledge of results against a difficult goal results in greater performance than KR alone.

This study provides writers with both knowledge of results and a sufficiently difficult goal. Writers in the knowledge of results only group are provided both a

readability analysis of their writing (KR) and a goal of reaching an acceptable readability level.

Achievement of this goal is dependent on the variables of sentence length and syllable count. Both the actual and goal levels are printed on the analysis for average sentence length, average syllables per word and grade level equivalent. Below is an example of the actual score and goal as found on readability analysis given writers as knowledge of results.

Average length per sentence	25.5	(should average 15-20)
Average number of syllables per word	1.7	(should average less than 1.6)
Grade level equivalent	16.4	(should match grade level of reader)

Based on these two theories it is hypothesized that the knowledge of results group will improve their writing skill performance. And it is also expected that, lacking learned techniques for improving writing, the group will not achieve as good a performance level as the trained group with knowledge of results.

It is also assumed that, although the training only group has the techniques to write clearly, without knowledge of results they will lose their skills. It is therefore assumed that this group will start out writing at an acceptable level and then regress to writing at their higher pre-training level.

The group with neither training nor knowledge of their level of writing performance will show no meaningful change in readability level.

Readability and Readability Measurement

Readability theories based on use of sentence length and syllable count have resulted in a dozen or more formulas for calculating the reading ease of writing. These theories hold that the longer the sentence the more difficult the writing is to read. And the higher the syllable count, the more difficult the writing is to read.

Those theories are the basis for both the training and knowledge of results presented in this study. Assuming their validity, training writers in techniques to reduce sentence length and replace many-syllable words with fewer-syllable ones should result in more readable writing. Use of computer readability analysis (KR) assumes that by telling writers their average sentence length, syllable count, and grade level equivalent, they will have the knowledge they need to reach and maintain readable writing.

Dissertation Overview

Chapter I has contained an introduction to the study of knowledge of results and its effect on the
maintenance of readable writing. Chapter I also reviewed the need for the study as it related to Michigan state government. This chapter also contains statements as to the significance of this study on the Michigan Department of Social Services and trainers and instructional developers. It ends with an explanation of the experiment's terms, hypotheses, limitations, and assumptions.

In Chapter II the pertinent literature is reviewed. The bulk of that review identifies the findings associated with knowledge of results/goal setting and readability theories. Also reviewed are sections dealing with the relationship of feedback on performance, references to knowledge of results as a reinforcer, and the use of the computer as a fast and accurate means to perform the computations necessary for determining the readability index.

The design of this study is described in Chapter III. Chapter III also contains (1) more detailed explanation of the two treatments--knowledge of results and training, (2) a full description of the writer population and samples used, (3) explanations of both the reading and writing measurement devices and the design for their use, and (4) the chapter ends with a description of the model used to test the hypotheses.

Chapter IV contains the results of the study. The results are then compared to each hypothesis statement.

The data from each group are analyzed separately and then summarized in terms of most meaningful findings.

The major summary covering the whole experiment is presented in Chapter V. This chapter also contains implications gathered from the results. It ends with recommendations for further research into the effects of knowledge of results and goal setting.

CHAPTER II

REVIEW OF THE RELATED LITERATURE

This review is divided into five sections which deal directly or indirectly with this research: Readability and Readability Measurement, Computer Produced Readability Levels, Effect of Feedback on Performance, Knowledge of Results and Goal Setting, and Knowledge of Results as a Reinforcer and Motivator. A summary follows the last section.

Readability and Readability Measurement

This section reviews what the literature says about readability and the means for measuring the readability of written material.

The term "readability" has been defined and interpreted in many ways. Edgar Dale and Jeanne Chall, two of the best known researchers in this area, state that in the broadest sense readability "is the sum total of all those elements within a given piece of printed material that affects the success which a group of readers have with it." They then define "success" as the

extent to which "they (readers) understand it, read it at optimum speed, and find it interesting."¹

The idea underlying readability measurement is the appropriate matching of reader and printed material. It assumes that readers differ in their ability to read. Printed material is therefore the object to vary in making the match. Suitable matching is essentially a problem of prediction and control. It implies that the writer knows something about the factors that make for ease and difficulty. The writer must then have a means for estimating these factors.

The earliest procedure for matching written material to the reader was based solely on judgment.² Thorndike's <u>Teachers'Word Book</u> (1921) was one of the first efforts to objectively measure the difficulty level of reading materials. Early attempts (before 1934) to analyze readability relied upon vocabulary variables as determiners. Between 1934 and 1938 readability research efforts broadened the number of factors considered. Reliance on word lists compiled by Thorndike and others diminished, and more attention was given to factors such

¹Edgar Dale and Jeanne S. Chall, "Predicting Readability," <u>Educational Research Bulletin</u>, January 1948, p. 38.

²Jeanne S. Chall, <u>Readability: An Appraisal of</u> <u>Research and Application</u> (Columbus, Ohio: Ohio State University Press, 1954), p. 9.



as sentence length and syntactical construction such as parts of speech.

One of the earliest works was done by Irving Lorge in 1939. He concluded that "probably the best predictor of passage difficulty is some measure of vocabulary."³ During the same year G. A. Yoakam made available a readability formula based on vocabulary.⁴ This formula was based on a weighted index of vocabulary difficulty.

The period between 1940 and 1950 was characterized by efforts to make readability formulas more efficient. Rudolf Flesch postulated that previous formulas gave too much emphasis to vocabulary. His factor analysis of reading-comprehension tests gave support to his hypothesis that, while reading comprehension at the lowest levels is largely "the perception and understanding of individual words" beyond that level it is "concerned more with relationships between the ideas."⁵

During that period Edgar Dale and Jeanne Chall, Irving Lorge, and G. A. Yoakam reduced the number of

³Irving I. Lorge, "Predicting Reading Difficulty of Selections for Children," <u>Elementary English Review</u> XVI (October 1939):231-32.

⁴G. A. Yoakam, <u>Basal Reading Instruction</u> (New York: McGraw-Hill Book Company, 1955), pp. 229-340.

⁵Rudolf Flesch, "A New Readability Yardstick," Journal of Applied Psychology XXXII (June 1948):221-33.

variables in their formulas. By this time there were at least a half dozen readability formulas in use. During this same period quantitative associational studies were done using these formulas. The results of these tests were reported by Chall in 1954.⁶ Her appraisal of the research on readability formulas up to that time reported that Flesch's formula discriminated difficulty at higher levels (ninth grade and above) than the Gray-Leary and Lorge formulas. The Flesch and Dale-Chall formulas, in seeking discriminated the difficulty of materials at lower levels.

Lorge reported a fundamental weakness in the use of tested comprehension as a criterion measure of readability in several formulas. Tested comprehension difficulty can be changed by the "ease or difficulty of the question asked."⁷

Chall's report goes on to cite the high correlations between the three factors common to all formulas in testing vocabulary workload: between the Dale and Flesch .7932, between the Flesch and Lorge .7441, and between

⁶Chall, op. cit., p. 90.

⁷Irving I. Lorge, "Readability Formulas--An Evaluation," <u>Elementary English</u> XXX VI (February 1949): 89.

the Dale and Lorge .7988.⁸ These data support Lorge's findings that a measure of vocabulary load is the important factor in rating reading difficulty and that all the measures of vocabulary are highly correlated.

Chall also analyzed the subject matter on which separate research studies were based. She found that most investigators used specialized materials. For example, Dale and Tyler used only health materials.

More recent studies have used letter redundancy and the independent clause as additional factors. A procedure using a patterned deletion of words, the Cloze procedure, has also gained recent interest by researchers.⁹

The study which used the most factors to determine readability was the Gray-Leary study.¹⁰ In the study, Gray and Leary attempted to relate 82 factors to determine readability. At the end of the investigation all but five were ruled out either because they could not be reliably measured, because they occurred only once in less than half of the selections, or because they could be combined with another factor.

⁸Chall, op. cit., p. 40.

⁹W. L. Taylor, "Cloze Procedure: A New Tool for Measuring Readability," <u>Journalism Quarterly</u> 30 (1953): 415-20.

¹⁰Chall, op. cit., p. 156.

Chall concluded her study of readability with the following generalizations. That of the four factors which contributed to reading difficulty (content, stylistic element, format, and organization), only stylistic elements could be reliably measured and verified. That within the four stylistic elements (vocabulary load, sentence structure, idea density, and human interest), vocabulary load "most significantly related to all criteria of difficulty."¹¹ Every study found some measure of vocabulary difficulty significantly related to comprehension. Vocabulary difficulty was measured either by reference to a word list or by word length. Studies show a moderate to high correlation between these two variables (between .67 and .84).

A study by Joseph Vaughan, Jr. yielded similarly high correlations.¹² Applying the Spearman rank correlations among the scores yielded by three formulas found Dale-Chall and Fry at .89, Dale-Chall and SMOG at .87 and Fry and SMOG at .80 (N=87 significant at .001 level).

One final caution is appropriate. Despite the evidence that each of the tested formulas was found valid, at least as valid as readability formulas can be,

¹²Joseph Vaughan, Jr., "Interpreting Readability Assessments," <u>Journal of Reading</u>.

¹¹Ibid., p. 157.

their scores are only an estimate of the reading difficulty. No discrete score can be considered definitive. Just as an IQ score is properly considered an indication of a general range of intellectual capacity, a readability score is most properly interpreted when it is taken to be a point within a range of grade levels.

The use of sentence length and vocabulary difficulty (as determined by syllable count) are still most widely used factors in readability measurement today. Both can be accurately measured and easily adapted to automated systems of measurement. For these reasons they were most appropriate for use in this research which required accurate, inexpensive, and quick determination of readability on a high volume scale.

The Dale and Flesch formulas were chosen for use in this study for three reasons: both formulas use sentence length and syllable count as determiners, they were validated on material similar to that in the study, and they accurately discriminate for the grade levels anticipated in the study.

Related to this research are numerous studies of the relationship between student reading levels, the readability of textbooks, and achievement. The

results of those studies are mixed. The following examples are typical. They are particularly worth noting because they were conducted less than a year apart. The subjects were nursing students, and they both measured readability of nursing texts.

In 1976 Gloria Kilian performed reading level tests on nursing students in Kansas.¹³ She then compared those scores with the readability level of nursing texts used and with scores received on the final nursing certification exam. Her conclusion was that "there was a significant relationship between achievement and reading ability." A very similar study by Musa Caffey performed on Mississippi nursing students found "no significant difference between reading scores and those who passed or failed that state's nursing exam.¹⁴

¹³G. L. Kilian, "The Relationship Between Readability of Assigned Textbook and Reading Levels of Students in a School of Nursing," Ph.D. dissertation, Kansas State University, 1976.

¹⁴Musa B. Caffey, "A Study of the Relationship Among Reading Levels of Practical Nursing Students' Readability Levels of Textbooks and Pass/Fail Scores on the State Board Examinations for Practical Nursing Students," Ph.D. dissertation, University of Southern Mississippi, 1977.

Although both studies were performed on essentially similar populations and texts, the different finding may be a result of other variables. The Caffey Study used the Nelson-Denny Reading Test and Flesch Reading Formula, while the Kilian Study used the California Advanced Reading Test and Fry Readability Graph. Because research documents high correlations between both reading tests and readability formulas, another difference is more likely. That difference is the one between Kansas and Mississippi State Board examinations for nursing. Differences between the two student populations, based on different entrance criteria, may also account for the discrepancy. Not enough documentation is given to tell for sure.

Another related question is the possible relationship between reading achievement and writing achievement. A research study conducted at the University of South Carolina in 1976 looked into that relationship and those between amount of reading and writing achievement. The results showed negligible to no relationship between reading achievement, amount, diversity, and overall writing achievement.¹⁵

¹⁵F. L. Thomas, "The Extent of the Relationship Between Reading Achievement and Writing Achievement Among College Freshmen," Ph.D. dissertation, University of South Carolina, 1976.

Computer Produced Readability Levels

Ever since the early days of computer use, dating back to the development of mathematical routines in the 1940's, great strides have been made in speed of computation. Advances in computer hardware have led to improvement of approximately 10⁶ in speed of computation.¹⁶ This has resulted in researchers developing many software packages to take advantage of calculating speed and accuracy. Developing a computer program to determine readability involves first identifying and counting words and sentences and second identifying and counting syllables. The former presents little difficulty. The latter presents more difficulty.

A word may be defined as any set of alphanumeric characters delineated by blanks and punctuation marks. The sentence is identified as all words occurring between two periods. The problem with counting syllables is one of defining syllables objectively. One reasonably complicated computer procedure to identify syllables was reported by I. E. Fang in 1968.¹⁷ This

¹⁶Morris Rubinoff and Marshall C. Yovits, eds., <u>Advances in Computers</u> (New York: Academic Press, 1977), p. 58.

¹⁷I. E. Fang, "By Computer Flesch's Reading Ease Score and a Syllable Counter," <u>Behavioral Science</u> 13 (1968):249-251.

program computes a great many exceptions to the rule which equates a single vowel with a syllable. The computer not only determines when an "e" ends a word, but also decides whether this letter has been preceded by one of several consonants and whether this consonant was preceded by a vowel.

Coke and Rothkopf compared Fang's program with three somewhat simpler ones which count vowels (A, E, I, O, U, Y) to get the number of syllables.¹⁸ Their computer estimated syllable count based on vowels yielded a correlation of .92 compared to a hand count.

General Motors' Simple Test Approach for Readability (STAR) program is programmed to do six things:

- A. Count end-of-sentence marks (periods, exclamation marks, question marks, semicolons, colons) to find the number of sentences in the sample.
- B. Count spaces after each word to find the number of words.
- C. Count vowels (A, E, I, O, U, Y) to get the number of syllables.
- D. Ignore final -E, -ES, -ED, which are not counted as syllables.
- E. Read any word of three letters or less as a single syllable (ARE, THE, A, OFF).

¹⁸Esther U. Coke and Ernst Z. Rothkopf, "Note on a Simple Algorithm for a Computer Produced Reading Ease Score," Journal of Applied Psychology 54 (1970):209.

F. See all double-vowels as one syllable (OO, EI, AY).

A study by Meri Coleman produced an even simpler method for estimating syllables based on word length alone.¹⁹ Her formula correlated at .92 compared with the Cloze readability formula. When T. G. Szalay cross validated the study with a new set of passages the correlation was .88, almost no shrinkage.²⁰ Coleman's premise was that the keypunching of texts required by computer-based formulas was generally more expensive than obtaining a readability score by hand counting.

Effects of Feedback on Performance

A study found to closely parallel this researcher's in purpose was one which provided college professors results of student ratings. A study by Daniel Smith in 1977 examined the extent to which feedback of student rating results to instructors would lead on its own to improvement in (teaching) performance.²¹ Smith's premise was that

¹⁹Meri Coleman, "A Computer Readability Formula Designed for Matching Scoring," <u>Journal of Applied</u> <u>Psychology</u> 60 (1975):283.

²⁰T. G. Szalay, "Validation of the Coleman Readability Formula," <u>Psychological Reports</u> 17 (1965):966.

²¹Daniel Lloyd Smith, "The Relationship of Feedback to Professors of the Results of Student Ratings of Their Teaching Effectiveness at Mid-Semester to Their End of Semester Ratings," Ph.D. dissertation, University of Southern California, 1977.

feedback of results of ratings would provide information to the instructor that he would use to modify his performance. The feedback would tell the instructor what he was doing well or poorly and what he could do to improve his teaching.

The results were not the same for all subjects. Smith summarized his findings by stating that feedback and non-feedback instructors did not differ significantly in their mean end-of-semester scores. A particularly interesting result was a negative correlation of -.25 found between the magnitude of the feedback effect and mean rating of the instructors' behavior. Smith is cautious in his conclusions stating that although not effective for all instructors feedback "may" have had a positive effect on some. He also concluded that those instructors initially given low or moderate ratings would be most likely to benefit from feedback.

R. B. Payne and G. T. Hauty in 1955 reported on the question concerning the effects of feedback on prolonged performance.²² These authors reported on an important distinction in providing feedback, that of identifying fact and location of error (incitive feedback) contrasted with merely appraisal of substandard

²²R. B. Payne and G. T. Hauty, "Effect of Psychological Feedback Upon Work Decrement," Journal of Experimental Psychology 50 (December 1955).

performance (directive feedback). The latter identifies only a shortcoming without suggesting some specific corrective action beneficial to performance. Previous findings resulted in the conclusion that supplementary auditory or visual feedback boosted proficiency but did not affect the rate of decline in overall performance (work decrement).²³ Certain differences in the behavior were reported as a result of this study. The results of incitive feedback were found to be diminished by task duration, while those arising from directed feedback were not. Directive feedback signals induced an immediate increase in task proficiency. The increment was a positive function of the specificity of the information supplied and was able to be maintained throughout the task length. While incitive signals also induced an immediate increase in proficiency, the effect declined after the first hour.²⁴

There is increasing evidence that delayed feedback is better than immediate feedback when one is learning meaningful material.²⁵ Such evidence is important to this study since feedback in the form of

²⁵J. M. Sassenrath, "Theory and Results on Feedback and Retention," <u>Journal of Educational Psychology</u> 67 (1975):894.

²³Ibid., p. 348.

²⁴Ibid., p. 351.

performance (directive feedback). The latter identifies only a shortcoming without suggesting some specific corrective action beneficial to performance. Previous findings resulted in the conclusion that supplementary auditory or visual feedback boosted proficiency but did not affect the rate of decline in overall performance (work decrement).²³ Certain differences in the behavior of directive and incitive feedback were reported as a result of this study. The results of incitive feedback are diminished by task duration, while those arising from directed feedback are not. Directive feedback signals induced an immediate increase in task proficiency. The increment was a positive function of the specificity of the information supplied and was able to be maintained throughout the task length. While incitive signals also induced an immediate increase in proficiency, the effect declined after the first hour.²⁴

There is increasing evidence that delayed feedback is reliably superior to immediate feedback when one is learning meaningful material.²⁵ Such evidence is important to this study since feedback in the form of

²⁵J. M. Sassenrath, "Theory and Results on Feedback and Retention," <u>Journal of Educational Psychology</u> 67 (1975):894.

²³Ibid., p. 348.

²⁴Ibid., p. 351.

knowledge of results is provided between three days and one week following the task performance. The fact that delayed feedback is superior seems contrary to the conventional behaviorists such as Hull, Skinner, and Spence. These theorists would say that learning was better when feedback followed immediately after one's task performance.

Sassenrath and Yonge explained this result by suggesting that with delayed feedback subjects have more time to make use of the cues provided.²⁶ Kulhavy and Anderson have another theory for the same result. They postulate that over the delayed feedback period wrong responses are forgotten more readily.²⁷

Knowledge of Results (KR) and Goal Setting

Most studies involving knowledge of results have focused on the information or cueing function of KR compared with the motivational function. The former refers to information given to subjects regarding the nature and locus of errors and the nature of the correct response. The premise is that given a constant

²⁶J. M. Sassenrath and G. D. Yonge, "Delayed Information Feedback, Feedback Cues, Retention Set, and Delayed Retention," <u>Journal of Educational Psychology</u> 59 (1968):70.

²⁷R. W. Kulhavy and R. C. Anderson, "Delay-Retention Effect with Multiple-Choice Tests," <u>Journal of</u> <u>Educational Psychology</u> 63 (1972):507.

motivational state the more information given to the subject about the task the better the performance.

In several studies illustrating this visual reaction times of subjects given no KR have been compared with those given KR.²⁸ Knowledge of results was administered by flashing a green light whenever the subject's reaction time was faster than a designated standard and a red light was presented when the reaction time was slower. In these studies, KR subjects showed faster reaction times. These findings have been interpreted by their authors as demonstrating that knowledge of results enhances reaction-time performance.

Much less is written about the manner in which knowledge functions to motivate the subject. Motivating is much more complicated than information giving, which may account for the lesser number of studies. In early studies dealing with motivational KR, subjects were given

²⁸R. M. Church and D. S. Camp, "Change in Reaction-Time as a Function of Knowledge of Results," Journal of Experimental Psychology 78 (1965):102-106; P. D. McCormack, "Performance in a Vigilance Task with and Without Knowledge of Results," <u>Canadian Journal of Psychology</u> 13 (1959):68-71; P. D. McCormack and W. G. McElheran, "Follow-up of Effects on Reaction Time with Partial Knowledge of Results," <u>Perceptual Motivation Skills</u> 17 (1963):565-566; P. D. McCormack, F. R. S. Binding, and Joanne Chylinski, "Effects on Reaction-Time of Knowledge of Results of Performance," <u>Perceptual</u> Motivation Skills 14 (1962):367-372; P. D. McCormack, F. R. S. Binding, and W. G. McElheran, "Effects on Reaction Time of Partial Knowledge of Results of Performance," Perceptual Motivation Skills 17 (1963):279-281.

explicit goals to reach. No KR subjects were told either not to think about trying to improve their score, or they were told to "do their best." Edwin Locke reports that in these experiments the effects of goal setting were confounded with the effects of knowledge of score, making it impossible to determine which was the critical variable.²⁹ Locke suggested that the effects of KR be separated from the effects of goal setting. The major premise is that the level of effort on a task is determined largely by the subject's conscious performance goals.

Locke's experiment used a 2 X 2 fixed model design. The variables were knowledge of results and type of goal (do-best goal, and hard goal was 10% higher than do-best). The problem involved simple addition. Measurement included (a) deterioration of scores, (b) problems correct per minute, and (c) percentage of errors. The results are shown in Figure 1.

Locke's results support the hypothesis that differences in KR are a function of differences in performance goals associated with KR conditions. The study is limited by the small number of goal classes. Even then it suggests that to predict the effect of KR on performance level it is not enough to know that the subject has

²⁹Edwin A. Locke, "Motivational Effects of Knowledge of Results: Knowledge or Goal Setting?," <u>Journal</u> of Applied Psychology 51 (1967):324.



Figure 1.--Relationship Between Number of Correct Problems and Variables of KR and Type of Goal.

such knowledge. According to Locke it is also necessary to know what the subjects do with the knowledge, including how they evaluate it and what goals they set in response to it.³⁰

Knowledge of Results as a Reinforcer and Motivator

Skinner wrote that all informative feedback related to the acquisition of motor skills should be classified under the category of reinforcing events.³¹

This study deals with the issue of feedback rather than reinforcement because no attempt has been made prior to the study to show that the feedback used has a reinforcing effect on the study subjects. Nonetheless, a review of reinforcement literature is necessary as a basis for expecting certain responses to occur (writing skills) following feedback.

A picture of feedback and knowledge of results theory comes out of the early studies on reinforcement. Beginning with Skinner's work in the 1930's, he considered reinforcement as the experimental operation of presenting a reinforcing stimulus contingent on a response.

³⁰Ibid., p. 327.

³¹Robert M. W. Travers, Ian E. Reid, and R. Keith Van Wagenen, <u>Reinforcement: A Review of Selected Research</u> (University of Utah, 1963), p. 109.

Hull attempted to explain reinforcing stimuli as they related to primary or secondary drives. For Hull reinforcement consisted of presenting, contingent upon some response, <u>a drive reducing</u> stimulus. Hebb on the other hand associated reinforcement with an increase in drive.³²

Premack left the concept of a reinforcing stimulus out of his definition. To him a behavior is reinforcing to other behaviors occurring close together in time if it has a higher rate of occurrence than the one which it is to reinforce.³³

In reviewing the various definitions, it is clear that some are tied to stimulus events while others are tied to response events.

Another problem with reinforcement research is the lack of a consistently applied definition of reinforcement. There is no lack of definitions, only in agreement on how to explain the phenomena.

In his survey of the literature, Travers puts the definitions up to that date into two classes, operational and theoretical. Operational definitions are those in which the term designates the process of presenting a stimulus condition that results in a measurable change in

³²Ibid., p. 111. ³³Ibid., p. 107.

some response characteristic. Definitions involving theoretical terms are those involving terms such as "habit strength" or "need reduction."

Bruner³⁴ calls reinforcement the Law of Effect. He defines it as the result of when a reaction is more likely to be repeated because of its being followed by a "satisfying state of affairs." He also expresses doubt that the "satisfying state of affairs" can be reliably found outside learning itself. In effect, he is saying that reinforcement is only effective when it is generated within the person. He cites such examples as curiosity, desire for competence, deep sense of commitment and aspiration to emulate a model as the sources for reinforcement. It is through the "intrinsic motive," says Bruner, that one finds both its source and reward in the exercise of learning.

It should also be noted that, although knowledge of results has been shown to reinforce correct responses, the two definitions are quite different. In fact, it has been noted that educators have used the terms differently from psychologists, resulting in some confusion. In KR and reinforcement a response is strengthened, but the procedures differ. Educators see repetition or

³⁴Jerome S. Bruner, <u>The Process of Education</u> (Cambridge, Mass.: Harvard University Press, 1960), p. 128.

rehearsal as the procedure denoted by reinforcement.³⁵ Knowledge of results, on the other hand, is generally considered a sub-class of reinforcement.

Knowledge of results is considered an essential element in improving performance. Travers said that if a person were to perform but never knew how well he was doing, he would have no basis for improving.³⁶ In his review of the literature he found a general agreement among researchers that KR provided either during or after performance "had the potential of programming behavior precisely to reach some end state."³⁷

In a study by Cummings, Schwab, and Rosen, the impact of goal setting on knowledge of results was compared. Are higher goals a function of KR manipulation or the higher performance? The researchers found that when performance effects are accounted for, KR still influences goals significantly.³⁸ When the effects of four forms of

³⁶Robert M. W. Travers, <u>Essentials of Learning</u> (New York: MacMillan Company, 1972), p. 50.

³⁷Travers, Reid, and Van Wagenen, op. cit., pp. 9-28.

³⁵Association for Educational Communications and Technology, <u>Educational Technology</u>: <u>Definition and</u> <u>Glossary of Terms</u> (1977), p. 224.

³⁸L. L. Cummings, Donald P. Schwab, and Marc Rosen, "Performance and Knowledge of Results as Determinants of Goal Setting," <u>Journal of Applied Psychology</u> 55 (1971):526.

KR were compared, correct KR increased goal level significantly above that generated by no KR. Incomplete KR increased goal level insignificantly. And erroneous low KR decreased performance below the level associated with no KR. This study demonstrated the independent and combined effects of KR and performance goals. The authors cited as implications the need to identify clear standards and measurement processes which are specific, since their results showed that providing subjects with results on the "upward side" may actually be dysfunctional.

Bruner, too, identified knowledge of results as a prerequisite for learning. He also recognized the importance of time and place when giving KR.

Knowledge of results is useful or not depending upon when and where the learner receives the corrective information, under what conditions such corrective information can be used and the form in which the corrective information is received.³⁹

The generally consistent rule seems to be that KR should come at a point in time when the person can compare past results with some criterion.

A typical experiment using knowledge of results is one such as Gibbs and Brown conducted in 1956. Subjects were to work at the repetitive task of copying pages on a copy machine. In half the trials the subjects could

³⁹Bruner, op. cit., p. 50.

see a counter which tallied each page. In the other half the counter was covered. The subjects in this study were not instructed to meet a set quota. They were paid uniformly and they were never supervised. The researchers tried to arrange conditions in such a way that the subjects' output would be determined entirely by self competition and self satisfaction. The results showed that when the subjects could see the counter their output was significantly higher (25%) than when they could not.⁴⁰

Summary

There has been long acceptance for use of sentence length and syllable count as a measure of readability. Research on readability formulas has found them to be reliable measures of readability. The Dale and Flesch formulas, which form the basis for this study's calculations, were found to accurately discriminate between levels of reading difficulty. Both were validated on material similar to that in this study.

The literature also supports the use of computer produced scores. This is based on the studies which show a high correlation between hand calculated and computer calculated scores using the Flesch Reading Ease formula.

⁴⁰C. B. Gibbs and I. D. Brown, "Increased Production From the Information Incentive in a Repetitive Task," as reported in "Knowledge of Performance as an Incentive in Repetitive Monotonous Task," by Alphonse Chapanis, Journal of Applied Psychology 48 (1964):263.

Feedback has long been regarded as having a positive impact on performance. Researchers such as Payne and Hauty have presented evidence to make a distinction between two types of feedback. One merely identifies that the performance is substandard; the other type identifies the location of the errors and other relevant facts. The latter has been shown to increase performance and maintain it over a longer period of time. This has a direct relationship to the form of feedback provided in this study. Writers receiving feedback will be given specific scores on sentence length, syllable count, and overall grade level of the writing.

The effect of goal setting on knowledge of results has the most guidance to offer of any section. Given the amount of writing on the subject, goal setting is the latest attempt to identify what makes feedback or knowledge of results work better in some cases when compared with others. The literature supports the theory that subjects perform best when given a sufficiently difficult goal as compared to a "do-best" goal or no goal at all. This study presents a goal to both knowledge of results groups.

Whether knowledge of results is a reinforcer of behavior may depend more on how one defines reinforcement and knowledge of results. Early research made no distinction between them. Traver's major review of reinforcement

research in 1963 devotes only six pages out of nearly two hundred to a discussion of knowledge of results. While Skinner classified all feedback related to the acquiring of motor skills under the category of reinforcing events, not everyone agrees. Carter V. Good states that feedback is not necessarily synonymous with reinforcement, "since it is not defined by its effect on the recurrence of the response."⁴¹ Markle, writing in the same publication, identifies knowledge of results as a subclass of reinforcement because of its having been shown to help learning. Markle reports that "when the student's correct response is followed by presentation of the correct answer, the probability that the correct response will recur is increased."⁴² A note in this Educational Technology, Glossary of Terms just quoted from further identifies the confusion posed by different definitions for the same term. The publication warns educators to be aware that the psychological use of the term reinforcement does not parallel their own use of it.

While there seems to be confusion among the ranks of educators concerning the specifics of these definitions, there does seem to be agreement that knowledge of results can be one element in improving performance.

⁴¹Association for Educational Communications and Technology, op. cit., p. 328.

⁴²Ibid., p. 334.

CHAPTER III

DESIGN METHODOLOGY

Introduction

This section contains a review of the techniques used to collect and analyze data reflecting the effect that training and knowledge of results had on the maintenance of an acceptable writing level among writers of manuals. The design methodology used four intact groups which were assigned to four experimental conditions:

- Group 1 Control group of thirty-five writers given neither training nor knowledge of results.
- Group 2 A group of twenty-five writers given knowledge of results.
- Group 3 A group of thirty-three writers given training in clear writing.
- Group 4 A group of twenty-eight writers given both training and knowledge of results.

It is believed that results from this study will enable the Michigan Department of Social Services and similar agencies to make more appropriate decisions in improving the readability of its written material. The decisions have direct impact on over two hundred full-time writers and over twelve thousand other employees in the department who write as a portion of their job. The

indirect impact affects over six million public citizens who use departmental services.

Managerial constraints on the project prevented a desired random selection and assignment of subjects. The consequences of that limitation seem less meaningful with the evidence of homogeneity of writers within the groups. That evidence, discussed in Chapter I (Table 2) showed that treatment sample and the target population differ very little.

Objective measurement of readability level variables was insured by use of computer computation. The knowledge of results took the form of a computer printout. That printout listed average syllable to word, average sentence length and grade level equivalent of the writing. Next to those three indicators were fixed goals which served as direction and reminders on how to achieve more readable writing.

Analysis was conducted by comparing group means and variability within and among groups.

The Population

The population for this study included full-time writers employed by the Michigan Department of Social Services. These writers number approximately 200 and are responsible for writing procedural material. The material interprets federal and state requirements defining eligibility and obtaining services from this government department. The writers were identified by their assignment to units responsible for preparation of manuals. This information was verified by review of their job descriptions from personnel files. Training data show that the average writer is male, age 38, having worked eight years in his present job. As full-time writers, they spend approximately 30% of their time writing new policy and procedures, 60% rewriting material to conform to changes and 10% doing nonwriting administrative duties.

The Sample

The sample is composed of writers from four groups within the department and working in Lansing, Michigan. Each treatment group was one which had responsibility for writing for a series of manuals. Group One, the control group, contained thirty-five writers from two departmental units, personnel and training. Group Two was composed of twenty-five writers who write the Administrative Manual series. This manual contains policy and procedures for operation of the department itself. Group Three was composed of thirty-three writers responsible for the Services Manual series. These manuals covered such services as foster care, adoption, and child abuse. In Group Four were writers

from the Assistance Payments section. These twentyeight writers supplied policy and procedure statements for the manual series dealing with direct payment of monies to welfare clients. None of the groups gained any new writers during the study. Two groups did lose one group member each. These slight changes are assumed to have no effect on group scores.

Hypotheses

Legend:	M ₁ =	=	Mean	for	no	train	ning	no	KR	group
	^M 2 [±]	=	Mean	for	no	train	ning	KR	gro	oup
	M ₃ =	-	Mean	for	tra	ained	but	no	KR	group
	M ₄ =	=	Mean	for	tra	ained	with	n KI	R gr	oup

Null Hypothesis 1: There will be no difference in readability level of writing between the group receiving training only and the group receiving both training and knowledge of results at each measurement point.

 $H_{0}: M_{3} = M_{4}$

Alternate Hypothesis: The group mean scores on readability of writing between the group receiving training and the group receiving training and knowledge of results will differ at each measurement point. The difference will be that the group receiving both treatments will show a lower mean readability level over each measurement point. Null Hypothesis 2: There will be no difference in readability level of writing between the group receiving knowledge of results only and the group receiving both knowledge of results and training at each measurement point after the first month.

$$H_0 = M_2 = M_4$$

- M₂ = mean score for knowledge of results
 group
- M₂ = mean score for trained plus KR group

Alternate Hypothesis: The group mean scores on readability of writing between the group receiving knowledge of results and the group receiving knowledge of results and training will differ at each measurement point after the first month. The group mean score for the KR group will drop showing improved readability but will not equal the readability level achieved by the group with both treatments (training and KR).

<u>Null Hypothesis 3</u>: There will be no difference in readability level of writing among the group receiving only knowledge of results, the training only group, and the group receiving both knowledge of results and training at each measurement point after the first month.

 $H_0 = M_2 = M_3 = M_4$

Alternate Hypothesis: The group mean scores on readability of writing among the three treatment groups will differ at each measurement point after the first month. Null Hypothesis 4: There will be no difference in readability level of writing between each of the treatment groups and the control group at each measurement point after the first month.

$$H_{O} = M_{1} = M_{2}$$
$$M_{1} = M_{3}$$
$$M_{1} = M_{4}$$

Alternate Hypothesis: The group mean scores on readability of writing between each of the treatment groups and the control group will differ at each time point. The scores will show that although the control group and the KR group begin the experiment with essentially similar mean scores the mean score of readability for the KR group will drop after the first measurement point. The trained group's mean score will begin at a level much lower than the control group's but will rise to a point close to the control. The mean score of the group receiving both treatments will remain essentially at its start level throughout the experiment.

Design

The design used in this research is quasiexperimental. It is best labeled as a nonequivalent control group design.¹ It involves three experimental groups and a control group. The groups do not have true

¹Donald T. Campbell and Julian O. Stanley, Experimental and Quasi-Experimental Designs for Research (Chicago: Rand-McNally, 1963), p. 47.
pre-experimental sampling equivalence. Rather, the groups constitute naturally assembled groups of writers working for the Michigan Department of Social Services in Lansing.

Despite the unmatched nature of the groups, the control for internal validity is increased in proportion to the confirmed similarity between groups. That similarity is confirmed by use of readability scores, reading level scores and the characteristics of age, sex, and work experience (Table 2).

With confirmed similarity this design can be regarded as controlling the main effects of history, maturation, testing, and instrumentation.² The question of how generalizable this design is, is answered by its external validity. Although the threat to external validity represented by reactive arrangements is present, Campbell and Stanley state that it is present "to a lesser degree than in most true experiments."³

Analysis

The data were analyzed by comparisons of central tendency and variability. The group means, taken once each month for six months, are used as the basic observations. These observations were plotted and compared. An

²Ibid., p. 50. ³Ibid.

Analysis of variance was conducted using means averaged over all months. A correction factor was also applied to account for the violation of the assumption of independence of scores within groups.

Despite the weaker nature of a "self-selected" design its alternatives are often found to be more reactive, thus creating more awareness of an experimental condition among subjects. This design has proved to be very popular, with Campbell and Stanley calling it "one of the most widespread designs in educational research."⁴

The Form of Knowledge of Results

The hard copy computer printout from the STAR program provided knowledge of results to individual writers. The printout was sent back to the writing unit and then distributed by supervisors to the writer responsible. It was this knowledge of results of prior work which provided the writer with information on whether he or she was meeting the readability guidelines of between 10 - 14 (the ideal being grade level 12). Shown in Figure 2 are instructions for interpreting the printout which served as the knowledge of results. On the reverse side of the instructions was a sample printout of text and analysis (Figure 3).

⁴Ibid., p. 47.

HOW TO USE THE MATSTAR PRINTOUT

The MATSTAR printout is divided into three parts: 1) The analysis giving grade level equivalent and other data, 2) the list of words in the passage containing three or more syllables, and 3) the sample passage which was analyzed.

1. The Analysis

The most important part of the printout is the analysis located at the bottom. Look at three things: grade level equivalent, average sentence length, and average syllables per word.

The <u>Grade Level Equivalent</u> gives the grade level to which the sample is written. For example, a grade level equivalent of 18.7 means that the passage is written for someone with an education level of between 18 and 19. If your intended readers are Assistance Payment Workers with a minimum of a 12th grade education, you have a good indication that the level of the writing is not matched with the level of the reader. Ideally, the level you should aim for is the lowest educational level for that particular reader group. Choosing the lowest level will best account for all persons in the group. If you do not know the lowest grade level for the group, a good rule-of-thumb would be 12 for employees and 9 for clients.

The Average (Avg.) Syllables Per Word tells you how hard your choice of words is. Syllable average should be less than 1.6. Check the words over three syllables for ones which are clear and simple. If you can't reduce the syllable count, try to reduce sentence length. A good rule-of-thumb is: the harder the words used, the shorter the sentence should be. The reader then gets some difficult words but in small and manageable amounts.

2. Words of Three or More Syllables

All the words of three or more syllables are listed. This lets you check for words which could be replaced with simple, clear words. A few misspellings or extra letters will not significantly change the results.

3. The Sample Passage

If the material you wrote is $1-1\frac{1}{2}$ pages, the entire text is analyzed. If 2-10 pages, two samples per page are taken. If 10 or more pages, two samples every other page are analyzed.

Typing errors may be introduced by the person who made input into the computer. A few misspellings, extra letters or syllables will not significantly change the analysis. Periods left out or inserted <u>can</u> significantly change the results.

Figure 2.--Computer Printout Explanation.

SAMPLE MATSTAR COMPUTER PRINTOUT

CONTENTS OF FILE:

WRITER'S SAMPLE ANALYZED

THE MATSTAR COMPUTER PRINT OUT GIVES FEEDBACK TO WRITER, AS A WRITER IT WILL HELP YOU TO TELL HOW WELL YOUR WRITING WILL COMMUNICATE. THE OUTCOME IS BASED ON HOW HARD A READER WOULD HAVE TO WORK TO UNDERSTAND WORDS AND SENTENCES. MATSTAR GIVES YOU THE GRADE LEVEL EQUIVALENT OF A WRITTEN SAMPLE. YOU CAN THEN COMPARE THE GRADE LEVEL OF YOUR WRITING TO THE GRADE LEVELS, THE GREATER THE CHANCE FOR MISUNDER-STANDING. ALL THE WORDS OF THREE OR MORE SYLLABLES ARE LISTED. THIS LETS YOU CHECK THE WORDS WHICH COULD BE REPLACED WITH SIMPLE CLEAR WORDS. A FEW MISSPELLINGS OR EXTRA LETTERS WILL NOT SIGNIFICANTLY CHANGE THE RESULTS. THE AVERAGE (AVG.) SYLLABLES PER WORD IS AN INDEX OF WORD DIFFICULTY. CHECK THE WORDS OVER THREE SYLLABLES FOR ONES WHICH ARE CLEAR AND SIMPLE. IF YOU CAN NOT REDUCE THE SYLLABLE COUNT THEN TRY TO REDUCE SENTENCE LENGTH. A GOOD RULE OF THUMB IS THE HARDER WORDS USED, THE SHORTER THE SENTENCE SHOULD BE. THE READER THEN GETS SOME OF THE DIFFICULT WORDS BUT IN SMALL AND MANAGEABLE AMOUNTS.

SYLLABLES

- 3 COMPUTER
- 4 COMMUNICATE
- 3 UNDERSTAND
- 4 EQUIVALENT
- 5 MISUNDERSTANDING
- 3 MISSPELLINGS
- 5 SIGNIFICANTLY
- 3 AVERAGE
- 3 SYLLABLES
- 3 DIFFICULTY
- 3 SYLLABLES
- 3 SYLLABLE
- 3 DIFFICULT
- 4 MANAGEABLE

```
AVERAGE SENTENCE LENGTH = 18.9SHOULD AVERAGE 15-20AVG. SYLLABLES PER WORD = 1.5SHOULD AVERAGE LESS THAN 1.60GRADE LEVEL EQUIVALENT = 10.2SHOULD MATCH GRADE LEVEL OF READERS
```

Figure 3.--Sample Computer Printout which Served as Knowledge of Results.

The Training

The two trained groups received twelve hours of instruction on how to write clearly. This researcher conducted all the training sessions using a slide tape program titled "Put It in Writing."⁵ The total amount of instruction is present on slides and tape. This fact contributed to maintaining equal treatment of both groups.

The training was presented one-third by audiotape, one-third through workbook exercises and one-third through self and colleague critiques of their own writing. Trainees were required to write a final report using all clear writing techniques taught. That report also had to meet a readability score of no more than twelve. Only four of the total number trained were required to rewrite the final paper in order to meet completion criteria.

There is considerable experience with the "Put It in Writing" course. Since its introduction to state government it has been given to over 1,800 employees. The focus of the course is clear writing. Simplicity is stressed. Don't use a big word when a small one will do the job is one of the many rules.

⁵Joseph, op. cit.

Testing of Writers for Reading Level

The study included pretesting all the group participants for reading level. The Nelson-Denny Reading Test was chosen over three other tests to measure the reading level because of three factors: (1) the test had a greater grade range over which it could measure reading level, (2) the resultant score was given as a grade level equivalent (GLE) and therefore could be equally compared to the writing level measure also expressed as GLE, and (3) the Nelson-Denny test had been used by the department in its Rapid Reading Course. It was felt that the scores dating back two previous years would be helpful in comparison.

The reading test did show its own inherent limitation. It did not identify grade level equivalent scores above level 15. This was also true for the other tests. Because of this limitation, it is not possible to identify the upper limits nor to average scores for comparison.

The results of this testing showed that all writers were reading at the tenth grade level or above with 71% reading above the fifteenth grade level (see Table 1).

Testing of Writers for Writing Level

Prior to training, all writers were tested for writing level. All tested above grade level 12 with 74% writing above grade level 14 as determined by analysis of the writings of each unit over a three-month period prior to the start of the study. The average GLE for each group was 16.8, 16.4, 17.4, 16.9 for groups 1-4 respectively. Those averages show no meaningful difference between groups in the sample.

Measurement of Writing Level

The measurement of writing level was critical to this study since it resulted in the means for providing knowledge of results. That measurement was provided by STAR (Simple Test Approach for Readability). STAR is the name for a computerized readability evaluation program designed by the Service Research Group of General Motors Corporation. It has been used by GM for more than three years. It insures that their owners and service manuals match the reading levels of their intended readers. GM makes the STAR program available to the public at no cost.

The basic formula used in the STAR program is the one developed by Rudolf Flesch in 1946. Flesch defined reading ease (RE) as RE = 206.835 - .846 WL - 1.015 SL, where WL (word length) is the number of syllables per 100

words, and SL (sentence length) is the average number of words per sentence. The number counts result in a readability index range of 0 to 100. The STAR program then translates the Flesch index to years of education (grade level equivalent).

The sensitivity between sentence length and syllable count in determining the grade level equivalent is shown in Table 3.

Average Sentence Length	Average Syllables Per Word	Grade Level Equivalent		
30	1.5	14.3		
30	1.6	15.6		
30	1.7	16.9		
25	1.5	13.5		
25	1.6	14.8		
25	1.7	15.1		
20	1.5	11.5		
20	1.6	12.7		
20	1.7	13.9		

TABLE 3.--Sensitivity Between Sentence Length and Syllable Count in Determining Grade Level Equivalent.

Given a one hundred word passage with five sentences and an average of 1.5 syllables per word the resulting grade level equivalent would be 11.5. Holding the number of sentences constant, adding ten more syllables (1.6 syllables per word) would result in a grade level equivalent of 12.7. In comparison holding average syllable count constant an increase in sentence length resulting in an average of twenty-five words per sentence would increase the grade level equivalent from 11.5 to 13.5. The increased sensitivity of syllable count is also evident in an example found in Appendix A and reviewed in Table 7.

STAR was ideal for use in this project. Besides its free price tag, it is a simple program and is easily adapted to a variety of inputs including batch, paper tape, and on-line. It also is available in all the basic computer languages. Probably its best feature is its simplicity of use. Anyone who can operate a typewriter keyboard can adapt to the six user commands with half an hour's practice. In this research all writing samples were verified (keyed again) by a second operator to reduce the probability of error.

The STAR program provided a fast and efficient means to analyze the hundreds of pages of written material. Although readability can be calculated by hand, the added time and reduced accuracy make it prohibitive for large volume input.

For at least the last eight years, all written manual material has undergone a review process prior to its final publication. This limited review includes checks by appropriate departmental administrators for information accuracy. No check or critique is made of clarity, although spelling errors are noted. The review process was used as the avenue for analyzing the written material in this study. One review copy was submitted for review. After analysis the computer printout of the results returned to the writing unit. This served as the knowledge of results for the study. All written samples and analyses of them were saved in computer memory. They were then periodically retrieved from the system. This became the raw data providing information on group progress. It also provided a double check on input accuracy. The system was purged about every three weeks to erase the text from memory.

CHAPTER IV

ANALYSIS OF RESULTS

Introduction

The results of this research are presented in three sections. The first presents the data that pertains to each of the stated alternate hypotheses. The second section covers the statistical treatment of the data that enables rejection of the null hypotheses. The third section presents a separate indepth analysis of each group.

In this research, analysis of data is accomplished by comparison of central tendency (mean) and variability (variance) of each of the four groups. Computation of both was accomplished by the Statistical Package for the Social Sciences (SPSS) computer program. SPSS was also programmed to draw the plots of the group means.

It was expected that although each started with the same scores the group receiving training (Group Three) and the group receiving training and knowledge of results (Group Four) would differ following training. It was hypothesized that the group given knowledge of results (KR) would main group mean scores on readability of writing within the acceptable limits of 10 - 14 grade

level equivalent (GLE) while the group without KR would not.

A difference in scores was apparent between Group Three which received training and Group Four which received training and knowledge of results. Figure 4 identifies the mean scores on readability at each time point for both these groups.



Figure 4.--Plot of Group Mean Scores on Readability Levels of Writing for Groups Three and Four.

While both groups started the experiment with essentially the same readability level score in January, all succeeding months showed a difference of at least four grade levels. Between January and March both groups increased in mean readability level. In that two-month period Group Three, the training only group, increased its writing difficulty level by six grade years of education (from 11.4 to 17.3 GLE). In the same period the training and knowledge of results group increased by only two and one half years (from 11.7 to 13.3 GLE). The sharp increase for Group Three was accounted for in their lack of knowledge of results. See "Analysis of Group Three Data" in the next section for a more detailed discussion of this increase.

At each of the other measurement points both groups showed differences in both mean readability score levels and in the fluctuation pattern for those scores. For the last three months of the experiment Group Three, receiving knowledge of results, shows almost no change in mean score (17.35, 17.02, 16.95, 17.04). The training and knowledge of results group shows a meaningful fluctuation during the same period (13.3, 11.0, 11.5, 11.9). Both of these differences and fluctuations are discussed in the next section.

Figure 5 shows the resulting mean scores on readability for Group Two, the knowledge of results group, and Group Four, the training and knowledge of results group.

Although both groups started at different readability levels (untrained at 17.6 and trained at 11.7) it was hypothesized that knowledge of results for Group



Figure 5 .--Plot of Group Mean Scores on Readability of Writing for Groups Two and Four.

Two would have a positive effect on reducing readability level. It was anticipated that this effect would lower the readability level of Group Two but not to the same extent achieved by the trained group.

Group Two reduced its readability level dramatically over the first two months (17.6 to 14.8). But subsequent measurements (14.3, 14.2, 15.1) indicate that with only knowledge of results they did not achieve readability levels reached by the trained group with knowledge of results (Group Four). While the lowest GLE reached by Group Two was 14.2, Group Four averaged 11.9. Further discussion of Group Two data will be found in the next section.

Figure 6 shows the group means for each of the treatment groups. It was hypothesized that each would have a different plot with respect to maintaining an acceptable readability level. Examination of each of the plots of group mean scores on readability shows that they do differ. The period between March and June shows a difference of between two and seven grade levels difference among the groups.

Table 4 identifies the differences in scores between the three treatment groups (nos. 2-4) and the control group (no. 1). It was hypothesized that each of the treatment groups mean scores on readability would



Figure 6.--Plot of Group Mean Scores on Readability Levels of Writing for Groups Two, Three, and Four.

differ from that of the control group which received neither training nor knowledge of results.

TABLE 4.--Group Mean Readability Level Scores* for Each Measurement Point (Scores Rounded Off to Nearest Tenth).

Measu	reme Mo:	ent Point nth	l JAN	2 FEB	3 MAR	4 APR	5 MAY	6 JUN
Group	1,	No T/No KR	16.6	17.6	17.2	16.4	17.4	18.0
Group	2,	No T/KR	17.6	16.6	14.8	14.3	14.2	15.1
Group	3,	T/No KR	11.4	14.9	17.3	17.0	17.0	17.0
Group	4,	T/KR	11.7	12.0	13.3	11.0	11.5	11.9

*Given as a grade level equivalent.

Group One mean score readability levels of 16.6, 17.6, 12.2, 16.4, 17.4, 18.0 show fluctuations within a range of what could be expected for a group which is unaffected by training or knowledge of results. Further analysis of Group One data is discussed in the next section.

Hypotheses

The following null hypotheses were the subject of experimental treatment.

- Null Hypothesis 1: There will be no difference in readability level of writing between the group receiving training only and the group receiving both training and knowledge of results at each measurement point.
- Null Hypothesis 2: There will be no difference in readability level of writing between the group receiving knowledge of results only and the group receiving both knowledge of results and training at each measurement point after the first month.
- <u>Null Hypothesis 3</u>: There will be no difference in readability level of writing among the group receiving only knowledge of results, the training only group, and the group receiving both training and knowledge of results at each measurement point after the first month.
- Null Hypothesis 4: There will be no difference in readability level of writing between each of the three treatment groups and the control group at each measurement point after the first month.

Table 5 presents the results of an analysis of variance of the data to determine whether any of the groups differ significantly from any other. The analysis of variance identifies whether the independent variable, time, has had an effect on the dependent variable, grade level equivalent. It also identifies the "equality" of the means of the four experimental groups. Any relation that exists between the independent and dependent variables will be reflected in the inequality of means. The

and Group.						
Source	Sum of Squares	df	MS	F	Sig. Level	
Main effects	3550.17	8	443.77	120.21		
Time	194.24	5	38.48	10.52	.001	
Group	3366.08	3	1122.03	303.95	.001	
Two-way interaction Time x group	1307.50	3*	87.17	23.613**	.001	

TABLE 5.--Between Group Analysis of Variances on Variables of Time and Group.

Greenhouse-Giezer correction factor $\frac{1}{\# \text{ groups } -1} = \frac{1}{6 - 1} = \frac{1}{5}$

*Corrected for non independence of scores within groups.

**Tabled value of F = 2.60.

more unequal the means, the wider apart they are the higher the relation, other things being equal. The statistical test, F test, confirms the visual observations of the group plots.

Looking at the main effects variables of time, group and combined interaction the resultant F value is larger than the tables value identifying a statistically significant difference between groups. All are significant well beyond the .05 level. Note (*) that the two way interaction between variables has been corrected for non independent groups. With each hypothesis there is a statistically significant difference supporting the rejection of the null hypothesis and acceptance of the alternative.

Analysis of Group One Data

Group One was composed of writers whose writing was analyzed, but who received neither training nor knowledge of results. Because their writing was analyzed without their knowledge, the effects resulting from interaction effect of testing and reactive effects of experimental arrangements are essentially eliminated.

The mean readability scores varied from a low of 16.4 to a high of 17.9. This group also had the highest variability of within group scores (see Table 5).

Analysis of individual writing scores shows them to range from 12.3 to 24.4. This fluctuation in mean

grade level equivalent scores compares favorably with writing samples from the entire population covering the six-month period prior to the conduct of this study.

Analysis of Group Two Data

Group Two was composed of writers who, like those in Group One, received no training. Their initial writing level was essentially the same (17.65 vs. 16.66 for Group One). Analysis of individual scores for measurement period one for both groups showed no meaningful difference. After that point both groups began to differ in mean grade level equivalent and variability of scores.

Following measurement one, Group Two mean scores take a shift downward for two months. They then leveled off before taking an upward turn just before the last measurement (Figure 7).

Goal setting theory is a plausible explanation for the performance improvement of group two. That theory states that KR in combination with a hard goal will have greater effect than without. Group Two had a sufficiently hard goal to go from 17.6 mean score on readability to 13. No other group had knowledge of such a large distance from present performance to reach the goal.

The failure of this group to improve beyond its lowest mean score of 14.2 can be attributed to its lack



Figure 7.--Plot of Group Mean Scores on Readability Levels of Writing for Group Two.

of training. Without the clear writing techniques presented to the trained groups, Group Two writers do not have the tools to do better than they did. Support for this can be seen when comparing an original writing sample and one rewritten using the knowledge of results computer printout. Appendix A is such an example of two written samples from the same writer within the knowledge of results group. The analyses are contained in Table 6.

	Average Sentence Length	Average Number of Syllables Per Word	Grade Level Equivalent	
Original Time Point 3	24.7	1.5	12.7	
Rewritten Time Point 4	15.7	1.6	12.5	
Difference	- 9.0	+0.1	+ 0.2	

TABLE 6.--Comparison of Sample Original and Rewrite from a Single Writer at Two Time Points.

This Group Two writer given knowledge of results was only able to reduce readability by 0.2 GLE. The resulting reduction in average sentence length was a much greater change (down 9.0 GLE). The small reduction in overall readability level was a result of an increase in syllable count. A reasonable explanation is that without training writers can reduce sentence length. But identifying simpler words (with fewer syllables) requires training.

The upturn in mean scores shown between measurement points 5 and 6 in Figure 7 cannot be explained given present data. The answer lies in replication and extension of the experimental period.

Analysis of Group Three Data

Group Three writers received clear writing training but were not given knowledge of readability level of their writing. As was expected in a previous study performance of these trained writers decreased. The resulting mean grade level score was close to the score before training (16.8 before, 17.4 two months after training). The reduced performance was anticipated; the rate of that reduction was not. Instead of the mean grade level equivalent gradually increasing over the entire six months, the change takes place within the first two months. After that time the scores remain almost constant through the last three months (Figure 8).

One might expect that once the peak GLE mean was reached the pattern would follow that of Group One, but it does not. The primary difference between Group One and Three is the degree of fluctuation of scores. One



Figure 8.--Plot of Group Mean Scores on Readability Levels of Writing for Group Three.

factor that may account for the resulting lack of fluctuation in mean scores is syllable count. Examination of individual printout records identifies the fact that that group maintained a relatively stable syllable-perword count, while the average sentence length sharply The latter accounted for the increase in increased. grade level equivalent score. Group One, by comparison, had a greater degree of fluctuation in both sentence length and syllable per word. Fluctuation of both variables is normal in situations where the subject is not conscious of them. The trained writer is initially conscious of both. In this researcher's training experience with writers from this population, they are most concerned with overly difficult words. It may very well be that they maintain low syllable counts out of choice and yet lose awareness of sentence length.

Analysis of Group Four Data

Group Four was the group which was trained and then provided ongoing knowledge of results. It was hypothesized that Group Four would maintain its post training readability level with little fluctuation. Analysis of the group means shows that the group's performance was maintained but with fluctuations greater than anticipated (see Figure 9).



Figure 9.--Plot of Group Mean Scores on Readability of Writing for Group Four.

The sharp increase in score between the second and the third month for the T/KR group (no. 4) dips sharply the next month. This sharp reduction in performance followed by a marked increase in performance can best be explained through reference to knowledge of results and goal setting theory.

That theory, as pioneered by Edwin Locke, identified the effect of goals on performance. His experiments showed that the level of effort on a task is determined largely by the subject's conscious performance

goals. Locke found that, to be meaningful, the goals must also be sufficiently hard as to pose a challenge to the subject.¹ To apply this theory to Group Four would be to first identify the goal and how difficult it was to achieve. The goal for both knowledge of results groups (nos. 2 and 4) was to achieve a grade level equivalent readability score of twelve. The difficulty in achieving that level depends on the group's distance from the goal and whether they have the knowledge and skill to achieve that performance level identified by the goal. Having just completed training by the start of the experiment, Group Four was already writing at level 11.7. During the first month of the experiment, Group Four was performing at the desired goal level. According to goal theory experiments, knowledge of results without a sufficiently difficult goal will result in a lower level of performance. Group Four's decrease in performance, shown by the increase in readability score, can be explained by the lack of a difficult goal.

The turning point at month three can be explained as that point where performance level is such that the goal becomes meaningful. At month three the readability score is 13.32. At that time their performance has decreased to such a point that they can consciously be

¹Locke, op. cit., p. 324.

aware of their decrease in performance. The goal of twelve then becomes meaningful, which results in an increase in performance skills and a decreased readability level in month four. Added support for this explanation is found by observing the decrease in variability of scores at month three with a subsequent increase the following month (Table 7). This same phenomenon looks as though it might be repeating itself with the decrease in performance following month four.

TABLE 7.--Within Group Variances in Grade Level by Time Periods.

Group No.	Treatment	Month						
		1	2	3	4	5	6	
1	NO T/NO KR	5.11	5.69	5.51	5.85	5.66	6.49	
2	No T/KR	5.82	3.00	2.03	1.12	1.84	2.56	
3	T/No KR	1.35	4.84	4.75	4.05	5.07	5.13	
4	T/KR	.89	1.96	1.12	1.88	1.44	2.31	

Summary

The data collected during this study enabled rejection of the three null hypotheses and acceptance of the alternatives. In general the findings were as expected (see Figure 10).





- The training only group reverted back to the mean grade level held prior to training.
- The knowledge of results group made meaningful reduction in difficulty of writing.
- The trained group receiving knowledge of results maintained their readability level following training.
- The control group receiving neither training nor knowledge of results maintained their high readability level as predicted.

The most meaningful results of this research were (1) identifying the apparent goal setting effect over time experienced by Group Four, and (2) the rate of loss of writing skill experienced by Group Three. The significance of these for this and other future studies is reviewed in Chapter V.

CHAPTER V

SUMMARY AND CONCLUSIONS

Purpose of Study

In 1974 the Michigan Department of Social Services began a program to train its manual writers in skills that would enable them to select words and compose sentences that would be clear to their readers. Tests immediately following the training demonstrated that indeed the trained writers could match their writing to their intended readers. In 1976 an evaluation was conducted of fifty writers who had completed training four to six months earlier. The idea was to confirm the effect of the training. Some reduction in performance was expected but not to the degree that was found. The finding showed that as many as 68% of the writers were writing at pre-training levels. Immediately following training they were writing at a twelfth grade level equivalent, considered appropriate for employee readership. After six months most of these writers were writing at levels between 16 and 19. It was apparent that the training was not paying off.

Shortly after this, two hundred samples of departmental writing were analyzed for readability. The samples ranged from level 17 to level 21. Subsequent

reading level tests of employees and clients (Table 1) showed that between 20%-100% of certain groups had reading abilities below the reading level of written material. The extent of the distance between readability level and reading ability was unfolding. Departmental manuals and other written communications were being written above the level of their intended readers. With over two hundred full-time writers and over twelve thousand employee readers there was a significant possibility of mass miscommunication.

The effects of difficult writing within the federal social welfare system came to light with publishing of the Report of the Commission on Federal Paperwork. That report estimated that unclear welfare forms and instructions were costing the federal government six billion dollars a year. Since those same federal forms and instruction form the basis for Michigan's welfare system it was reasonable to assume that there was waste attributed to written material of Michigan's Department of Social Services. The report also indicated that, besides imposing unnecessary hardship and confusion on readers, difficult writing conveys feelings of indignity and appears demeaning to the reader. If that wasn't reason enough, the study found evidence that such writing invited welfare fraud.

Internal audits of state welfare case files showed such a high percentage of errors that Michigan's

federal share of the welfare cost was being cut as a penalty. Make your written material easier to read was the message being given to the Department of Social Services by the state legislative committee that oversees the department's affairs and appropriates its funds.

The result was a program to assess readability of all manual and client material within the department. Manual material included instruction and procedures for case workers, and client material included forms and eligibility criteria sent to the clients and public. This research study was an outgrowth of that readability assessment project.

The research was undertaken to identify an effective means to maintain a writer's clear writing skills after training. Another question to be answered was the extent to which knowledge of results could achieve acceptable levels without the training component.

Major Summary

The research had two treatment components-training and knowledge of results. Training for writers included a twelve-hour clear writing course titled "Put It in Writing." Knowledge of results was provided by computer readability analysis of writers' work. This analysis included a measurement regarding average

sentence length, average syllables per word, and average grade level equivalent. Opposite those three measures was a goal which would insure that the writing would be at a difficulty no greater than what the reader could understand.

A quasi-experimental design was chosen as most appropriate for use on intact groups. Four writer groups were identified and assigned to the following treatments.

- Group 1 A control group of thirty-five writers given no training and no knowledge of results.
- Group 2 A group of twenty-five writers given knowledge of results.
- Group 3 A group of thirty-three writers given training in clear writing.
- Group 4 A group of twenty-eight writers given both training and knowledge of results.

Threats to internal validity were controlled through confirmed similarity between subject groups (see Table 2). Although the threat to external validity represented by reactive arrangements is present, Campbell and Stanley state that it is present "to a lesser degree than in most true experiments."¹

Knowledge of results is defined as information as to whether or not a response is correct and whether

¹Campbell and Stanley, op. cit., p. 50.

progress or improvement are being made. There is general agreement among theorists that knowledge of results can improve performance of a task. A good number of those theorists credit the internalization of difficult goals as the determining factor. This research provides writers with both knowledge of results and a goal. For the untrained knowledge of results group writing at level 17 a goal of 12 is assumed to be sufficiently difficult. The trained group receiving knowledge of results began at the level of 12.

The design limitations included lack of random sampling and assignment to groups because of use of intact groups. Evidence to confirm similarities between samples and between groups was used to strengthen the design. Inability to identify individual writer samples prevents checking of individual progress and amount of writing produced.

Because the researcher cannot identify which persons received more than one analysis during each time period, independence within groups cannot be confirmed. However, analysis of the number of samples submitted per group per time period indicates that all writers wrote at least one sample (and therefore received knowledge of results) during each time period. The rates of samples submitted per month (Table 2) indicate that less than one fourth would have received more than one sample per period.
Information gathered from trained writers and supervisors from each group indicates that it was unlikely that any writers submitted more than two samples per time period. This experiment assumes that both amount and use of knowledge of results were equally distributed between groups. In performing an analysis of variance the Greenhouse-Giezer correction factor was applied to account for this non independence of some scores within groups.

Conclusions

Keeping the above limitations in mind the following conclusions are drawn.

Group One, the control group which started at a grade level equivalent of 16.6 fluctuated up one grade level at month two and down to 17.2 the next. It reached a low of 16.3 at month four and then climbed to 17.4 and ended at 18. Given measurements of writing for the one year preceding the study these fluctuations were normal.

Group Two was also untrained, but its subjects did receive knowledge of results. The knowledge of results on their writing performance, coupled with a sufficiently hard goal, resulted in a meaningful improvement in readability from a beginning mean score of 17.6 to 16.6 the next month and to 14.7 the third month. The leveling off during the next two months can be attributed to a lack of skill to further improve on readability. For the knowledge of results group (Group Two), knowledge of scores and polysyllable words provides only limited help when compared to the skills available to a trained group. The increase in reading difficulty level from 14.2 at month five to 15.1 at month six is more difficult to explain. It may be a result of reaction to an inability to meet the set goal. Or it may be the beginning of a normal fluctuation for this group.

Group Three, the training only group, decreased its performance level as expected. But the surprise was the rate at which it occurred. This group went from a post-training level of 11.4 to 15 the next month and peaked at 17.3 the third. It reached its pre-training level in 60 days. This reduction in performance gives support to those who have identified significant pressures on writers to conform to the difficult writing which may be the norm (see Appendix B). This group also shows almost no change in mean readability score for the last three months.

Group Four was composed of writers who were trained and received knowledge of results throughout the experiment. The data showed overall performance maintenance during the experimental period as hypothesized. The most meaningful observation was a decrease in performance between month two and three followed by a sharply increased performance level the next month. The

best explanation for this fluctuation comes from goal setting theory (compare with Locke's results, Figure 1). That theory says that for knowledge of results to be effective there must be a hard goal associated with it. Having just completed training, Group Four was already writing at level 11.7. At this level there is no goal. The result is that KR has no effect until at point three performance has decreased enough to result in a difference between present performance now 13.3 and the goal of 12.0. Now the writer can set a meaningful goal. At the next time point a group mean of 11.5 is achieved. The goal is reached. Then the direction changes with a decrease in performance.

Implications Gathered from Results

Two major implications have been identified. One implies that knowledge of results alone is a reasonable alternative. The other implies that self administered knowledge of results in combination with external KR may provide better performance than either one alone.

Although the size of the fluctuations was found to be consistent with that of the total population under similar circumstances, how were they caused? The implication is present that without a performance goal and a measure of performance a writer would have no reason to pay attention to how they expressed their ideas. Content

would be important, but how one goes about communicating content would not necessarily be important. In fact, my experience in training writers is that very few are aware that there is a difference between what we say (content) and how we say it (words and sentences). Without that separation words and sentences are content. This belief is more apt to lead to such behavior as fulfilling a writing requirement by finding a similar document in the file and changing a few words and figures. Such behavior most often shows little regard for how well the writing communicates.

Providing writers with a means to measure their own performance and a significantly difficult goal is a reasonably effective means to improve readability levels without providing training. The results of Group Two data show that even without training knowledge of results provided performance improvement. Although the improvement gained in this study did not meet the goal, the fact that the goal remained sufficiently difficult may account for the evenly maintained performance level over the last three months.

Providing writers with opportunities to measure their own performance (and provide KR to themselves) is an option to consider. Though not investigated in this study, writers in classes previous to this experiment were given a copy of the Fry Graph (see example in

Appendix C). This device is a quick and easy way to figure readability level of writing without the calculations required by other formulas. The resultant score is in grade level equivalent units. Although follow-up showed that very few writers used the device on their own after training, its use could be increased when combined with external KR. The combination is an alternative to increasing the frequency of external KR provided. Such an option would not increase the costs for externally provided KR but would provide the opportunity for writers to measure the readability of material while in the draft stage. While KR provided solely by the subject may not be effective, the combination with external KR should increase its use. This option is presently being incorporated into the readability improvement program of the Michigan Department of Social Services.

Implications for the Michigan Department of Social Services

The Michigan Department of Social Services was an ideal agency in which to initiate this study. With over 12,000 employees it is the second largest in state government. And it has one of the largest numbers of writers with over two hundred devoting fifty percent or more of their time on the job to writing. The readership is the largest of any department. Besides its employees, readers include several thousand nondepartmental

employees who are providers of services such as private social workers, local government workers, professionals in health care, and others. Including adult clients who use the direct services of the department, the total number of readers of departmental materials approximates one-fifth of the adult population of Michigan.

The policy and procedural writings of the department are crucial to running foster care facilities, treating Medicare and Medicaid patients, providing persons with direct payment of money and emergency food. A single written document dealing with the eligibility requirements for receiving welfare aid is sent to over 6,000 social workers and has impact on 100,000 clients. The large number of readers and significant impact in determining whether or not services are provided make that single document a very powerful communications device.

Training writers in the department was well underway when it was discovered that readable writing levels were not being retained over an extended period. When the knowledge of results writing maintenance program was suggested it was seen as an inexpensive means toward resolving the problem of loss of writing skill over time.

The results of the reading and writing tests and the six months of training and KR maintenance have already had their effect on decision-making. As a result of the tentative results available in July, 1978, the department administration decided to continue the sending of computer readability analyses to writers. In addition they decided to institute a management report to the directors of writing units to give them knowledge of how well their writers are doing. The goals as set forward in the project have been established as standards for all departmental writing. Internal departmental writing can be no higher than a 14 grade level. Grade level 12 was set as a maximum for material leaving the department for reading by the general public. And grade level 9 has been established as the highest level for client-read publications.

Another good sign is the formation of guidelines dealing with forms development, and publications and manual design. This project has also received good publicity in the media. It has been written about in several editorial columns (see Appendix D) and was featured in a television documentary. Several government agencies in other states have indicated their interest in using the project model. And the developer of the training program is considering additions to his material based on these findings.

The benefit for the Michigan Department of Social Services has been and will be threefold:

- A greater awareness by employees, administrators, and the public of the significance of readability in writing.
- 2. A training and skill maintenance program which has been successful in doing what it set out to do.
- 3. An opportunity to benefit other agencies and large organizations through providing data on a successful application of knowledge of results to improve on writing skill and maintain readable writing after training.

Recommendation for Further Research

This section contains six recommendations for further research on the effects of knowledge of results and goal setting on the writing task:

- Use random sampling and assignment
- Extend time period
- Add a group given KR without a goal
- Increase the frequency of KR
- Add a group with self administered KR
- Study effect of KR on syllable count and sentence length

Use of intact groups had its advantages. The group effects were what would occur upon implementation of knowledge of results on similar intact groups. Knowledge of results was given to everyone in a particular working group minimizing interaction effects of selection biases. Repeating this experiment with randomization would be well warranted. It would permit use of more powerful statistical tests. Group equivalence reduces the effects of chance events.

The plots in Figure 10 show group means over the time periods may be cyclic in nature. The upturn of the Group One mean grade level equivalent at the last measurement point is an example. Would readability level have continued to rise given an extension of time? Or would it have gone down? The same question can be asked regarding Groups Two and Four. And for how long would the group means for Group Two have continued to remain flat? Those questions could be better answered by extending the experiment's time period to ten months.

In this experiment knowledge of results was provided with a goal. A question could be raised concerning how writers would have performed given a readability level without a goal. The effects on the no-goal/KR group were the same in all the experiments reviewed; inclusion of such a group should strengthen the findings.

Increasing the frequency of KR may very well result in more stable performance in the training KR group. To have done so would have required that all draft material undergo readability analysis. Increased frequency of KR would be more realistic for occasional writers of letters and memos where the frequency of

writing is greater than manual writers, who like report writers, must research their writing. Increased frequency might be better researched using another task more repetitive in nature.

Self administered knowledge of results in combination with external KR is one way of increasing the frequency. Most readability formulas are time consuming to compute because they require at least counting the syllables and sentence lengths. Fry's Readability Graph is a good alternative.¹ It will result in a grade level equivalent by simple addition. (See sample, Appendix C.) The Readability Graph issued to writers during previous training sessions was found not to be used by fewer than ten percent. In combination with external KR wider use should result.

After having reviewed experiments used to test the effects of goal setting, I feel that writing is particularly well suited. Locke's experiments used tasks such as addition or reaction time. These are subject to the effects of testing. Repeated testing of subjects' reaction times resulted in raising performance. Subjects who know they are being tested perform differently from those who do not know. Writing is a skill

^LE. Fry, "A Readability Formula that Saves Time," Journal of Reading II (April 1968):514.

which lends itself well to "observation" without letting subjects know. And there is no evidence that repeated practice alone improves writing performance.

Another effect worth studying is the interaction between syllable count and sentence length in determining the grade level equivalent. As identified in Chapter IV (Table 5) both variables are not equal in effect. A design for such a study might include providing training to one group on the skills to reduce sentence length. Another group would be trained using simpler words and a third group trained in both. Knowledge of results might provide information in only one of the areas, either syllable count or sentence length.

APPENDICES

•

.

APPENDIX A

ORIGINAL DOCUMENT AND REWRITTEN DOCUMENT

ORIGINAL DOCUMENT

.

ORDER TO BE ELIGIBLE FOR AD A LEGAL PARENT. VINUED ABSENCE OF A PARENT PARENTAL SUPPORT OR CARE UN	C, A CHILD MUST BE D			
AUCHA TO BE ELIGIBLE FOR AU A LEGAL PARENT. NTINUED ABSENCE OF A PARENT PARENTAL SUPPORT OR CARE UN	C, A CHILU MUSI BE U			
PARENTAL SUPPORT OR CARE UP		ELKIVED OF THE CARE ON SUFFOR		
	DER THE FOLLOWING CIT	ES THE BASIS FOR DEPRIVATION RCUMSTANCES:		
WHEN THE PARENT IS OUT OF	THE HOME; AND	T DI		
OVIDER OF MAINTENANCE, PHYSI	CAL CARE OR GUIDANCE	FOR THE CHILD: AND		
THE ABSENCE, BECAUSE OF 11	S KNOWN LENGTH OR IN	DEFINITE DURATION, DOES NOT		
Y ABSENCE HAVING ALL OF THES	E CHARACTERISTICS CO	NSTITUTES DEPRIVATION OF PAREN		
L SUPPORT OR CARE. WHILE IT	MAY AID IN A DETERM	INATION OF DEPRIVATION, THE		
SENCE.	BE USED TO LIMIT THE	DEFINITION OF A CONTINUED		
RENT OUT OF HOME. A PARENT	IS CONSIDERED TO BE	OUT OF THE HOME UNLESS		
IS APPLIES TO PARENTS WHO LI	VE ELSEWHERE AND PAR	ENTS WHOSE WHEREABOUTS ARE		
IKNOWN. A PARENT MAY LIVE EL	SEWHERE AND STILL MA	INTAIN REGULAR CONTACT WITH		
TE CHILD.	allagate av on incl	TIGH OD TEDNINATION		
THE PARENT'S FUNCTIONING OC	CURS WHEN THE ABSENC	E RESULTS IN THE DISRUPTION OF		
LE PARENT'S ACTIVITIES IN PRO	VIDING MAINTENANCE,	PHYSICAL CARE, OR GUIDANCE FOR		
VINTENANCE MEANS PROVIDING FI	NANCIAL SUPPORT TO M.	EET THE CHILD'S NEEDS FOR		
DOD, SHELTER AND CLOTHING.	THE ANT IS NOT TOTAL	THE PARTY OF PROPERTY OF THE		
ICH AR FFFDING DRFSSING PRO	VIDING MEDICAL CARE	AND ATHER DIRECT. PERSONAL		
RAINING AND OTHER ACTIVITIES OCIAL DEVELOPMENT.	WHICH TEND TO PROMOT	E A CHILD'S PSYCHOLOGICAL AND		
VLLABLES	SALLABLES		SYLLABLES	
4 ELIGIBLE		CONTINUED		DEPRIVATION
3 PARENTAL		FOLLOWING TERMINATES	• •	CI RCUMSTANCES FUNCTIONING
3 PROVIDER	e	HAINTENANCE		PHYSICAL
4 INDEFINITE		DURATION		CHARACTERISTICS
4 DEPRIVATION		DEFINITION		CONTINUED
3 CONSIDERED		PERSONAL		PROPERTY
3 ELSEWHERE		INTERRIPTION		FUNCTIONING
4 INTERRUPTION		TERMINATION		FUNCTIONING
3 DISRUPTION		ACTIVITIES		PROVIDING
D MAINTENANCE		FINANCIAL		PUVSICAL
3 PROTECTION		PHYSICAL		PROVIDING
3 MEDICAL		PERSONAL		PHYSICAL
		ACTIVITIES		PSYCHOLOGICAL

REWRITTEN DOCUMENT

APPENDIX B

SOME OBSERVATIONS REGARDING WRITERS

SOME OBSERVATIONS REGARDING WRITERS

Without knowledge of results the writer is likely to take cues from other writing close at hand. The writer may think that because most of the other writing is written that way his or hers must be that way too. Taking that view to an extreme the organizational writer does not communicate as an individual, but as a functioning part of the whole. John B. Haney reached this view.¹ He sees the writer as only following standard operating procedures so that he/she will not unnecessarily disturb "some remote part of the organization." When it comes time to submit a report the writer gets a similar document from the file and just changes the dates, adds a few new sentences, and sends it on.

While Haney's observations offer an explanation why revisions of material are perpetuation in a heavy style, it does not account for newly written material. During the last four years this researcher has taught clear writing no one section has brought more resistance than having to identify oneself in writing. Writers avoid calling attention to themselves in writing. And

¹John B. Haney, "Readable Writing: Retrospect and Prospect," <u>Journal of Communication</u> 9 (1959):186.

heavy writing supports that style. The indirect expression "it is recommended that . . ." is an example of how the writer protects him or herself.

Heavy writing covers up weaknesses in the message. Albert Joseph, an industrial writing consultant and author of the training course used in this research, lists this as one of his reasons for heavy writing.² He says that writers who know their ideas are weak may hope to deliverately cover this up. He or she hopes that if the writing can be made to sound so complicated that no one can understand it, no one will recognize it has said very little. According to Joseph, the reason people rarely question the writing of others is that they have a profound respect for things they cannot understand. They would be embarrassed to admit they could not understand it.

²Joseph, op. cit., p. 8.

APPENDIX C

FRY READABILITY GRAPH

HOW TO TELL THE GRADE LEVEL OF YOUR WRITING A SHORT CUT FOR FIGURING THE FOG INDEX

. •

Determine the average sentence length (count the number of words in the passage and divide by the number of sentences). Next count the number of words containing three or more syllables contained in the first one hundred words. Do not count word endings such as -es, -ed or -ing in the syllable count. Also omit from the count capitalized names of people, places, programs and products. Then, lay a ruler or pencil across the two outside columns representing those numbers. The grade level of the writing appears on the scale in the center column.

Average Sentence Length (words)	Grade Level of Writing	Number of words per 100 words containing 3 or more syllables
15	5	ľ°
20	10	5
25	. 15	10
30	20	15
35	20	20
40	.25	25
	30	30

The higher the writing's grade level is above the reader's reading level, the more difficulty the reader will have reading that material.

Tips for improving the readability of your writing can be found on the other side.

APPENDIX D

NEWS CLIPPING REGARDING MICHIGAN DEPARTMENT OF SOCIAL SERVICES READABILITY PROJECT

Index tries lifting fog from writing

By JOHN H. O'BRIEN News Editorial Page Columnia

Having promised to deliver a system whereby clarity in writing can be tested, I now have second thoughts about the idea.

Mathematical issues would be a would be a solution of the start of a set respective careful and to a set of set respective careful and the set of the heavy hand of the back or coming down on, and cuting off, the sources of inspiration and imaginal respective careful and cuting the defined of the sources of style a faral blow.

ETML AS TRAVA A fromte is a promuse and you may deal with the System written by our J Gordon Schlecher, an in written by our J Gordon Schlecher, and Management berelopment accelarity of its Management Development accelarity of the Management of Schlecher and Management Development accelarity accelarity of the McMang Department of Social Services and also coordination and in Services and also coordination and in else are writting available to state office feature retriety dear?

His archie appeared in Outraed, a monthy publication of the Social Services Department, and describes the "Comming Fog Index." amed for the creater "Ruber of the service of the creater of the creater safe for ictar writing a sys goor writing fogs up desa and had here reserves the amount of fog in a prose sample. Schlecher gives an erample of the process by analyzing this own prose.

"To figure the Fog Index you must compute the average words per sentence and the percentage of 'polysyllables' in the



OHN H. O'BRIEN

passage. A 'polysyllable' is any word of three or more syllables, with three exceptions.

"First, proper names, no matter how mage Second, combination words itse borspond, combination words itse "borspond" or "whole smaller words itse "borspond" or "and the "borspond" of ed. es, or ing. When figuring the Fog index it helps to underime figuring the Fog index its the bor underime figuring the polysibable that meet the criteria."

Now then, targe the Cummg formula, we not the number of words in the passage and get 46. The average number of words per sentence is 14.3. The physiliable, words are "average, percentage, polysyliables, opsyliables, addition, polysyliables, addition, polysyliables, toon, syliables, addition, polysyliables, addition, pol

criteria... a total of 12 The polysyllabic content is threatore 139 percent. Add the 14.3 to the 13.9 and we get 32.2 Multiply this sum by 4 and get 11.26, which is rounded of to 11.3. This is the Fog Indervof the passade. According to Schleicher, the ideal in clarity in wring; as score between 10 and 12 on Gunning; 8 Index Anything between eight and 14 is accorptiable. Over 14, the writing is "unreadable"

Infallibility is not claimed:

"The Fog Index cannot measure all the characteristics of writing. Therefore, a good readability score is no guarantee of good writing. But a poor score, one above 14, is a good indication of poor writing."

All CARLCHER NUTRS HIVE states to a state to a state to the state to the state to the state to the state of the state of

Where had grow evenget 700 many polyyilables because it took seven of them to explain Schecher's position and procetors: "instructional development perceitar" and "coordinator" and instruction" and "available ... personnel did me in

OK. so I tried the first 102 words of Lincoln's Gettysburg Address. Old Abe scored a 14.2. Unreadable. BIBLIOGRAPHY

BIBLIOGRAPHY

- Association for Educational Communications and Technology, <u>Educational Technology: Definition and Glossary</u> <u>of Terms</u>, 1977.
- Bruner, Jerome S. The Process of Education. Cambridge, Mass.: Harvard University Press, 1960.
- Caffey, Musa B. "A Study of the Relationship Among Reading Levels of Practical Nursing Students' Readability Levels of Textbooks and Pass/Fail Scores on the State Board Examinations for Practical Nursing Students." Ph.D. dissertation, University of Southern Mississippi, 1977.
- Campbell, Donald T., and Stanley, Julian O. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand-McNally, 1963.
- Chall, Jeanne S. <u>Readability: An Appraisal of Research</u> <u>and Application</u>. Columbus, Ohio: Ohio State University Press, 1954.
- Church, R. M., and Camp, D. S. "Change in Reaction-Time as a Function of Knowledge of Results." Journal of Experimental Psychology 78 (1965):102-106.
- Coke, Esther U., and Rothkopf, Ernest Z. "Note on a Simple Algorithm for a Computer Produced Reading Ease Score." Journal of Applied Psychology 54 (1970):209.
- Coleman, Meri. "A Computer Readability Formula Designed for Matching Scoring." Journal of Applied Psychology 60 (1975):283.
- Cummings, L. L.; Schwab, Donald P.; and Rosen, Marc. "Performance and Knowledge of Results as Determinants of Goal Setting." Journal of Applied Psychology 55 (1971):526.
- Dale, Edgar, and Chall, Jeanne S. "Predicting Readability." Educational Research Bulletin, January 1948.

- Fang, I. E. "By Computer Flesch's Reading Ease Score and a Syllable Counter." <u>Behavioral Science</u> 13 (1968):249-251.
- Flesch, Rudolf. "A New Readability Yardstick." Journal of Applied Psychology XXXII (June 1948):221-33.

. The Art of Readable Writing. New York: Harper and Row, 1949.

- Fry, E. "A Readability Formula that Saves Time." Journal of Reading II (April 1968):514.
- General Motors Corporation. S.T.A.R. General Motors Computerized Simple Test Approach for Readability, 1976.
- Gibbs, C. B., and Brown, I. D. "Increased Production From the Information Incentive in a Repetitive Task." Reported in "Knowledge of Performance as an Incentive in Repetitive Monotonous Task," by Alphonse Chapanis, Journal of Applied Psychology 48 (1964):263.
- Haney, John B. "Readable Writing: Retrospect and Prospect." Journal of Communication 9 (1959): 186.
- Joseph, Albert. <u>Put It in Writing</u>. Cleveland, Ohio: Industrial Writing Institute, 1972.
- Kilian, G. L. "The Relationship Between Readability of Assigned Textbook and Reading Levels of Students in a School of Nursing." Ph.D. dissertation, Kansas State University, 1976.
- Kulhavy, R. W., and Anderson, R. C. "Delayed Retention Effect with Multiple-Choice Tests." Journal of Educational Psychology 63 (1972):507.
- Locke, Edwin A. "Motivational Effects of Knowledge of Results: Knowledge or Goal Setting?" Journal of Applied Psychology 51 (1967):324.
- Lorge, Irving I. "Predicting Reading Difficulty of Selections for Children." Elementary English Review XVI (October 1939):231-32.

_____. "Readability Formulas--An Evaluation." <u>Ele-</u> mentary English XXX VI (February 1949):89.

- McCormack, P. D. "Performance in a Vigilance Task with and Without Knowledge of Results." <u>Canadian</u> Journal of Psychology 13 (1959):68-71.
- ; Binding, F. R. S.; and Chylinski, Joanne. "Effects on Reaction-Time of Knowledge of Results of Performance." <u>Perceptual Motivation Skills</u> 14 (1962):367-372.
- ; ; and McElheran, W. G. "Effects on Reaction Time of Partial Knowledge of Results of Performance." <u>Perceptual Motivation Skills</u> 17 (1963):279-281.
- _____, and McElheran, W. G. "Follow-up of Effects on Reaction Time with Partial Knowledge of Results." <u>Perceptual Motivation Skills</u> 17 (1963):565-566.
- Newman, Edwin. <u>Strictly Speaking</u>. New York: Warner Books, 1974.
- Payne, R. B., and Hauty, G. T. "Effect of Psychological Feedback Upon Work Decrement." Journal of Experimental Psychology 50 (December 1955).
- Rubinoff, Morris, and Yovits, Marshall C., eds. Advances in Computers. New York: Academic Press, 1977.
- Report of the Commission on Federal Paperwork. Frank Horton, D-NY, chairperson. Washington, D.C.: U.S. Government Printing Office, 1976.
- Sassenrath, J. M. "Theory and Results on Feedback and Retention." Journal of Educational Psychology 67 (1975):894.
- _____, and Yonge, G. D. "Delayed Information Feedback, Feedback Cues, Retention Set, and Delayed Retention." Journal of Educational Psychology 59 (1968):70.
- Schleicher, J. Gordon. Michigan Department of Social Services, memo to Director of Bureau of Program Operations, 9 November 1976.
 - _____. Michigan Department of Social Services, memo to the Director, Bureau of Management and Staff Development, 17 November 1977.

Smith, Daniel Lloyd. "The Relationship of Feedback to Professors of the Results of Student Ratings of Their Teaching Effectiveness at Mid-Semester to Their End of Semester Ratings." Ph.D. dissertation, University of Southern California, 1977.

State of Michigan. Senate Bill 1061, 1977.

Sticht, Thomas G., et al. <u>Determination of Literacy</u> <u>Skill Requirements</u>. Cited in "Maybe They Can't Read the Manual." <u>Training in Business and</u> Industry, June 1974.

______. Determination of Literacy Skill Requirements in Four Military Occupational Specialties. Alexandria, Virginia: Human Resources Research Organization.

- Szalay, T. G. "Validation of the Coleman Readability Formula." Psychological Reports 17 (1965):966.
- Taylor, W. L. "Cloze Procedure: A New Tool for Measuring Readability." Journalism Quarterly 30 (1953): 415-20.
- Thomas, F. L. "The Extent of the Relationship Between Reading Achievement and Writing Achievement Among College Freshmen." Ph.D. dissertation, University of South Carolina, 1976.
- Travers, Robert M. W. Essentials of Learning. New York: MacMillan Company, 1972.

; Reid, Ian E.; and Van Wagenen, R. Keith. Reinforcement: A Review of Selected Research. University of Utah, 1963.

- Vaughan, Joseph, Jr. "Interpreting Readability Assessments." Journal of Reading.
- Yoakam, G. A. <u>Basal Reading Instruction</u>. New York: McGraw-Hill Book Company, 1955.